The protistan microbiome of German grassland soils

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Paul Christiaan Venter

aus Durban, Südafrika

Köln 2017

Berichterstatter: Prof. Dr. Hartmut Arndt

Prof. Dr. Michael Bonkowski

Tag der mündlichen Prüfung: 18.01.2018

Danksagung

Als erstes bedanke ich mich bei Gott für seine Hilfe.

„ Ich vermag alles durch den, der mich mächtig macht.” Philliper 4:13 (Lutherbibel 2017)

Meinem Betreuer, Prof. Dr. Hartmut Arndt, gilt mein ganz besonderer Dank, da er ein großes Beispiel der Wissenschaft und Inspiration für mich und meine Arbeit war. Unter anderem, ein „Danke schön“ für die Möglichkeit eine Promotion zu machen, die Betreuung meiner Arbeit, die viele Diskussionen und Anregungen zu höheren Zielen und auch die Unterstützung. Er war immer freundlich und bereit mich zu Helfen.

Ich danke ganz herzlich Herrn Prof. Dr. Bonkowski für die Erstellung des Zweitgutachtens.

Ich möchte dem Deutschen Akademischen Austauschdienst (DAAD; Abteilung ST32) danken für das Forschungsstipendium (Kennziffer: 57048249) für Doktoranden und Nachwuchswissenschaftler für den größten Teil meines Aufenthalts in Deutschland, bzw. für die Ermöglichung, mein Studium abzuschließen.

Zunächst möchte ich mich bei den Projektmitgliedern der Deutschen Biodiversitäts-Exploratorien-Initiative (http://www.biodiversity-exploratories.de/) (Fischer et al. 2010) bedanken. Besonderen Dank an die Leiter der drei Exploratorien, Swen Renner, Sonja Gockel, Martin Gorke und alle ehemaligen Manager, die die Grundstücks- und Projektinfrastruktur aufrecht hielten; Simone Pfeiffer für die Betreuung durch die Zentrale, Jens Nieschulze für die zentrale Datenbankverwaltung, Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser und die verstorbene Elisabeth Kalko für ihre Rolle beim Aufbau des Projekts Biodiversitäts Exploratorien. Die Untersuchungen wurden durch das DFG Schwerpunktprogramm 1374 "Infrastruktur-Biodiversitäts-Exploratorien" (in Teilen) gefördert (AR 288 16-2). Alle notwendigen Genehmigungen der zuständigen Umweltämter von Baden-Württemberg, Thüringen und Brandenburg lagen vor.

Ins besondere möchte ich Peter Heger, Sarah Carduck, Maja Ilic, Johannes Schöneich, Frank Nitsche und Anja Scherwaß ein größes Dankeschön für ihre emotionale Unterstützung und oft mühsamen technischen Beistand durch das ganze ausrichten.

Zuletzt aber am wichtigsten, mein Panda. Martin N. Venter, du bist mit mir durch Dick und Dünn gegangen. Ich freue mich, dich in meinem Leben zu haben.

„Vervielfältigung bedeutet das Ende von Vielfalt.“
MICHAEL RICHTER
(Einspruch, Halle (Saale): Mitteldeutscher Verlag, 2009, S. 25)

„Es hatte aber alle Welt einerlei Zunge und Sprache. Als sie nun von Osten aufbrachen, fanden sie eine Ebene im Lande Schinar und wohnten daselbst. Und sie sprachen untereinander: Wohlauf, lasst uns Ziegel streichen und brennen! – und nahmen Ziegel als Stein und Erdharz als Mörtel und sprachen: Wohlauf, lasst uns eine Stadt und einen Turm bauen, dessen Spitze bis an den Himmel reiche, damit wir uns einen Namen machen; denn wir werden sonst zerstreut über die ganze Erde.“
1. MOSE 11:1-4
(Lutherbibel 2017)

Inhalt

Zusammenfassung

Hoch-Durchsatz-Sequenzierungsverfahren wie NGS (Next Generation Sequencing) ist in ökologischen Untersuchungen und Biomonitoring als Methode einsetzbar, um große Probenentnahmegebiete und Probenzahlen zu parallelisieren. Bei der Untersuchung von sehr kleinen einzelligen Protisten wird diese Methode am besten als eine Kombination aus morphologisch-kulturbasierten und molekularen Techniken verwendet, damit die Defizite der jeweils anderen Methode ausgeglichen werden können. Das molekularen Verfahren, wurde jedoch bis jetzt oft als eigenständige Methode verwendet – daher sind die Verzerrungs- und Einschlusskriterien der NGS-Ergebnisse umso wichtiger. Aus diesem Grund konstruierten wir eine eigene Pipeline mit sehr konservativen Kriterien, damit sowohl sehr spezifische (Einzigartige Einzelne Sequenzen, Unique Individual Reads - UIRs) als auch generellere (Clustern in Operationelle Taxonomische Einheiten, Operational Taxonomic Units - OTUs) Treffer zur nächsten Referenzsequenz in der Protist Ribosomal Reference (PR^2) –Datenbank innerhalb paarweise ähnlicher Sequenzen-Schwellenwerte zu erhalten. Die Ergebnisse sind mit ähnlichen Studien vergleichbar, allerdings insofern einzigartig, dass eine detaillierte Analyse der ursprünglichen Sequenzen mit Umweltsequenzen aus anderen Studien direkt vergleichbar sind, um die Biogeographie der unbekannten Diversität zu vergleichen. Obwohl nur wenige Sequenzen (~1%) genau zu Protisten-Referenzsequenzen passten, wurden mit Schwellenwerten von Sequenz-Ähnlichkeiten eine große verborgene Diversität ohne Repräsentanten in der PR^2-Datenbank festgestellt. In dieser Dissertation wird die erste Taxa-Areal-Beziehung für Protisten in der Mesoskala (1 – 1000 km Entfernung zwischen Probenahmestellen) beschrieben, die unerwarteterweise größeren Tier- und Pflanzenarten ähnlicher sind als anderen Mikroorganismen (Pilze und Bakterien). Die Überlappung der

Taxa-Flächenarten wurde mit zunehmender Landnutzungsintensität (land-use intensity LUI - Weidevieh, Mähen und Düngung) geringer.

Durch die Kombination des Protisten-Datensatzes für die 150 Grünlandstandorte in der Mesoskala mit georeferenzierten Daten für insgesamt 12 unter- und oberirdische trophische Gruppen konnte eine echte multitrophische Homogenisierung gemessen werden, wenn sich die Diversität mit der Intensivierung der Landnutzung verändert. Die Hauptschlussfolgerung dieses multitrophen Diversitätsvergleichs war, dass die α-Diversität in unterirdischen Taxa mit einer erhöhten Landnutzung im Vergleich zu einer Abnahme der α-Diversität von oberirdischen Taxa zunahm, obwohl in beiden Fällen eine Homogenisierung stattfand.

Erneut wurde der georeferenzierte Grünland-Bodenprotisten-Datensatz in zwei weiteren Multifunktionalitätsstudien, Artenreichtum und Abundanz für neun unter- und oberirdische trophische Gruppen verglichen und zwei weitere erstaunliche Entdeckungen gemacht. Einerseits hatten die ober- und unterirdischen Arten gegensätzliche funktionelle Effekte, bei denen die seltenen Arten eher als die gewöhnlichen Arten mit einem erhöhten Ökosystem funktionieren und in ihrer Abundanz mit der Landnutzungsintensivierung abnahmen. Basierend auf einem angenommenen funktionellen Abwägungsprinzip bei seltenen Arten gehen wir davon aus, dass eine große Vielfalt an seltenen Arten mehr Vorteile für die Multifunktionalität bietet als eine große Vielfalt an verbreiteten Arten, unabhängig von der Intensivierung der Landnutzung und der untersuchten Region. Zweitens bietet der kombinierte multitrophische Reichtum eine bessere Erklärung für die Wirkung auf 14 Ökosystemvariablen (Dienstleistungen) als jede einzelne trophische Gruppe alleine, wo die kombinierte Bereitstellung von Diensten und Funktionen in trophischen Gruppen stärker war, wenn die Diversität hoch war. Dies unterstreicht nicht nur die funktionelle Bedeutung der Biodiversität, sondern auch den Fehler, der mit Analysen auf der Basis einzelner trophischer Gruppen verbunden ist.

Ein genauerer Blick auf die Vielfalt der gut dokumentierten monophyletischen Ciliaten und die vergleichbar weniger gut studierten polyphyletischen heterotrophen Flagellaten im Bodenprotisten-Datensatz deuteten in beiden Fällen auf eine große verborgene Vielfalt innerhalb der seltenen Arten hin. Die meisten UIRs mit 100 % paarweiser Identität stimmten eher mit anderen Umweltsequenzen als mit morphologisch beschriebenen Arten überein. Phylogenetische Analysen wiesen darauf hin, dass selbst UIRs, die den beschriebenen Referenzarten sehr nahe kamen, Varianten sein konnten, da sie eine individuelle Biogeographie aufwiesen. Eine weitere große verborgene Gemeinschaft könnte anhand einer unbekannten Umwelt-Diversität in der Referenzdatenbank (PR^2) und der bereits renommierten "seltenen Biosphäre" in diesem Datensatz beschrieben werden. Aus dieser Studie kann die Schlussfolgerung gezogen werden, dass die Entdeckung von Organismen im Boden zusätzlich zu technologischen Mängeln auf die Analysemethode zugeschnitten ist. Des Weiteren ist die Wiederfindungsrate der Arten von der ursprünglichen Fundstelle höher als die von entfernten Standorten.

Abstract

High throughput next generation sequencing (NGS) is a method used in ecological impact studies and biomonitoring to survey large sample numbers. When studying very small unicellular protists, this methodology is most optimally used in combination with morphological culture-based or other molecular techniques, to compensate the shortcomings of each method. In most studies to date, it was however used as a standalone method – therefore the biases and inclusion criteria of the NGS results become important. For this reason, we applied an own pipeline and very conservative criteria to be most inclusive (include all Unique Individual Reads – UIRs), but also very conservative (cluster to Operational Taxonomic Units - OTUs) within sequence pairwise similarity cut-offs to the closest reference sequence in the Protist Ribosomal Reference (PR^2) database. The results were comparable to other similar studies, but very unique in that a detailed analysis of the true sequences was possible, and pyrotags could be compared to environmental sequences of other studies to compare the biogeography of the unknown diversity. While only a very few sequences (~1%) strictly matched protist reference sequences, pairwise identity inclusion cut-offs identified a large hidden diversity with no representatives in the PR^2 database. In this dissertation, the first taxa-area relationship for protists in the mesoscale (1 – 1000 km between sampling sites) is described, being unexpectedly more similar to large animal and plant species than to other micro-organisms (fungi and bacteria). Taxa-area relationship of species overlap was discovered to decrease with increased land-use intensity (LUI – grazing livestock, mowing and fertilization).

Combining the protist dataset for the 150 grassland sites in the mesoscale with georeferenced data for altogether 12 below- and aboveground trophic groups, true

multitrophic homogenization could be measured as diversity changes with land-use intensification. A major conclusion of this multitrophic diversity comparison was that the α-diversity in belowground taxa increased with increased land-use, as compared to decreases in α-diversity of aboveground taxa, even though in both cases homogenization occurred.

Once again, including the georeferenced grassland soil protist dataset in two more multifunctionality studies, species richness and abundance for nine below- and aboveground trophic groups were compared and two more discoveries were made. First of all, the above- and belowground species had opposing functional effects, where the rare species rather than the common species associated with high ecosystem functioning, and declined in their abundances with land-use intensification. Based on a presumed functional trade-off principle among rare species, we assume that a high diversity of rare species is more advantageous for multifunctionality than a high diversity of common species, irrespective of land-use intensification and region studied. Secondly, the combined multitrophic richness had a stronger explanatory effect on 14 ecosystem variables (services) than any single trophic group alone, where the combined provision of services and functions in trophic groups were stronger when diversity was high. This not only underlined the functional importance of biodiversity, but also the error associated with analyses based on single trophic groups alone.

A closer look at the diversity of the well documented monophyletic ciliates and also the comparatively less studied polyphyletic heterotrophic flagellates in the soil protist dataset indicated a large hidden diversity in the rare species range, in both cases. Most of the pyrotags with 100% pairwise identity matched other environmental sequences rather than morphologically described species. Phylogenetic analyses indicated that even UIRs that were close matches to described refenerence species could be variants, because they displayed an individual biogeography. A further large hidden community could be described in terms of an unknown environmental diversity in the reference database (PR^2) and using the already renowned “rare biosphere” in this dataset. This study concludes that the discovery of

organisms in soil is tailored to the analysis method used, in addition to technological shortcomings. Furthermore, the recovery rate of species from the original site of discovery is higher than for taxa from distant sites.

General introduction

Protists – mostly single-celled eukaryotes - include photosynthetic, heterotrophic free-living and parasitic life forms (e.g. Adl et al. 2012). They form an important part of the microbial loop in terrestrial systems between bacteria and plants (e.g. Bonkowski 2004; Bonkowski and Clarholm 2012), as well as the phytoplankton of the open oceans and biofilms in freshwater aquatic systems (e.g. Azam et al. 1983; Arndt et al. 2000). They can feed on bacteria and other protists and relay the energy to higher trophic levels, although there are known exceptions to this over-simplified scheme (Geisen et al. 2016; Hess et al. 2012). In soils, protist communities may influence the growth rate of the aboveground plant species, like grasses, by its capacity to selectively graze on rhizobacterial communities and stimulating plant roots (e.g. Bonkowski and Clarholm 2012; Bonkowski and Roy 2005). In turn, aboveground human activity and environmental parameters have an impact on protist community structures, like that of grassland use in commercial farming (e.g. Allan et al. 2014; Gossner et al. 2016). Due to their small size, high turnover rate and abundance, protists can easily disperse and collect in soil like they do in aquatic sediments, forming a "seedbank" of living and dormant forms waiting to populate the environment under favorable conditions (Finlay 2002; Foissner 2006).

Protists are recorded to be the most diverse eukaryotes in soil (Mahé et al. 2016), with predicted dimensions as high as 85% of all soil taxa (see summary by de Vargas et al. 2015). Previous morphological soil surveys from Central European habitats yielded protozoan abundances of 1.7 to 12.7 x 10^4 individuals per gram dry weight (Domonell et al. 2013). These protist numbers included typical dominant soil taxa, like amoebae (~50%), cercozoans (~32%), stramenopiles (~8%), euglenozoans (~7%), apusozoans (~1%) and ciliates (~0.5%). In the rhizosphere naked amoebae and flagellates graze bacteria on the root surfaces,

stimulating root growth by releasing nitrogen compounds such as ammonia from the bacteria (Bonkowski and Clarholm 2012). Interactions of the belowground to aboveground trophic groups are however affected in both directions (vertically) and changes can also be indirectly induced. As such, a top-down and bottom-up relationship has been recorded to occur in terrestrial systems, between plants and micro-organism species, in that a loss in diversity in one trophic level will also affect the other (Scherber et al. 2010). Likewise, bottom-up effects can start with the bacterial chemical warfare against predators, where defense mechanisms of some prey safeguard their survival and grazers shape their community structure (e.g. Jouset et al. 2009). Diversity loss studies incorporating multiple trophic groups are therefore very important for our understanding of ecosystem functioning in response to ecosystem variables (Allan et al. 2014; Gossner et al. 2016; Soliveres et al. 2016a, b). This protozoa-bacteria-plant interactome needs more studying, where molecular, rather than morphological techniques may be more successful in future studies.

Due to their small size (about 0.002 to 2 mm; e.g. Finlay 2002) protists are traditionally studied under the microscope using morphological classification and cultivation based methods (e.g. Ekelund and Patterson 1997; Esteban et al. 2006; Foissner et al. 2004). These methods are often timeous and require expert taxonomic skill, because some protists may be overlooked when they are hidden behind soil particles in culture, dormant forms are not always present and most protists are simply uncultivable (e.g. Domonell et al. 2013; Ekelund et al. 2002; Foissner 2006). These difficulties led to the use of combined morphological and molecular techniques to accurately identify (Brabender et al. 2012; Foissner et al. 2004) and classify species based on marker genes, e.g. the 18S rRNA gene (Adl et al. 2012; Pawlowski et al. 2012). For this purpose, Sanger reference sequences were collected in curated databases (e.g. Protist Ribosomal Reference - PR² data base). In this way, the recovery of described species from environmental samples are made possible for meta-barcoding studies amplifying environmental sequences using high throughput sequencing

(HTS) technology (Guillou et al. 2013). The challenge now is to combine the semi-quantitative and biogeographic knowledge associated with HTS reads and the taxonomic knowledge of morphological studies. This combined knowledge is needed to gain insight into the ecology of protists, e.g. their adaptations to environmental conditions (Boenigk and Arndt 2002; Bonkowski and Clarholm 2012). In fact, a combination of methods is necessary to optimally study species diversity in natural habitats (Jeuck et al. 2017; Schoenle et al. 2016).

High throughput sequencing

HTS is especially powerful in contributing to the knowledge of the unknown diversity (Berney et al. 2004), the ecology and distribution of previously isolated reference sequences in environmental samples (e.g. Bates et al. 2013; del Campo and Massana 2011; Foissner et al. 2014; Geisen et al. 2015). There are, however, a number of pitfalls to HTS that need to be addressed, that can cause misinterpretation of the data. One is the fact that only a subgroup of the actual diversity is ever retrieved when sequencing an ever insufficient number of environmental samples, as is visually portrayed by species accumulation or rarefaction curves (Egge et al. 2013; Forster et al. 2016). Finding novel taxa in environmental samples, usually indicated by diverging similarity to known reference sequences, is most probably due to database shortfalls for the taxon in question (Berney et al. 2004; Forster et al. 2016). And even if the taxa that existed in reference databases covered all possible species in environmental samples, no single universal barcode primer can cover them all (Pawlowski et al. 2012), meaning that one is always limited to some subgroup of the actual community (Hadziavdic et al. 2014). All this implies that just because an organism was not detected, does not prove its absence in the environment sampled (Berney et al. 2004).

Only a hand full of comparative molecular surveys using next-generation sequencing (NGS) techniques have been performed on the earth's soil surface (e.g. Bates et al. 2013;

Geisen et al. 2015; Lara et al. 2011; Lentendu et al. 2014; Mahè et al. 2017), often disregarding scaling aspects and in many cases too few samples are taken to rule out undersampling. HTS surveys are more often employed to survey large areas of the earth, especially aquatic (marine and freshwater) environments (de Vargas et al. 2015; Forster et al. 2016; Lima-Medez et al. 2015; Medinger et al. 2010; Stoeck et al. 2010). Despite the above mentioned pitfalls, HTS is rapid, can accommodate large sample numbers in parallel to cover large sampling areas, making it a cost-effective alternative or supplement to culture dependent methods and record the presence of even uncultivable taxa (Hadziavdic et al. 2014; Schlebusch and Illing 2012). By using 454 and PCR amplifying a sequence length of ~700bp, amplicons covering the V4 region of the 18S rRNA gene are large enough to be used for reliable clustering methods and phylogenetic analyses among reference sequences (Dunthorn et al. 2014; Schlebusch and Illing 2012). Because of the error rates of PCR and the 454 sequencing platform, a number of robust filtering and analysis approaches have to be followed (Berney et al. 2004; Schlebusch and Illing 2012).

Sequencing errors and clustering

Next generation sequencing (NGS) is a term denoting all HTS platforms post-Sanger sequencing that makes use of a PCR amplification step (Schlebusch and Illing 2012). In this study we overcame many of the large hurdles associated with 454 HTS platforms and PCR errors, such as error rate and chimeric sequences (e.g. Berney et al. 2004) through strict quality filtering. First of all, a single PCR step amplified a ~710bp amplicon covering the variable 4 (V4) region of the SSU rRNA gene (Hadziavdic et al. 2014; Niklas et al. 2013), large enough to skip the library preparation by random shearing step (Schlebusch and Illing 2012). The multiplex identifier adapter primers were ligated to either end of the sequences, but sequencing was only performed in the forward primer direction, therefore no post

sequencing contig-forming step was necessary. A maximum sequence length exclusion step during data quality filtering excluded chimeric sequences, where conserved regions could anneal between sequences from distantly related taxa (Berney et al. 2014). Other filtering steps are mentioned under the Methods sections in Venter et al. (see publications) and not repeated here. Since the V4 region of the 18S rRNA gene is the most variable, it contains polymorphism induced length differences across individual taxa (Hadziavdic et al. 2014). Sequences were, therefore, cut to a maximum length of 530bp before the dereplication step to combine comparable data for the V4 region across taxa.

In a 530bp HTS query sequence, one single basepair difference to the original sequence will result in a 99.82% pairwise similarity between the two sequences. This new sequence is called a pyrotag variant. The 454 platform is known for its error rates caused by the inaccurate distinction between single insertion and deletions when long homopolymer runs occur during pyrosequencing (Schlebusch and Illing 2012). While studies claim that 454 technology error rates lie between 99.75 and 99.82% pairwise identity (Huse et al. 2007; Niklas et al. 2013), a drop in accuracy of 1% is only recorded to occur after the 400th base pair (Schlebusch and Illing 2012). When the PCR error rate caused by the polymerase enzyme is added to this error rate, one can safely assume an error rate of 0.03% as explained under the methods section. This makes the evaluation of individual HTS sequences so much more interesting and worth reporting on, because clustering these sequences with such a low variant rate at a common 97% sequence similarity will not only reduce the diversity, but also mask conspecific, or even congeneric species (Caron et al. 2009; Nebel et al. 2011). On the other hand, the risk of artificially inflating diversity has to be considered when matching pyrotags directly to reference sequences (Huse et al. 2010) and some compromise is needed. So we included and reported on both. We even included pairwise identity cut-offs for inclusion and very few sequences were lost (Venter et al. 2017).

Rare species

Independent of the clustering method used, HTS results usually produce a pyrotag consortium consisting of an abundant few and many rare species (Huse et al. 2010), where more than half of the pyrotags occur less than 10 times in the dataset (Venter et al. 2017). These rare pyrotags, may be representatives of the "rare biosphere" (species with low abundance) or variants of the original sequence caused by artifacts of PCR and pyrosequencing (Huse et al. 2010; Mahè et al. 2015). These variants often cluster to singleton operational taxonomic units (OTUs, Fig. 1) that occur in one sample only, making estimates on diversity and species pool size hard to predict (Chao 1987; Huse et al. 2010). Some researchers argue that all possible sequence variants must be eliminated by clustering into OTUs (Huse et al. 2010). This brings us back to the question regarding clustering or not. An alternative method is to evaluate these pyrotags (partial sequences) in a phylogenetic context with the appropriate reference sequences, as determined by an initial blasting protocol (Dunthorn et al. 2014). This method is database biased because it is a closed-reference protocol, but may make an accurate description of the rare biosphere possible.

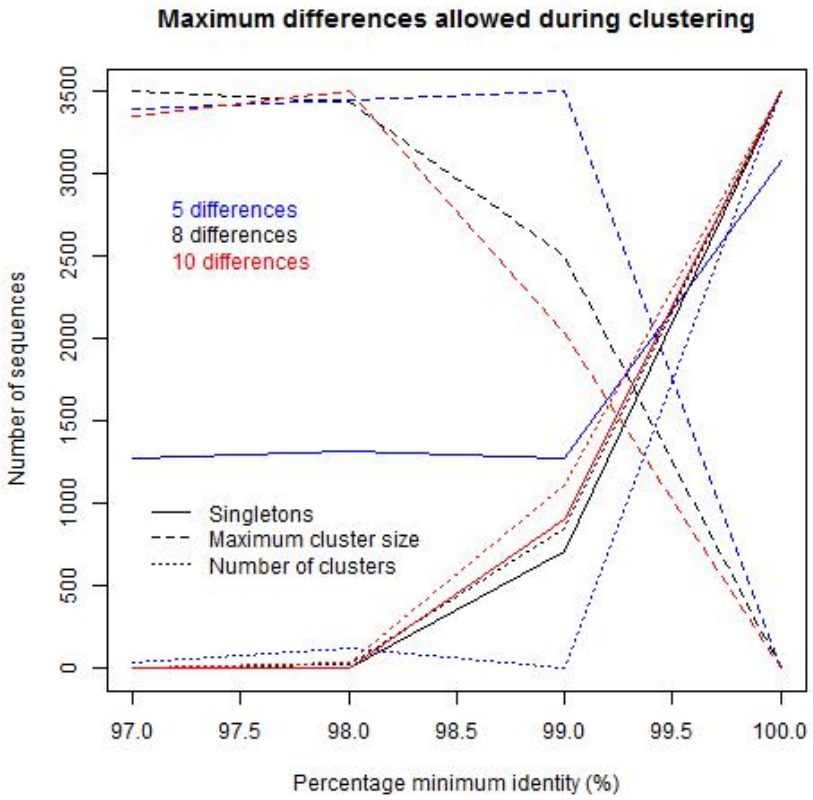


Figure 1: Clustering beyond a 99.7% pairwise identity is a pivotal point for operational taxonomic unit (OTU) formation. The analysis is based on a comparison of the number of differences between pyrotags included within OTUs and the resulting clustersizes being produced. Beyond 1% dissimilarity, the number of singletons increases drastically. Immage based on Swarm analyses of a subset of the data (Mahé et al. 2015).

Spatial scaling and taxa-area relationships

That organisms exhibit genetic divergence across space and time and therefore a definite distribution or biogeography is a law in ecology (Barreto et al. 2014). Taxa-area relationships ($S=cA^z$) consider the latitudinal gradient and effect of habitat loss on species diversity (Arrhenius equation; Oksanen 2017). This relationship between area size (A) and the number of species (S) that reside in it, is used to study the gradient (z) which depends on the overlap in species composition between sampled localities. This beta-diversity measure is taxon specific (c). The taxa-area relationships of microbial taxa such as Bacteria (z=0.02 - 0.04), Ciliates (z=0.04 - 0.08) and Fungi (z=0.07) are known (Green and Bohannan 2006), but no or very few data on the taxa-area relationship for protists exists or is yet understood. We

generally know that the z-value is larger for macrofauna (0.90) and flora (0.78) compared to micro-organisms (Barreto et al. 2014; Collins et al. 2002; Oksanen 2017). Geographic separation and isolation can lead to genetic distance, possibly driving the speciation process, because organisms are capable to adapt to extreme environmental niche areas like hot springs, hot and cold desserts and salt water lakes (Barreto et al. 2014). Ecologists are especially interested in this measure, with emphasis on a defined scale (Zinger et al. 2013; Green and Bohannan 2006), especially for protists – long thought to have ubiquitously biogeography (e.g. Esteban et al. 2006; Finlay et al. 2002).

Land use

Intensified land-use, by humans in an agricultural setting, may disrupt local species biodiversity, increasing species overlap and thus destabilize ecosystem functioning (e.g. Allan et al. 2014; Blüthgen et al. 2015; Fischer et al. 2010). Land-use intensity (LUI) is a well defined index for grasslands of Germany, including measures of fertilization, mowing frequency and livestock grazing intensity (Blüthgen et al. 2012). This standard has been used to study possible selective pressure on community structure, composition and distribution patterns of many taxa (above and below ground) within the mesoscale (Allan et al. 2014), but not for protists. Little is known about the effect of LUI on protists, other than that some taxa like cercozoans seem to associate with low land use and amoeba are more dominant under high land use conditions (Bonkowski and Clarholm 2012; Domonell et al. 2013). This information is paramount to formulate optimal conservation strategies and policies on agricultural activities to achieve effective sustainable land use (Fischer et al. 2010).

Aims and hypotheses of the study

The general aim of this dissertation was to increase the knowledge on protist diversity in

grassland soils. Within the scope of a temperate climate in middle Europe and within a defined mesoscale using HTS, it was aimed to describe how protist communities are influenced by geographic distance (biogeography), determine their actual taxa-area relationship and the influence of land-use.

Chapter 1 – The protistan microbiome of grassland soil: diversity in the mesoscale. In order to analyze the original HTS sequences for 150 samples, an alternative pipeline was developed to study more than one facet of HTS results. With the robust pipeline, it was aimed to gain high taxonomic resolution up to the genus and species level and report on similar sequences (matching UIRs) found elsewhere and thus gain insight into distribution issues such as local to global distribution patterns. To analyze the HTS pyrotags, an all-inclusive (unique individual reads - UIRs) and a more conservative (operational taxonomic units - OTUs) level was applied. Using universal primers for the SSU rRNA gene, we hypothesized to find not only the identity of typical soil protistan groups (Rhizaria, Alveolata, Stramenopiles) covered by the primers, but also their distribution patterns in central European grasslands. Using the taxa-area relationship for small protists, we aimed to find the distribution of protists in the mesoscale and hypothesized that this should be similar as for other small eukaryotes (e.g. fungi, ciliates) and prokaryotes (bacteria). By extension, this taxa area overlap should decrease with increased geographic separation. Regarding the influence of land-use intensity (LUI) on species diversity, it was hypothesized that more homogenous communities with more dominant species and higher overlap at sites with increased land use should be found. We hypothesized to uncover a large hidden diversity compared to morphological and cultivation based studies, where the rare biosphere would include the overlooked diversity, including some parasitic groups, not previously discovered in HTS soil studies with many samples in a defined mesoscale.

Chapter 2 – Land-use intensification causes multitrophic homogenization of grassland communities. This was planned as one of the first studies to survey the loss of biodiversity by homogenization due to increased land use intensity (LUI) across multiple communities in grasslands on such a large scale as the mesoscale. Selecting samples representative of a range of land uses and using the geo-referenced information on 12 trophic groups, the hypotheses were that LUI has a homogenizing effect with regards to species turnover (β-diversity). To evaluate this biotic homogenization: 1.) the effect of LUI on β-diversity was measured; 2.) the exact point of change in β-diversity with increasing LUI had to be found; and 3.) correlations between trophic groups were made to assess changes in β-diversity between them. The aim of the study was to assess local species loss (α-diversity) within each of the above- and belowground trophic groups with changes in increased LUI in the mesoscale. It was hypothesized that increased LUI will decrease β-diversity or compositional dissimilarity between sites, e.g. become more homogeneous, but that this should not necessarily be accompanied by reduced local or α-diversity. A hypothetical loss in specialist species during homogenization occurring linearly with increased LUI will reduce correlations between the β-diversity of different trophic groups and in such a way change the multitrophic community structure. Protist microorganisms, as one of the belowground groups, were hypothesized to be less affected by increased LUI to aboveground trophic groups such as plants and animals. Our contribution to this study was the georeferenced protist microbial dataset, to be considered as the bacterivorous trophic group.

Chapter 3 – Locally rare species influence grassland ecosystem multifunctionality. The aim of this part of the study was to assess the relative functional importance of rare species, compared to common species, in driving the biodiversity-multifunctionality relationship. Therefore the local richness and abundances for nine trophic groups were measured to assess how they associated to supporting, provisioning, regulating and culturing services (altogether

14 ecosystem factors) across a LUI gradient. The hypotheses were: 1.) that the diversity of common species, and not rare species, drive ecosystem multifunctionality; 2.) that multifunctionality driven by high diversity will decrease with increased LUI due to functional composition changes; 3.) that aboveground organism diversity is a stronger predictor of ecosystem multifunctionality; and 4.) that there are specific important indicator species that point out high multifunctionality. We contributed the georeferenced protist dataset as a belowground trophic group.

Chapter 4 – Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. This study hypothesized that ecosystem functioning cannot be simplified to single important drivers, but rather to multifunctionality within broader ecosystem functioning categories. This implies that simultaneous provisioning (called multifunctionality) is a concerted effort among trophic groups, rather than one indicator species and that species loss across many trophic groups, may cause stronger consequential loss of ecosystem functioning. The aim was to measure the relationships of the richness and abundance between trophic groups and ecosystem services. Again we contributed the georeferenced protists dataset as a belowground trophic group to this study. It was hypothesized that the combined richness of the trophic groups measured alongside 14 ecosystem services would have a stronger effect on ecosystem services than any individual trophic group. Hypothetically, the effects on individual ecosystem and categorical service types would be indicated by changes in richness and abundance of the multitrophic group, rather than by one individual trophic group. Multifunctionality can only be proven when biodiversity loss occurs across many taxa, indicating that the functioning of a single trophic group is dependent on the diversity of the other trophic groups.

Chapter 5 – Methodological studies on estimates of abundance and diversity of heterotrophic flagellates from the deep-sea floor. This study aimed at comparing the methodological issues between morphological (culture-dependent) and molecular (HTS) methods. Issues were discussed as they pertain to protist communities and sampling strategies as exemplified by studies of sediments of the deep-sea. The study aimed at supporting a combination of both methods for future studies.

Chapter 6 – Discrepancies between molecular and morphological databases of soil ciliates studied for temperate grasslands of central Europe. The aim was to attain phylogenetic and taxonomical monophyletic associations between the ciliate taxa within the protists, even when including pyrotags with reference sequences in a phylogenetic analysis. We hypothesized that, because this group contains so many morphologically described species, it is especially likely that this group can be used to address questions regarding restricted biogeography. Here we aimed at describing the knowledge gap that exists between existing morphological knowledge and the molecular entries of known taxa deposited in the Protist Ribosomal Reference (PR^2) molecular database. We aimed to show that this may hamper describing the biogeography of ciliate taxa groups in georeferenced environmental samples and finding gradients in community structure due to environmental variables (e.g. LUI and edaphic parameters). The PR^2 database is limited in that it containes all available reference sequences for morphological described species in GenBank and only a curated subset of the sequences from environmental samples. In our dataset, less than 1% of UIRs and OTUs had 100% matches to these reference sequences. Because ciliates are a monophylotic group, we aimed at relating these UIRs and OTUs to a taxonomic lineage based on phylogenetic association and pairwise distance to the closest reference sequence. We hypothesized that because the ciliate group contains many flagship species and has been studied for a very long time in central Europe, 1.) an increased recovery rate of known

flagship and well-studied species is to be expected, 2.) most pyrotags should have close sequence similarity to reference sequences in the PR^2 database with many exact reference sequence matches. We also hypothesized that a deviation from the previous hypotheses would most probably be due to discrepancies between the molecular and morphological databases. Regarding the reference database, we hypothesized that: 1.) taxa for which there are many reference sequences in the PR^2 database should also be equally well presented in our data, and 2.) a phylogenetic analyses would be more accurate at determining species richness, and 3.) recovered ciliate richness should be comparable to that recovered in other molecular and morphological studies.

Chapter 7 – The hidden diversity of flagellated protists in soil. In this part of the study, we aimed at gaining deeper knowledge on individual taxa grouped under the flagellated protists. It was hypothesized that an in-depth study of the HTS pyrotags for typical rare and dominant soil flagellates may uncover a large unknown diversity. Due to the often polyphyletic associations of flagellate taxa within the same lineages, phylogenetic affiliation could not be used to elaborate on ecological function. To strengthen our methods and underscore the uniqueness of this study, we aimed to determine the recovery rate of 10 cercozoan species previously isolated, cultured and described from the sampling sites in the same region. Having the original reference sequences and comparing the HTS pyrotag variances to these, it was hypothesized that the recovery rate should be high for described species, when the original site of discovery is surveyed compared to sites located further away. It was aimed to investigate the biogeographical importance of individual pyrotages and their variants associated with known (ten described species), typical abundant (cercomonads - *Sandona*) and rare taxa (apusozoans) in soil to evaluate the recovery rate of described to unknown groups in soil.

Ultimately, it was hypothesized that many unknown species, not previously reported in

other molecular HTS studies, will be discovered. Using the robust pipeline and critically analyzing the pyrotags, it was hypothesized that many pyrotags have an own unknown biogeography, which may be masked by commonly used clustering methods in OTUs. These variants to centroids in OTUs may indicate a deepened hidden diversity. Using phylogenetic analyses, the aim of this study was to prove the importance of the large hidden diversity of flagellated protists in the rare biosphere, which could potentially be masked by OTU-clustering.

Chapter 1 - The Protistan microbiome of grassland soil: diversity in the mesoscale

Protist, Vol. 168, 546–564, xx 2017
http://www.elsevier.de/protis
Published online date 10 April 2017

Protist

ORIGINAL PAPER

The Protistan Microbiome of Grassland Soil: Diversity in the Mesoscale

Paul Christiaan Venter[a], Frank Nitsche[a], Anne Domonell[a], Peter Heger[b], and Hartmut Arndt[a,1]

[a]University of Cologne, Department of Biology, Institute of Zoology, General Ecology, Zuelpicher Str. 47b, D-50674 Koeln (Cologne), Germany
[b]University of Cologne, Department of Biology, Institute for Genetics, Bioinformatics & Population Genetics, Zuelpicher Str. 47a, D-50674 Koeln (Cologne), Germany

Submitted October 10, 2016; Accepted March 30, 2017
Monitoring Editor: David Moreira

Genomic data for less than one quarter of ~1.8 million named species on earth exist in public databases like GenBank. Little information exists on the estimated one million small sized (1–100 μm) heterotrophic nanoflagellates and ciliates and their taxa-area relationship. We analyzed environmental DNA from 150 geo-referenced grassland plots representing topographical and land-use ranges typical for Central Europe. High through-put barcoding allowed the identification of operational taxonomic units (OTUs) at species level, with high pairwise identity to reference sequences (≥99.7%), but also the identification of sequences at the genus (≥97%) and class (≥80%) taxonomic level. Species richness analyses revealed, on average, 100 genus level OTUs (332 unique individual read (UIR) and 56 class level OTUs per gram of soil sample in the mesoscale (1–1 000 km). Database shortfalls were highlighted by increased uncertain taxonomic lineages at lower resolution (≥80% sequence identity). No single barcode occurred ubiquitously across all sites. Taxa-area relationships indicated that OTUs spread over the entire mesoscale were more similar than in the local scale and increased land-use (fertilization, mowing and grazing) promoted taxa-area separation. Only a small fraction of sequences strictly matched reference library sequences, suggesting a large protistan "dark matter" in soil which warrants further research.

Key words: Molecular ecology; soil; diversity; land-use intensity; spatial distribution; taxa-area relationship.

Introduction

Molecular surveys in planktonic marine systems covering two thirds of the Earth surface have unveiled a large diversity of small protists (de Vargas et al. 2015). Conversely, much less information exists on the remaining one third of the earth surface covered by soil. Studies on soil ecosystems have typically focused on the smallest (prokaryotes) and larger organism size classes (nematodes, insects) (Allan et al. 2014). Recent global surveys of organismal diversity in the ocean revealed that small eukaryotes comprise by far the major part of eukaryotic genotype diversity (de Vargas et al. 2015; Sunagawa et al. 2015) indicating an enormous functional diversity within this size class (Lima-Mendez et al. 2015). During the last 5 years, soil protist biodiversity has increasingly been studied using high through-put next-generation sequencing (NGS) techniques

[1]Corresponding author; fax +49 221 470 5932
e-mail hartmut.arndt@uni-koeln.de (H. Arndt).

http://dx.doi.org/10.1016/j.protis.2017.03.005

(e.g., Bates et al. 2013; Geisen et al. 2015; Lara et al. 2011; Lentendu et al. 2014). Although NGS studies covered large areas in some cases, often few samples were taken and scaling aspects were not taken into consideration.

The issue of spatial scaling and distribution patterns, i.e. taxa-area relationships, is a central topic in ecology (Green and Bohannan 2006). Martinus Beijerinck (1898) hypothesized that microbial life is uniformly distributed due to its small size (<2 mm) and that the 'environment selects' its inhabitants, implying high turnover rates (Fenchel and Finlay 2006). According to this hypothesis, local sites are expected to harbor 10^4 to 10^7 active individuals of live protist species per gram of soil (Bates et al. 2013) and a huge number of encysted species in their "seedbank", representing a significant proportion of global diversity (Finlay 2002). This implies a moderate slope in the species-area curve resulting from low global species richness and the absence of endemic species for geographically distant sites. It remains a matter of sampling intensity to prove this high local/global species ratio (Woodcock et al. 2006). In contrast to this a moderate endemicity model for protists was recently proposed by Foissner (2006). Apart from that, many rare taxa and cryptic species may exist as a result of both diversification and endemism of dispersed organisms across spatial scales, prompted by their short generation times and rapid genetic divergence (Horner-Devine et al. 2004). Because no taxa-area curve is available for protists, we investigated aspects of biogeographical scaling in soil protist communities by analyzing sampling sites in distances of one up to 1 000 km.

Here, we present a comprehensive study of protist diversity in soils, using a large sample size (n = 150) within a defined mesoscale. We would like to study whether local to global ratios of soil protists was comparable to records for marine systems (de Vargas et al. 2015) including new or seldom recorded lineages. In order to obtain optimal resolution of most protistan supergroups, we used the most widely studied SSU (small subunit) rRNA gene as a marker gene in metabarcoding approaches together with a curated database for the SSU rRNA gene (Guillou et al. 2013). We focused on the common soil protistan supergroups Rhizaria, Alveolata, Stramenopiles, Excavata and Opisthokonta, which are well-covered by general SSU primers (Pawlowski et al. 2012). We aimed to obtain high taxonomic resolution and hence possible unknown protistan diversity by using long sequence read lengths (>500 bp). This increased the reliability of taxonomic assignment (Hadziavdic et al. 2014; Lentendu et al. 2014; Wang et al. 2014) and resolution at protistan supergroup level (de Vargas et al. 2015; Taib et al. 2013).

Results and Discussion

NGS and Barcoding of Protists

Environmental sequences from ecological studies using next-generation sequencing (NGS), usually bin these query sequences into meaningful operational taxonomic units (OTUs). In these approaches many cleaned sequences are lost (Caron et al. 2009). To optimize our analyses and include, if possible, all of the unknown diversity of soil protists, we had to modify existing pipelines (Supplementary Material Fig. S1). This is because sequences that bin into OTUs are thought to share similar ecological affiliation and hence are suitable to model ecological consistency (Koeppel and Wu 2013; Preheim et al. 2013). In this closed-reference based approach, trimmed NGS sequences (maximum length 530 bp) were filtered and dereplicated to produce unique individual reads (UIRs). UIRs identifed singleton reads (abundance = 1) which were removed. During de novo sequencing a singleton is an OTU represented by a single NGS sequence, and the sequence could have an abundance of >1. This singleton OTU could be informative of a species and an indicator of rare biosphere lineages (Zahn et al. 2013). In our study, singletons were also retained, because we identified it differently. We defined singletons as dereplicated reads with an initial abundance of 1. These singletons with an initial abundance of 1 were removed to circumvent the dangers of pyrosequencing related artifacts (Tedersoo et al. 2010). UIRs were classified directly to the closest reference sequences (accession number) in the curated Protist Ribosomal Reference (PR^2) database (based on GenBank, version 203, downloaded June 2016) (Guillou et al. 2013; Supplementary Material Figs S2, S3), the largest database for 18S rDNA sequences. This form of closed-reference-clustering to an accession number also identifies operational taxonomic units (OTUs), because more than one UIR can be linked to a single accession number. In order to retain the most accurate information in complex communities, compensate for errors and artifacts of pyrosequencing and to get an upper and lower estimate of the real species richness of protists, the dereplicated query sequences or UIRs (upper estimate) were analyzed alongside the number of OTUs (conservative estimate).

As an alternative to *de novo* clustering, we applied a closed-reference approach and assigned query UIRs to the taxonomy of the first blast hit (Supplementary Material Figs S1, S2). These hits could be clustered to a proxy level within the Linnaean taxonomy after blasting them, where sequences similar to previously observed taxa returned higher pairwise identity values as compared to novel diversity (Edgcomb et al. 2011; Forster et al. 2016; Schloss and Westcott 2011; Westcott and Schloss 2015).

Binning UIRs into OTUs is dependent on pairwise identity as high cut-off values can artificially inflate diversity and therefore species richness. On the other hand lower levels might include congeneric UIRs (Caron et al. 2009; Nebel et al. 2011). Because of the unresolved species concept (Caron et al. 2009) and differences in the rates of evolution within the V4 region of the SSU across taxa (Stoeck et al. 2010), we classified UIRs to be an exact match, or approximate species taxonomic level, when UIRs were identified at ≥99.7% sequence similarity. This took into account: the quality score of 454 sequencing technology (between 99.75%; Huse et al. 2007 and ∼99.82%; Niklas et al. 2013) and the expected *Pfu* polymerase error rate (2.6×10^{-6}; Thermo Scientific) during PCR (Huse et al. 2007; Stoeck et al. 2010). For our 530 bp reads, this translated into a difference of a single base pair between two UIRs. The ≥97% pairwise identity level, commonly used as a binning value for species taxonomic level was used to allow comparisons with other studies and therefore used as a conservative proxy level to separate genera (Caron et al. 2009). UIRs were included up to the final cut-off of ≥80% pairwise identity, a level which stood as proxy for class taxonomic level (Stoeck et al. 2010) and an attempt to avoid false placement in higher taxonomic ranks (Nebel et al. 2011). We only removed very few reads <80% pairwise identity (11 OTUs (or 652 UIRs to be exact)), all other UIRs passing the filter criteria were retained.

Based on common blast hit results and pairwise identity thresholds for UIRs, the lineages for these assigned UIRs could be grouped into OTUs for respective taxonomic levels (Supplementary Material Fig. S2). In conventional molecular ecology using de novo clustering, UIRs would be binned into OTUs where similar sequences fall within the cut-off sequence similarity (e.g. 97%) of a centroid sequence, hence masking other UIRs with possible higher dissimilarity to one another and disregarding differences in the rates of evolution across lineages (Nebel et al. 2011). Even though direct closed-reference approaches may be influenced by database quality and therefore classification errors, assignments are comparable with other studies and do not change OTU numbers centroids as sequences are added to the dataset (Caron et al. 2009; Westcott and Schloss 2015). On the other hand, OTU centroids are assigned to a taxonomic lineage that is used to identify the entire OTU. Therefore, erroneous identification due to database insufficiencies may have a far greater effect (Westcott and Schloss 2015). In this closed-reference method, all UIRs above thresholds associated with the same reference sequence are included in a single OTU. Therefore de novo clustering is still possible for sequences with higher dissimilarity to reference sequences and, in contrast to open-reference methods, OTUs may contain UIR sequences across several levels of sequence similarity or taxonomic ranks. Only 11% of all UIRs could not be matched with 100% query sequence coverage to a reference sequence in the PR^2 database. The number of reads (UIRs) to percent similarity with the reference database (Supplementary Material Fig. S2) is similar to that for open marine ecosystems (Mahé et al. 2016), indicating a shared eukaryote discovery profile and confirming assumptions that the PR^2 database is biased toward marine and temperate terrestrial reference sequences (Mahé et al. 2016). After rigorous filtering of the obtained >1.2 million raw reads (mean length = 548 bp), about 35% of all eukaryotic OTUs identified in our study belonged to protists when removing fungi, plant and animal-associated OTUs (Supplementary Material Table S1). Although retrieved sequencing read numbers and protist supergroup proportions are within a range similar to previous soil studies (Bates et al. 2013; Geisen et al. 2015; Lentendu et al. 2014), this value is far lower than that obtained from tropical forests (Mahé et al. 2017) and marine plankton studies (de Vargas et al. 2015), even when species richness is inflated to include non-fungal protist UIRs (20% protist UIRs) – the focus of our study. However, the average ratio of "local" ($1 m^2$ mixed sample) richness to "global" (mesoscale of 1,000 km) richness was $14.4 \pm 7\%$, similar to the local to global species ratio estimated for oceanic systems ($9.7 \pm 4\%$) (de Vargas et al. 2015). A frequency distribution curve of these OTUs indicated that protistan species richness was slanted towards the rare taxa range (Supplementary Material Fig. S3). The molecular survey of soil samples from 150 grassland sites distributed over a distance of about 1,000 kilometers revealed 8,407 UIRs for protists (Supplementary Material Table S1). A species accumulation curve analysis of these reads was able to approach protist

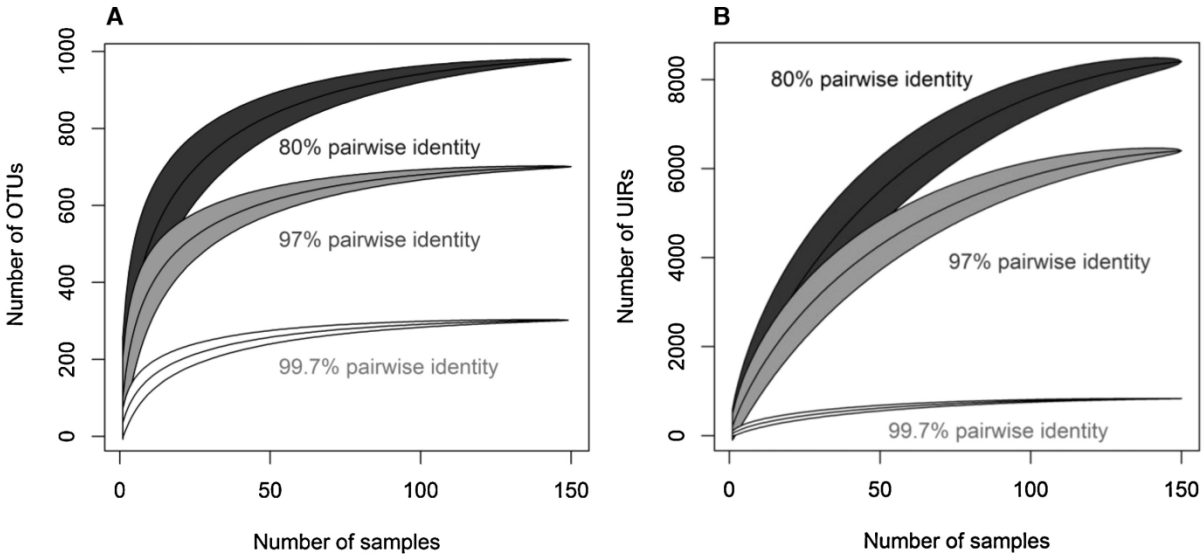


Figure 1. Species accumulation curves representing (**A**) the number of operational taxonomic units (OTUs) (**B**) and unique individual reads (UIRs) at ≥80%, ≥97% and ≥99.7% sequence pairwise identity for all 150 sites. Species accumulation curve and standard error lines at respective taxonomic levels are according to Kindt's exact accumulator rarefaction-like model (32). Cut-offs represent proxy taxonomic levels of class (≥80%), genus (≥97%) and species (≥99.7%) according to pairwise identity of OTUs/UIRs to reference sequences in the PR^2 database.

richness saturation at ≥97% reference sequence similarity cut-off, but only when UIRs clustered to known taxa (accession numbers – OTUs) in the PR^2 database (Fig. 1A) and not if UIRs alone were used (Fig. 1B). The UIR set may therefore contain many previously overlooked protist species. In agreement with this assumption, the taxa richness at ≥80% sequence identity (as opposed to ≥99.7%) in the accumulation curves (Fig. 1A) indicated that with increased sampling effort still more OTUs and UIRs (Fig. 1B) can be discovered (see also: Massana et al. 2011). The Chao species pool-size estimate (Chao 1987; Oksanen et al. 2015) confirms this assumption (Supplementary Material Table S1), although probably underestimating it, due to singleton elimination (Supplementary Material Table S1).

To describe the protist community structure obtained for the mesoscale, we created a multi-value schematic tree (Fig. 2). Although ~137,000 protist reference sequences have been catalogued in the PR^2 database (v203), only about 0.6% could be recovered in our soil UIR data set at 99.7 – 100% nucleotide similarity. Among the 55 deep protistan class lineages in our data set, 20% belonged to uncertain taxonomic lineages, suggesting a high number of understudied taxa. This is typical for environmental high throughput sequencing studies (Bates et al. 2013; de Vargas et al. 2015; Guillou et al. 2013) and indicates that most soil microbiome protists have not previously been sequenced. Because a sequence identity distance of more than 3% roughly implies that no biological sequence for a ribotype exists in the database (see Methods), their affiliation to known taxonomic groups has to be inferred using phylogenetic means (Forster et al. 2016; Mahé et al. 2016; Šlapeta et al. 2005; Stoeck et al. 2010) or de novo clustering (Westcott and Schloss 2015). By open-reference methods, these de novo sequences are clustered to OTUs. However, it was realised/revealed early that such OTUs are rarely biologically meaningful and can erroneously group different morphological species into a single OTU (Caron et al. 2009). Even rare taxa occurring only once, but present in more than one sample (Wang et al. 2014), are excluded as singletons (1 OTU represented by 1 sequence) and doubletons (1 OTU represented by 2 sequences) in some studies (Lentendu et al. 2014; Simon et al. 2015) because they are likely to be artefacts of pyrosequencing (Huse et al. 2010). Depending on the sequence similarity cut-off value of inclusion into clusters and the clustering method settings, these OTUs can be artificially inflated (Forster et al. 2016; Mahé et al. 2015) and therewith the number of singletons. Therefore, we stuck

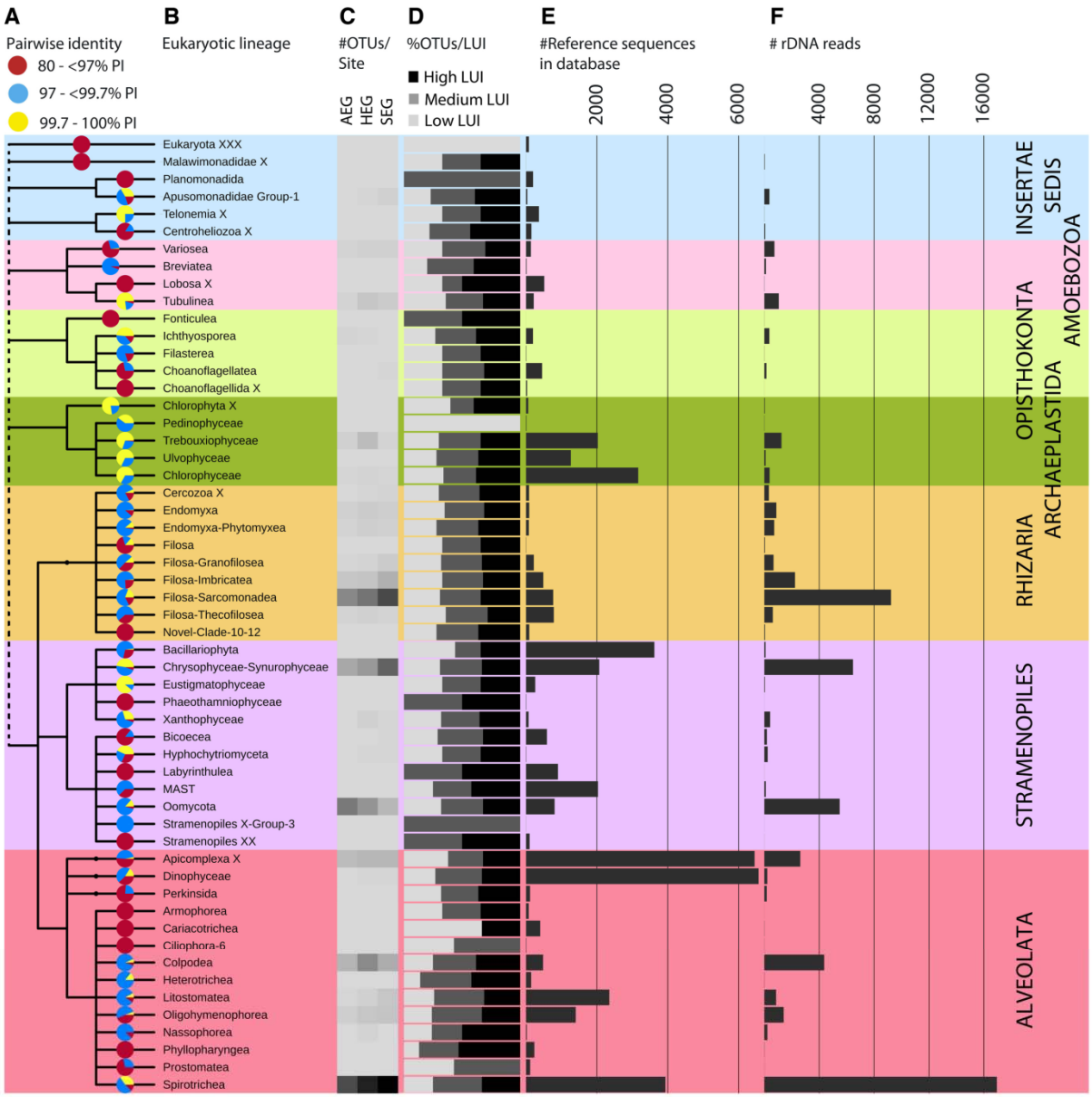


Figure 2. Taxonomic distribution of assignable protist operational taxonomic units (OTUs) from 150 Biodiversity Exploratory samples within a phylogenetic context. (**A**) Pie charts superimposed on the unrooted schematic phylogenetic tree indicate the percentage of OTUs with ≥80 to <97% (red), ≥97 to <99.7% (blue) and ≥99.7 to 100% (yellow) pairwise identity to the Protist Ribosomal Reference (PR^2) database. (**B**) Tree leaves indicate class (≥80% reference sequence identity) taxa (PR^2 database) to which OTUs were assigned. Unresolved eukaryotic classes are indicated by an X after the class name, e.g. Eukaryota_XXX from the Eukaryota_X supergroup are putative eukaryotic ribosomal DNA (rDNA) sequences. (**C**) Heat maps for the number of site specific OTUs (≥80% sequence identity) separated for Schwaebische Alb (AEG), Hainich-Duen (HEG) and Schorfheide-Chorin (SEG) (total read numbers in F). (**D**) Bar chart for the percentage of OTUs (≥80% sequence identity) associated with land-use intensity (LUI) index levels: Low LUI (light grey), medium LUI (dark grey), high LUI (black). (**E**) Bar chart indicating the number of class specific reference sequences present in PR^2 database. (**F**) Bar chart indicating the abundance of rDNA reads assigned to reference sequences in PR^2 database at ≥80% sequence identity.

to the direct annotation of UIRs and comparing data with previous studies for sequences with >3% dissimilarity.

In our study, more OTUs were observed in the data sets at the <97% sequence identity level compared to the dataset for sequences with ≥97% sequence identity to a reference sequence, because UIRs with lower pairwise identity to a reference sequence were considered (Fig. 2A and B). At this lower resolution (≥80% to <97% sequence identity) more class taxa were observed for ciliates, lobose amoebozoans and opisthokonts. Class level (≥80%) OTUs for Hacrobia and the two supergroups Excavata (Malawimonadidae) and Apusozoa (Apusomonadidae) disappeared with increasing sequence identity. Although being less similar to reference sequences, we found this level of resolution to be more reliable when considering that query sequences are less sensitive to unequal supergroup entries in the PR^2 database (Hadziavdic et al. 2014). Therefore, the OTUs at lower resolution (≥80% to <97%) become very important in illuminating gaps in the database and represent an inviting starting point for more research in soil molecular ecology (Pawlowski et al. 2012; Stoeck et al. 2010; Taib et al. 2013).

Hereafter, we concentrate mainly on the conservative ≥97% sequence identity OTU estimate to allow dataset comparisons with other published data. Among protists, the major part of OTUs comprised rhizarians (35.3% of OTUs), followed by alveolates (28.2%), stramenopiles (18.3%), archaeplastids (12.7%), amoebozoans (3.0%), opisthokonts (1.6%), hacrobians (0.6%), and apusozoans (0.4%; Table S2). Ribosomal barcodes for heterotrophic protists, including plant and animal parasites (~11%) and free-living heterotrophic protists (~76%), were much more diverse than autotrophic protists (~13% of OTUs) across all soil sites (Fig. 2). This is in consensus with studies of other biomes such as marine pelagic systems (de Vargas et al. 2015), rainforest soils (Mahé et al. 2017) and peat bogs (Lara et al. 2011), reenforcing the challenges with traditional concepts on heterotroph-to-autotroph ratios (de Vargas et al. 2015). Due to the coverage limitation of single primers for variable regions of the SSU rRNA gene (Hadziavdic et al. 2014; Pawlowski et al. 2012), only few OTUs were discovered for groups like Amoebozoa and Opisthokonta. The dominating rhizarian, alveolate and stramenopile OTUs of the terrestrial protist microbiome (Fig. 2) were also well represented in the PR^2 database (>30%) implying that results are highly database dependent.

Group-specific Diversity

Rhizarians: An interesting finding contradicting aquatic (de Vargas et al. 2015; Šlapeta et al. 2005) and other soil studies (Bates et al. 2013; Lentendu et al. 2014) was the prevalence of amoeboid heterotrophic rhizarians (245 OTUs) over alveolates (196 OTUs). Although this finding is not unusual for forests and grassland soils (Geisen et al. 2015), this is contrary to recent findings by Mahé et al. (2017) which indicated the importance and overwhelmingly dominant contribution of alveolate apicomplexans in Neotropical forest soils, and to findings by Bates et al. (2013) who reported a dominant contribution of alveolates over rhizarians in all kinds of soil ranging from arid to humid conditions. The relative contribution of the different phylogenetic groups might not only be attributed to the diversity of communities studied, but may also be due to the selection of the "universal" primers used in the different studies. The primer selection may significantly influence the resolution of the community and coverage of the different organisms (e.g. Pawlowski et al. 2012). For a more complete coverage of protistan communities in soil, it is suggested to use group-specific primers (Lentendu et al. 2014). In comparative studies comprising a large number of samples, the shortcomings of a universal primer cannot be solved yet. But at least a comparison of different communities analyzed in exactly the same way can be carried out.

Here, we found that Filosa-Sarcomonadea, a spore forming phylum within Cercozoa, contributed the highest number of OTUs (175) of all rhizarians. Presence/absence data indicated that three of the most common genera (taxa on the ≥97% identity level) belonged to rhizarians represented by an undescribed cercomonad (at 116 of 150 sites), an undescribed allapsiid (at 134 sites) and an undescribed leptophryid vampyrellid (at 115 sites) algal feeder. Although OTUs from grassland sites were more cosmopolitain up to the species level than in previous soil studies (Bates et al. 2013; Lentendu et al. 2014), no single OTU or UIR sequence among all rhizarians was ubiquitous. As expected, foraminiferans did not occur in our samples, because group-specific primers would have been necessary for their detection (Lejzerowicz et al. 2010; Pawlowski and Lecroq 2010).

Alveolates: This taxonomic group contributed the highest number of unique sequences peaking in the phagotrophic ciliates (175 OTUs; 2 317 UIRs). Assuming that small sequencing errors inflated UIRs, the total number of OTUs may not be inferred due to the general multinuclearity of ciliates as com-

pared to flagellates (Weber and Pawlowski 2013). Ciliates peaked in the class Spirotrichea (65 OTUs) of which one undefined stichotrich species was present at 139 of 150 investigated sites. Among the ten most common OTUs in grassland sites was the stichotrich ciliate *Bistichella cystiformans*, previously described from Chinese soil (Fan et al. 2014). It was found at 135 sites (one 100% identical UIR was found at 125 different sites). Other commonly found ciliates were members of the genera *Oxytricha* (at 125 sites) and *Colpoda* (at 119 sites). Colpodid ciliates, like most other soil ciliates, can form cysts. This is a very powerful property that allows the survival of taxa in dynamic soil communities over time (Dopheide et al. 2008; Šlapeta et al. 2005). Besides ciliates, many alveolate sequences in our study belonged to known parasitic groups of apicomplexans (15 OTUs), gregarines (9 OTUs) and perkinsids (2 OTUs, see below). Dinoflagellate sequences were much less frequently found (4 OTUs). One dinoflagellate OTU was 100% identical to the photosynthetic planktonic marine microalga *Lepidodinium chlorophorum* (Hansen et al. 2007), here discovered in soil.

Stramenopiles: The most frequently recorded OTU across all three Biodiversity Exploratory sites was a stramenopile. The oomycotan plant parasite *Pythium* was identified in at least 141 of the 150 sampling sites. Accurate identification could be a pivotal diagnostic step, since this pathogen may cause damage to a wide range of plant species (Robideau et al. 2011). In some instances, more than one UIR was clustered to the same OTU, keeping in mind that each UIR had its own similarity values to reference sequences. These singlet UIRs themselves where widely distributed within the mesoscale. E.g. one of the most abundant UIRs of the oomycotan plant parasite *Pythium* OTU (0.0% pairwise distance to *Pythium attrantheridium* [HQ643476] according to the K2P (Kimura 2-parameter model)) was identified at 62 of the 150 sites. In the same way, one MAST-12C associated UIR, occurring at three sites, closely matched a sequence (0.002 divergence, >99.6% sequence similarity) from GenBank previously detected from groundwater limestone aquifers. This implied that even though stramenopile OTUs were the most frequently detected, the diversity of this supergroup in soil was most probably underestimated.

In our below-ground samples, we also found several photosynthetic protists such as diatoms, eustimatophyceans, chrysophyceans and xanthophyceans, typically present in high numbers in freshwater and marine plankton (Berney and Pawlowski 2006; Šlapeta et al. 2005). As proof of this, at least 21 UIRs resembling *Spumella*-like flagellates were discovered at 108 of the 150 sites. The same *Spumella* sp. OTU, formerly isolated form freshwater (Findenig et al. 2010), was present at 128 of 150 sites, together with an unknown chrysophycean clade C_X (at 135 of 150 sites), previously isolated from river water (Nakamura and Suto 2012), were found among the ten most abundant organisms in our grassland samples. Chrysophytes are characteristically cyst formers and their presence could also be used to infer past climatic conditions (Findenig et al. 2010). Four of the 17 marine stramenopile OTUs (MAST-12C) in the PR^2 database typical for freshwater, marine plankton and oxic non-marine sediments (Massana et al. 2014) were discovered in all three biodiversity exploratory sites.

Important sub-dominant sequences: Apart from the dominating SAR supergroup, some lineages owe the scarcity of their detection to the coverage limitation of single primers for variable regions of the 18S rRNA gene (e.g. Amoebozoa, Opisthokonta) (Hadziavdic et al. 2014; Pawlowski et al. 2012) or to a lack of described taxa (e.g. Hacrobia). Hacrobian OTUs detected included some phagotrophic lineages of Centroheliozoa (3 OTUs) and Telonemia (1 OTU) mostly known from the ocean and freshwater. Amoebozoans, too, were only scarcely detected (*Hartmanella, Flamella*) although they resemble one of the most dominant protist groups in soil (Geisen et al. 2015). Two important opisthokontan phyla were detected using SSU rRNA primers. These were choanoflagellates (5 OTUs), which can adapt to oxic and anoxic conditions (Šlapeta et al. 2005), and the mesomycetozoans (5 OTUs) represented by ichthyosporean animal parasites (Ichthyophonae and Rhinosporidaceae) and filastereans (*Capsaspora*). The ichthyosporean OTUs detected (*Anurofeca* sp.) are rarely detected using sequencing techniques (Taib et al. 2013). In seven samples, we discovered that nine *Capsaspora owczarzaki* UIRs existed for the one OTU. Capsasporans are predacious aerobic amoeboid producing pedunces that can penetrate animal cells. In addition, three apusomonad OTUs were detected for *Apusomonas proboscidea*, however, distance matrices (K2P >0.03) indicated the possibility that UIRs may belong to more than one species (Massana et al. 2011). This confirmed the general problem in metabarcoding of soil samples, regarding incomplete databases.

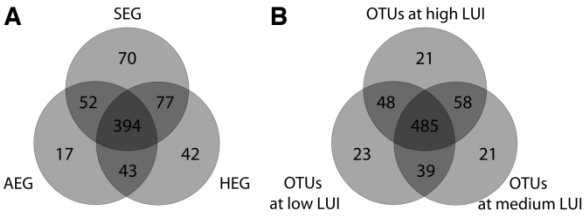


Figure 3. Venn diagram indicating soil protist OTU overlap between (**A**) geographically distant sites (AEG, HEG, SEG), and (**B**) degrees of LUI. For A and B, OTUs were identified at ≥97% sequence similarity.

Diversity of parasites in soil: Oomycetes are plant pathogens of agronomical importance, seldom detected using NGS (Bates et al. 2013; Geisen et al. 2015). Here, universal primers detected dominating oomycete communities in grassland samples (see *Pythium* sp. above). Another pathogen devastating potato crops, *Phytophthora infestans*, was detected as a dominant part of the Plasmodiophorida community (9 UIRs with <0.01% divergence to the reference sequence). One OTU for a plasmodiophorid protist, which is responsible for the transmission of several plant viruses (*Polymyxa graminis*) was identified in 101 of 150 sites with >99.8% sequence identity (84 UIRs with <0.02% divergence), of which one UIR was detected at 52 sites. *Polymyxa* primarily proliferates in grasses and is an obligate biotrophic plant root parasite and vector of various plant viruses like rosette disease. Other spore forming plant parasite genera of the Plasmodiophorida (*Spongospora* and *Plasmodiophora*) (Kanyuka et al. 2003) were also detected. Again, correct identification would be of pivotal prophylactic importance because this oomycete can cause damage to a wide array of plant species in agriculture (Robideau et al. 2011). Our study of 150 soil samples supports the recent findings of Geisen et al. (2015) from the analysis of 12 soil samples that oomycotan plant parasites are ubiquitous components of soil protistan communities.

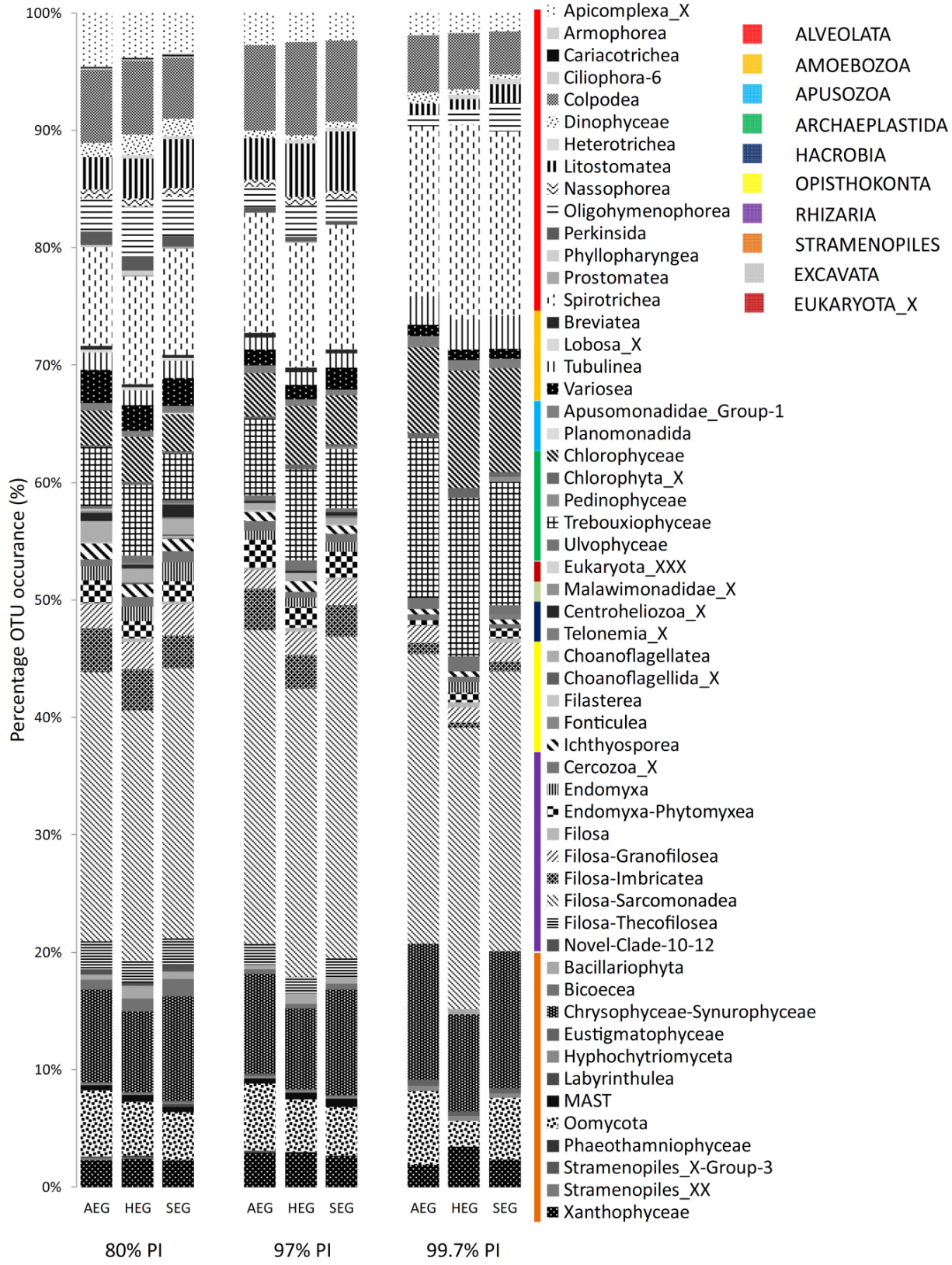
As animal parasites, ichthyosporeans can play an important role in arable soil (Lentendu et al. 2014). Detecting ichthyosporeans (Ichthyophonae and Rhinosporidaceae) was previously regarded as problematic using sequencing techniques, because of their low environmental abundance (Taib et al. 2013). We, however, detected 5 OTUs (clustering 43 UIRs), one of which occurs at 64 sites, and of this OTU, a single UIR with 100% sequence identity to *Anurofeca richardsi* was identified at 54 of the 150 sites. This parasite of anuran larvae may influence the anuran community composition aboveground (Baker et al. 1999), while the reference sequence (GenBank accession number: JN054668) was previously isolated from an activated sludge system for a waste water treatment plant in Australia.

Besides ichthyosporeans, apicomplexans and perkinsidans (Alveolata) are known as common parasites of invertebrates (Adl et al. 2012) and have been reported to be abundant in high throughput sequencing of terrestrial habitats (Geisen et al. 2015). Some of them may not only act as parasites, but also as symbionts (e.g. gregarines). In our grassland samples, apicomplexans (15 OTUs) were ubiquitous across all sites, although at higher resolution no ubiquitous OTUs were found and several clusters of orphan parasite sequences (>0.01 divergence) were discovered with high incidence (a gregarine taxon occurred at 54 of 150 sites), closely matching (0.01% divergence) environmental sequences from intracontinental (e.g. Switzerland, Lara et al. 2011) as well as intercontinental sites (e.g. USA, Lesaulnier et al. 2008). The perkinsids (2 OTUs) are not only intracellular pathogenic parasites of animals, but may often occur in their free-living stage as part of their lifecycle in freshwater systems (Lara et al. 2011). Their presence in soil may be associated with host activity or widespread cyst dispersal. This is in accordance with recent studies from different soil sites (Bates et al. 2013; Geisen et al. 2015; Lentendu et al. 2014).

For various freshwater systems, parasites were found to contribute to OTU diversity between 2 and 7% and it was assumed that these systems can stably tolerate less than 10% parasites (Simon et al. 2015). Here we report a much higher percentage of up to 14.1% for middle European terrestrial systems (0 – 34.6%; SD = 5.2; CL = 0.29). Recent soil studies have already indicated the importance of protistan parasites in soil (Bates et al. 2013; Geisen et al. 2015). In a recent metabarcode study of tropical rainforest soils using the V4 region of the SSU and universal primers, Mahé et al. (2017) reported that around 84% of all their high throughput reads came from hyperdominant OTUs affiliated to apicomplexan parasites. However, these samples have to be put into the context of the Neotropical climate where they were sampled among other hyperdiverse and dominating plant and animal species. These sequences included a large range of taxa, most of which are unknown and possibly new to the middle European climate zone. This indicates the need for further comparative studies.

Dominance of rare species: In our comparative study of 150 grassland sites, about 52.5% of

554 P.C. Venter et al.



the total sequences occurred less than 10 times at ≥97% sequence similarity (Fig. S3), declining with an increase in the percentage of sequence similarity from ≥80% (∼88%) to 99.7% (∼74%). Regardless of the methods applied, protists in soil seem to consist mostly of rare taxa (Van Dorst et al. 2014) with low-abundance, whether native or non-indigenous, and more sensitive to environmental change (Fig. 3, Supplementary Material Fig. S4). It has often been reported that the few dominant core species in an ecosystem are also the main drivers of ecological patterns (Allan et al. 2014; Van Dorst et al. 2014). The majority of taxa were rare and mostly unknown, however, our study indicates their importance among several known protist groups (Fig. 4, Supplementary Material Fig. S3). This phenomenon has also been observed for non-protist taxa (macrophytes, bacteria, invertebrates) across grassland sites (Allan et al. 2014) and in ecosystems in general (e.g. Woodcock et al. 2006).

Influence of Land-use

Intensified land-use influences biodiversity by disrupting local communities (Allan et al. 2014) and therefore presents a major threat to biodiversity. To measure the effect of land-use, a land-use intensity (LUI) index was devised to measure the different kinds and intensities of fertilization, mowing frequency and livestock grazing intensity (Bluethgen et al. 2012). This standardization reduces the complexity of identifying the influence of land-use on protist taxa richness and its possible selective pressure on community structure, composition and distribution patterns (Fig. 4, Supplementary Material Fig. S4). When we grouped samples according to their LUI index values (differentiating three levels of intensity – low, medium and high), few non-overlapping protist OTUs were identified (Fig. 3B) compared to geographically separated sites (Fig. 3A). The most dominant non-overlapping OTUs belonged to the SAR group, but occurred in only four samples at high LUI (unknown oomycotan stramenopiles). Correlating the LUI index with OTU richness per site led to a significant, but weak positive correlation (0.173, $p < 0.05$, Pearson), which suggested an increase in protist diversity with intensified land-use – as was found in a multitrophic level study by Soliveres et al. (2016b). When the correlation between OTU richness to LUI index within each locality was tested, significant but low positive correlations were found for Schwaebische Alb (0.549, $p < 0.01$, Pearson), slightly positive correlations were obtained for Hainich-Duen (0.217, $p < 0.05$, Pearson) and no correlations existed for Schorfheide-Chorin (−0.044, $p = 0.762$, Pearson). We suspect that the abundance of rare taxa and the increase in non-overlapping OTUs (especially in the Schwaebische Alb) work together as a response mechanism to the negative effects of increased land-use (Soliveres et al. 2016a,b). This may be seen as a possible buffering response to retain multifunctionality in soil systems when land-use increases make communities less homogeneous (see taxa-area relationships) and retain the capacity of ecosystems to provide services necessary to agricultural needs (Allan et al. 2014; Soliveres et al. 2016b). Additionally, dormant protist stages (detectable by rRNA gene sequencing, Santos et al. 2015) may temporally be insensitive to unfavorable land-use, but ready to proliferate upon environmental cues (Simon et al. 2015).

Taxa-area Relationship and Geographical Distribution

Organisms exhibit genetic divergence across space and time and therefore a distinct biogeographical distribution (Barreto et al. 2014). This ecological law is best described by the taxa-area relationship, the measure of change in community composition across space (Whittaker 1960). The Arrhenius taxa-area model ($S = cAz$) is a measure of beta diversity. It relates the number of species or taxa (S) discovered to the area size (A) sampled and is sensitive to environmental parameters (c) like land-use, soil characteristics and climate which all influence the microbial turnover rate (z-value) across space (Terrat et al. 2015).

Taxa-area relationships between sampling sites were statistically standardized to the same sequence number size to avoid biased community comparisons (see Methods). This left 32 samples within each of the three Biodiversity Exploratory sites, comprising 19,200 sequences. Contradictory to studies on bacteria in soil (Horner-Devine

Figure 4. Bar charts indicate the relative number of unique operational taxonomic units (OTUs) for all samples ($n = 50$) within each of the three Biodiversity Exploratories for each class taxonomic group at ≥80%, ≥97% and ≥99.7% pairwise identity (PI). Class level taxa are grouped within eukaryotic supergroups. Unresolved eukaryotic classes are indicated by an X after the class name, e.g. Eukaryota_XXX from the Eukaryota_X supergroup are putative eukaryotic ribosomal DNA (rDNA) sequences.

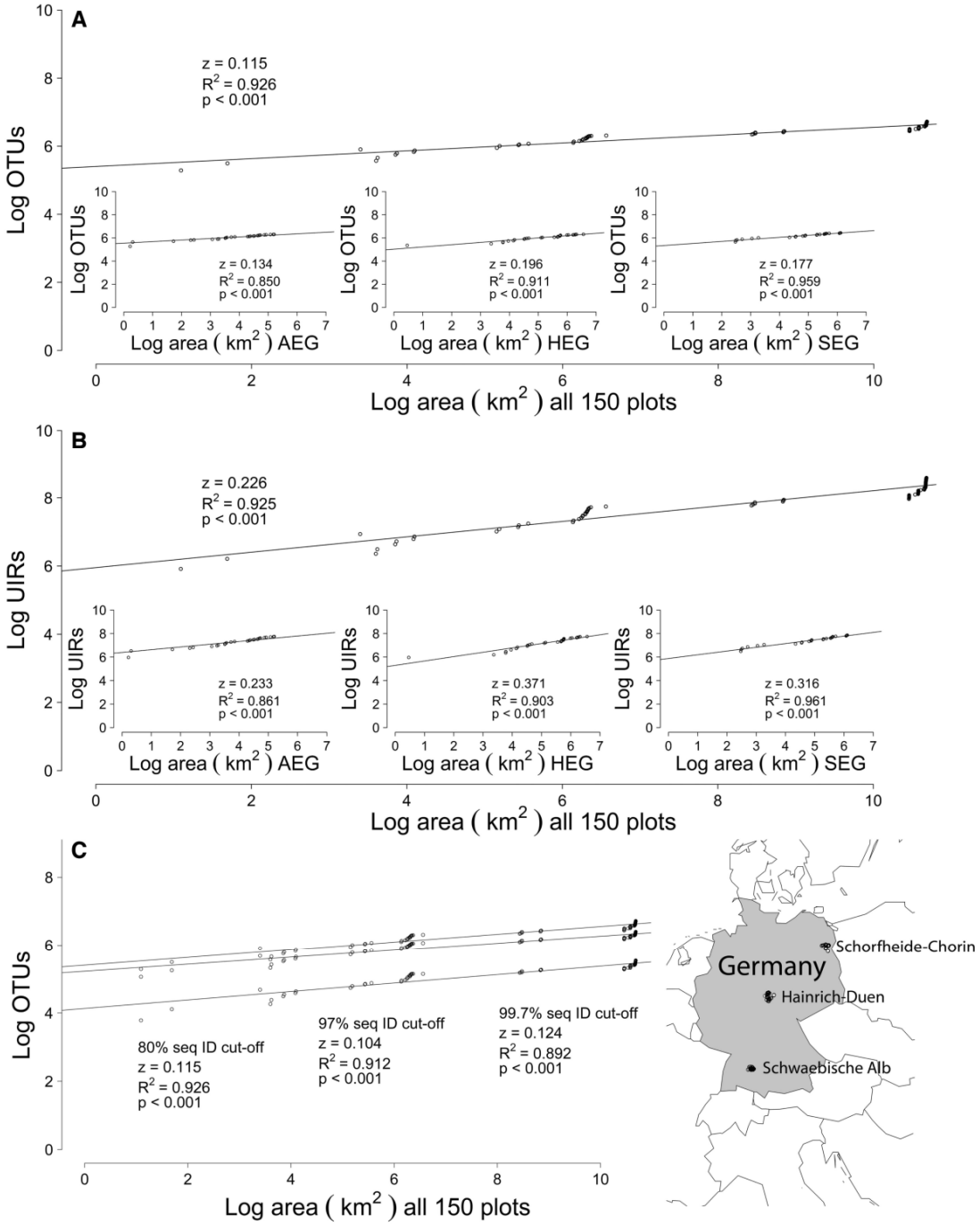


Figure 5. Log-log transformed taxa-area curves for all Biodiversity Exploratory sites (n = 150) in Schwaebische Alb (AEG), Hainich-Duen (HEG) and Schorfheide-Chorin (SEG). (**A**) Taxa-area relationship of operational taxonomic units (OTUs) for all sites and individual exploratories. (**B**) Taxa-area relationship of unique individual reads (UIRs) for all sites and individual exploratories. (**C**) Taxa-area relationship of OTUs according to ≥80% (upper line), ≥97% (middle line) and ≥99.7% (lower line) sequence pairwise identity cut-offs (seq ID cut-off). Distances are based on geographic coordinates and data for OTUs/UIRs based on presence/absence data, calculated at ≥80% reference sequence similarity. The slope of the taxa-area curve (z-value), variance in OTUs/UIRs by area (R-squared) and the significance (p-value) is given (SI Methods).

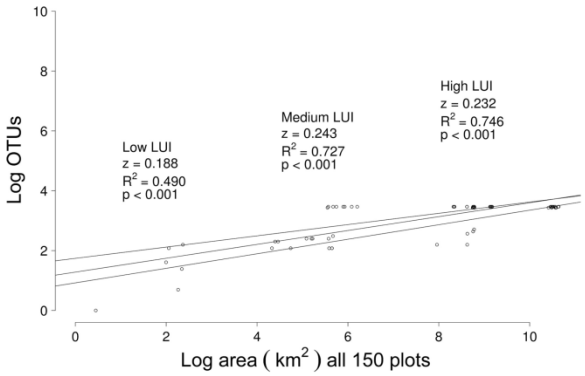


Figure 6. Taxa-area curves for all Biodiversity Exploratory samples (n = 150) at various levels of land-use intensity (LUI). Geographic distance was calculated using the coordinates for all sites and presence/absence data for protist operational taxonomic units (OTUs) determined at ≥80% reference sequence similarity. The slope of the taxa-area curve (z-value), amount of variation in species by area (R-squared) and the significance (p-value) are given (low LUI = upper line, medium LUI = middle line, high LUI = lower line). Presence/absence data for OTUs indicating taxa-area gradients (*z*-value) in relation to area size (km^2) were adjusted for statistical evenness (see Methods). Mean distance between Biodiversity Exploratory sites: ø311 km (<1 – 626 km).

et al. 2004), our study revealed that communities (OTUs and UIRs) separated by thousand kilometers were more similar to each other (mesoscale $z = 0.115$, $R^2 = 0.926$) than the taxa within each of the biodiversity exploratory sites (Fig. 5A and B). Even more contradictory, the degree of overlap (z-values/gradient) we obtained at all sequence similarity cut-offs (Fig. 5C) were comparable with that for macro-organisms like plants and animals (z = 0.1 – 0.3; Green and Bohannan 2006) in a more contiguous habitat (Barreto et al. 2014; Oksanen et al. 2015) and not as small as that of bacteria (z = 0.02 – 0.04), ciliates (z = 0.04 – 0.08) and fungi (z = 0.07) (Green and Bohannan 2006). The highest taxon overlap (OTUs) was observed among plots related to low LUI; indicated by a lower z-value (Fig. 6). Although high land-use indicated fewer taxa in common and therefore an increase in beta-diversity (heterogeneity index), the beta-diversity Bray-Curtis indicated that the OTU composition was similar amongst samples (Supplementary Material Fig. S4A). Between sites comparison of the OTUs communities using the Jaccard similarity index (Supplementary Material Fig. S4B) indicated that the OTUs communities overlapped greatly amongst exploratory sites. Gossner et al. (2016) found that among identical sampling sites, the alpha-diversity also increased with LUI (Gossner et al. 2016) which explains the increase in non-overlapping taxa in our results, when between sample comparisons are omitted. The damage induced by land-use (low, medium and high) is biotic homogenization, which causes subtle changes in beta-diversity (Supplementary Material Fig. S4A and B). This decrease in multidiversity with increased land-use intensity has been detected for above and below ground taxa among plants, invertebrates and vertebrates too (Allan et al. 2014; Gossner et al. 2016). We also found that species accumulation curves at various sequence similarity cut-offs were not saturated at UIRs level, implying that an insufficient number of samples were analysed in this study to draw final conclusions about the full extent of soil protist diversity, and that the slopes of taxa-area relationships may be overestimations (Fig. 1). Nevertheless, the z-values estimated in our study corresponded well with a recent study by Terrat et al. (2015) where microbial community z-values were in the range of that for macro-organisms. High z-values may be related to the accuracy of molecular barcoding methods. Analysing microbial sequences is more sensitive than microscopic methods in discriminating species from one another, thus reducing the community overlap between sites (Woodcock et al. 2006). These results challenge the hypotheses related to organism body size and distribution (Finlay 2002) and support earlier studies based on morphotypes (Foissner 2008). Protists obviously display a spatial structure in grassland soil.

Geographic separation and isolation may infer genetic distances and drive speciation processes. For instance, extreme environmental niche areas like hot springs, hot and cold deserts and salt water lakes are proven to contain adapted protist communities with non-cosmopolitan distributions (Barreto et al. 2014). When comparing OTU richness (presence/absence) data between the three Biodiversity Exploratory sites, dominant OTUs within each exploratory mainly belonged to the SAR supergroup (Fig. 2A, C and E). Even so, OTUs were significantly different between the three sites ($F\ ratio = 6.88$, $F\ crit = 3.06$, $p < 0.01$). At ≥97% sequence identity, up to 12% of the taxa were exclusive to a single Biodiversity Exploratory site (Fig. 2C and F). Among these, a stramenopile oomycote (*Phytophthora* sp.) dominated in Schwaebische Alb (2 OTUs). In Hainich-Duen, three ciliate taxa (*Vorticella microstoma, Vorticella* sp. and *Phacodinium*

sp.), an unknown diatom and four archaeplastidan taxa (*Bracteacoccus* sp., *Coenocystis* sp., *Desmochloris* sp., *Deasonia* sp. (Chlorophyta)) were dominant. Remarkably, within a rhizarian vampyrellid genus (*Platyreta* sp.), some phylotypes (UIRs) occurred in as many as 15 sites of the Schorfheide-Chorin exploratory only.

Even though supergroup-specific taxa were ubiquitous, selective global transport due to organism size and encystment may have led to clear species distribution patterns seldomly reported or possible to record in such depth of taxonomic resolution when applying barcoding data for protists. Several protists such as the ciliate *Colpoda* are extremely well adapted to be distributed by air (Mueller and Mueller 1970), while there is obviously a selective transport of most of the other protists via air (Feldmann 2007; Rogerson and Detwiler 1999). Several of the group-specific taxa with 100% identity to marine taxa most probably occurred in the salt-rich soil crusts which may have been transported to the Ah horizon via bioturbation.

Conclusions

Identifying ecological patterns and community structures in ecology is method-biased, even when using the most modern NGS technique. We suppose that undersampling, based on UIR species accumulation curves and competing sequences from algal, fungal, plant and metazoan organisms allowed only few OTUs to be retained after strict filtering steps. Here we confirm that DNA databases lack a huge amount of protist reference sequences and that protists, especially parasitic species, are much more diverse than previously thought (Bates et al. 2013; Mahé et al. 2017). All this warrants a robust approach in sequencing analysis, with access to dataset-true sequences (UIRs) as fingerprints for species (as an upper limit) in addition to database defined OTUs (as a lower limit) to estimate taxa richness. In our study, using UIRs revealed up to ten times more taxa than OTUs equaling the expected load per gram of soil (10^4-10^7) (Bates et al. 2013). Parasites, mostly dominant in soil, are heterogeneously spread; some sequences could be detected across continental distances. Their correct identification is of conservational and agricultural importance. Our modified approach in UIR sequence annotation is an attempt to limit inaccurate assignments allowed by sequence similarity cut-offs, revealing an enormous functional diversity, previously overlooked for protists in soil. We suppose that parasites form a significant part of the “rare biosphere” and “seed bank” (e.g. Ichthyosporea), as well as the dominant taxa (e.g. the stramenopile *Pythium* sp.) – and not just Apicomplexa, as was expected, waiting to proliferate upon favourable temporal environmental change. Understanding protist community response to environmental conditions is paramount to promote conservation and agricultural efficiency (Allan et al. 2014). In this study we found a higher species overlap between sites associated with different LUIs and not geographic separation. Overlapping taxa included almost all known and dominant species; where OTUs occurring >10 times in all samples made up 52.5% of all known taxa (≥97% pairwise identity). This is confirmed by the lower taxa-area relationship *z*-values across the mesoscale compared to the individual biodiversity exploratory sites (AEG, HEG and SEG). The greater overlap occurred mainly among sites associated with lower land-use, where this homogeneity is disturbed with increased land-use. The greater part of these protists are heterotrophic, compared to the proportion of phototrophic protist OTUs (∼13% of sequences). Although no single taxon or UIR was present in all samples, most known taxa were distributed across all sampling sites. This indicates that spatial separation does not dominate species richness, but rather availability of nutrients (e.g. LUI). There seems to be little geographic limitation and dominant as well as rare species are mainly limited by extrinsic factors like LUI. The results of the analysis of spatial distribution patterns contradict the current idea of a ubiquitous distribution of single protist taxa and support the moderate endemicity or cosmopolitan model (Foissner 2008). Less than 1% of genotypes matched curated library sequences, illuminating a large protistan “dark matter” in soil.

Methods

Data collection and soil sampling: Soil samples for the mesoscale were collected in May 2011 as part of the German Biodiversity Exploratories initiative (http://www.biodiversity-exploratories.de/) (Fischer et al. 2010). They cover 150 grassland soil samples from three temporally and spatially scaled geo-referenced study plots: the UNESCO biosphere reserve Schorfheide-Chorin (SEG) in north-eastern Germany, Hainich-Duen national park (HEG) in central Germany and the Schwaebische Alb UNESCO biosphere reserve (AEG) in south-western Germany. The co-ordinates and parameters at the time of sampling are given in Supplementary Material Table S1 (also see map in Fig. 5). Standardized field sampling (Brabender et al. 2012; Fischer et al. 2010) was performed. To summarize the procedure, fourteen soil cores (diameter, 8.3 cm) were

taken from 20 × 20 m size subareas, selected to represent a range of land-use intensities (LUI, Bluethgen et al. 2012). Soil samples were cored out from the upper 10 cm of the A horizon (core size 8.3 cm), the top most 5 cm root-layer was removed as well as any deadwood and roots larger than 2 cm in diameter. Samples were homogenized and stored at 4 °C while still at field moisture content. LUI index for the year 2011 was calculated from fertilization intensity (organic and mineral fertilization excluding livestock dunging), mowing frequency and grazing intensity (livestocks). Data was obtained from land owners by questionnaires (Bluethgen et al. 2012) and was applied in this comparative study to test land management and its effects on species richness. The three regions differ in climate, geology and topography and are representative of large parts of Central Europe (see Supplementary Material Table S2).

DNA isolation, PCR amplification and NGS: Whole genomic DNA was extracted from 1 g of each composite soil sample using the PowerSoil® DNA isolation kit (Mo Bio Laboratories, Inc., Carlsbad, CA) and quantified using a Nanodrop 100 spectrophotometer (Thermo Fisher Scientific, Germany). DNA quality was assessed by 2% Agarose gel electrophoresis. Overall DNA concentration was adjusted to 100 ng/μl using DEPC deionized water (ddH_2O). The highly variable V4 region of the 18S rRNA gene was directly amplified from the samples using the eukaryotic specific primers 590F (5′-3′:CGGTAATTCCAGCTCCAATAGC) and 1300R (5′-3′:CACCAACTAAGAACGGCCATGC). To separate the sequences, the Titanium primer design and the recommended multiplex identifier (MID) adaptor complex design (Roche, Germany) method was used. The pre-454 sequencing PCR reaction contained: 2 μl (100 ng/μl) DNA, 2 μl 10x DNA polymerase buffer with 20 mM $MgSO_4$, 2 μl (1 μM) 590F primer [590F + Key + MID + Adaptor A] and 2 μl (1 μM) 1300R primer, 2 μl (2 mM) dNTP each and 0.4 μl (2.5 U/μl) Pfu (*Pyrococcus furiosus*) DNA polymerase (Fermentas, Germany) and filled up to a total volume of 25 μl with ddH_2O. Pfu polymerase was used because of its high fidelity (2.6×10^{-6} error rate) through 5′ to 3′ exonuclease activity. Cycling conditions were: initial denaturation at 95 °C for 3 min; 30 cycles at 95 °C for 30 s, 55 °C for 30 s, 72 °C for 2 min; and a final extension at 72 °C for 10 min. PCR product quality was evaluated by agarose gel electrophoresis (2%) and its quantity was determined by spectrophotometry on a Nanodrop 100 (Thermo Fisher Scientific, Germany). The pre-454 sequencing PCR reaction was amplified in triplicate and pooled to a final concentration of 20 ng/μl to eliminate possible PCR bias. NGS using the GS-FLX sequencer with Titanium sequencing kit XLR70 (Roche, Germany) was performed by GATC Biotech AG, Germany. Sequencing was done as from the forward primer (adaptor A).

Removal of biased amplicons: Unidirectionally sequenced DNA traces received back from GATC Biotech AG were demultiplexed by means of the barcode in the pyrotags (adaptor A) and received in FastA format. Using standard command line, raw sequences were filtered for (1.) 100% forward primer match to remove false positive PCR amplifications of non-rRNA genes; (2.) minimum sequence length of 200 bp to remove possible artifacts; (3.) a maximum sequence length of 710 bp; and (4.) ambiguities (N's), to exclude sequences containing uncertain base pairs. Raw sequences are available from the authors upon request. Sequences reported in this paper have been deposited in the GenBank database under accession numbers SRP101780.

Bioinformatical identification of OTUs: Raw sequences were scanned for chimeric sequences against the curated Protist Ribosomal Reference (PR^2 v203) database (Guillou et al. 2013) using the uchime_ref algorithm in the USEARCH v. 7.0.1090 package (Edgar et al. 2011). All sequences were trimmed to a maximum length of 530 bp to avoid terminal read errors and focus downstream analyses on the V4 region of interest (Niklas et al. 2013). Dereplication into unique individual reads (UIRs) was performed using the VSEARCH script to cluster 100% identical amplicons (Rognes et al. 2016), to identify singletons (read abundance = 1; see Results and Discussion). Singletons are likely artifacts of pyrosequencing (Edgar et al. 2011) and were therefore removed. UIRs were aligned to the PR^2 (v203) database (Guillou et al. 2013) using the nucleotide basic local alignment search tool (BLASTn v. 2.2.31+) algorithm. Default BLASTn parameters (open gap penalty 5, cost gap extension penalty 2, nucleic match 2, nucleic penalty mismatch -3 and word size 11) were applied. A single hit for an UIR was retained if E-value $\leq e^{-100}$. UIRs that clustered to the same accession number in the PR^2 database were clustered to that accession number. One accession number was counted as a single operational taxonomic unit (OTU) and only UIRs with 100% query coverage (100% length of the query sequence matched a part of a full length reference sequence) were considered suitable for further analysis. This included most of the reads (78.5%). Even though UIRs (unique individual reads) may be 100% identical to a reference sequence, ambiguous identification is still possible. Some OTUs (~530 bp) containing UIRs with 100% sequence identity may represent more than one species because the sequence similarity is limited to the barcoding region only. All hits were inspected for accuracy using the metagenome analyzer (MEGAN v. 5) program for conserved sequences (Huson et al. 2007). Using 50 BLASTn hits per UIR sequence, conserved sequences were correctly identified to the high-order taxa in database, due to the lowest common ancestor (LCA) algorithm. On the other hand, although 454 sequencing presents a low error rate (Huse et al. 2007; Niklas et al. 2013), different query sequences might represent the same morphospecies; where genetic delineation is not known yet or genetic diversity is unclear due to the accumulation of "neutral" mutations in a single species (Škaloud and Rindi 2013). To manually inspect and weight the pairwise distances between UIRs and centroid reference sequences within each OTU or taxonomic group, the Kimura 2-parameter (K2P) in MEGA6 (Tamura et al. 2013) was used. Distance values could be multiplied by 100 to obtain the percent pairwise distance between two sequences (e.g. 0.03 = 3% pairwise distance). Furthermore, OTUs with high similarity annotations to (1.) taxa for which the V4 region of the 18S rRNA gene is not a suitable barcoding gene (Pawlowski et al. 2012) and (2.) non-protistan taxa were removed. OTUs with tags for Metazoa and Fungi within the supergroup Opisthokonta, and Streptophyta within the supergroup Archaeplastida (Adl et al. 2012) were removed.

Taxonomic identification: By grouping UIR sequences into OTUs at pairwise identity cut-off values, we were able to identify known diversity from unknown diversity. UIRs with ≥99.7% pairwise identity to a reference sequence identified known taxa and OTUs were proxy for species taxonomic level However, to be able to compare our results with literature data, analyses were also performed on presence-absence data for ≥97% sequence similarity which separated OTUs at least to the right genus taxonomic level. Sequences with ≥80% pairwise identity separated UIRs into the right taxonomic class level OTUs (see Results and Discussion). Reads with more than 20% dissimilarity to any reference sequence in the PR^2 database may have derived from so far unsequenced or even undescribed lineages (Stoeck et al. 2010) and were therefore removed from the analyses. The number of OTUs at ≥80% pairwise identity matching reference sequences from undescribed environmen-

tal samples and lineages with uncertain placement (taxa with suffix: "X" in PR^2 library; Guillou et al. 2013) increased inversely proportional to the sequence similarity cut-offs (Fig. 2A and B). A community consisting mainly of rare taxa that escaped previous research and formal description was revealed; far richer and more diverse than that visible at the ≥99.7% sequence similarity threshold. We assume several sequences may belong to inactive forms from the soil "seedbank" (Finlay 2002) which are included in molecular soil studies (Santos et al. 2015). On the other hand, some OTUs with 100% sequence identity may have been represented by more than one species due to conserved sequences (Huson et al. 2007).

Statistical analysis: Data for annotated UIRs, at class taxonomic level (≥80%), genus or general OTU taxonomic level (≥97%) and species taxonomic level (≥99.7%) were each converted to presence-absence matrixes (binary data) to circumvent the effect of inflated abundance. The number of individual UIRs or OTUs (conservative or database dependent) per sample therefore indicated the species richness. Collective binary data within the above mentioned levels for each exploratory site was then considered as relative abundance data for each taxonomic level and used to equate correlations with LUI. All statistical analyses were conducted in R version 3.0.2 (Core Team R 2013). The following analyses were performed using functions in the 'Vegan' package version 2.3-5 (http://cran.r-project.org/package=vegan; Oksanen et al. 2015). Rank abundance analysis (frequency distribution) was performed using the "fisherfit" function for the Fisher log-series model to ascertain the number of species with *n*-individuals. The extrapolated species pool size (Chao value) to estimate the number of unobserved species was computed using the "specpool" function (Chao 1987; Oksanen et al. 2015). This "specpool" function uses a matrix of the species per site occurrence and is therefore incidence-based on presence-absence data. Species accumulation curves with standard deviation generated from random permutations of the data to study species richness and estimate the number of unseen species by the asymptote was done using the "specaccum" function. Venn diagrams, analysis of variance (ANOVA) and correlations (Spearman) were also performed using the 'Vegan' package and illustrated using the 'gplots' package (http://cran.r-project.org/web/packages/gplots/index.html). The beta diversity was additionally analyzed using the "vegdist" and "cmdscale" functions in the R package 'Vegan', to equate the Bray Curtis and Jaccard dissimilarity indices between community OTUs.

Taxa-area relationship of protists: In order to compare datasets between sites and scales without biasing community compositional comparisons (Terrat et al. 2015), samples were homogenized to include a retained number of UIRs - a value close to the site with the lowest number of UIRs (min = 3UIRs/2OTUs). A maximum of 200 UIRs were resampled by random selection for each of the remaining samples (n = 32 per each Biodiversity Exploratory) containing ≥200 sequences. All remaining high quality reads were moved into one mesoscale file from which the taxa-area relationship was calculated as follows: Pivot tables were constructed for ≥80%, ≥97% and ≥99.7% sequence identity groupings, so to compare community composition at each level of resolution. To avoid up-weighting of rare UIRs in the dataset, the pivot tables were converted to presence/absence data frames (Terrat et al. 2015). The same procedure was applied to compute OTU-area-relationship for comparative purposes. Taxa-area curves were constructed from geographic coordinates (GPS) data and a species by localities dataframe using the "sac" function of the 'fossil' v. 0.3.7 package in R (http://matthewvavrek.com/programs-and-code/fossil/). The data for total area vs. the total species present was log transformed before plotting, approximating a linear transformation. The linear model "lm" was equated to find the incline "*z*". The goodness of fit (R^2 and significance) was equated using the "coefficients" function in R. The relationship of the number of taxa (UIR/OTU) (S) per sampled area (A) was derived from the slope (z) of the taxa-area graph. Sites were then separated into three levels of land-use based on the land-use intensity (LUI) index values associated with each site. Taxa-area relationships were subsequently equated in the same way as for the geographically separated sites, disregarding for geographical separation.

Acknowledgements

We acknowledge bioinformatical and programming help by Johannes Schoeneich. We greatly appreciated helpful criticism and comments by Micah Dunthorn and one anonymous reviewer. The work was supported by grants from the German Research Foundation (DFG) to H.A. (grant number AR 288/16-1,2, SFB 1211, INST 216/862-1), a stipendium of the German Academic Exchange Service (DAAD) to P.V. (grant number 91525927) and an exclusive research topic (SFB 680) project fund at the University of Cologne to P.H.

We thank the managers of the three Exploratories, Swen Renner, Sonja Gockel, Martin Gorke and all former managers for their work in maintaining the plot and project infrastructure; Simone Pfeiffer for giving support through the central office, Jens Nieschulze for managing the central data base, and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project. The work has been partly funded by the DFG Priority Program 1374 "Infrastructure-Biodiversity-Exploratories". Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg (according to § 72 BbgNatSchG).

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.protis.2017.03.005.

References

Adl SM, Simpson AGB, Lane CE, Lukes J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A,

Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EAD, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick L, Schoch CL, Smirnov A, Spiegel FW (2012) The revised classification of eukaryotes. J Eukaryot Microbiol **59**:429–493

Allan E, Bossdorf O, Dormann CF, Prati D, Gossner MM, Tscharntke T, Blüthgen N, Bellach M, Birkhofer K, Boch S, Böhm S, Börschig C, Chatzinotas A, Christ S, Daniel R, Diekötter T, Fischer C, Friedl T, Glaser K, Hallmann C, Hodac L, Hölzel N, Jung K, Klein AM, Klaus VH, Kleinebecker T, Krauss J, Lange M, Morris EK, Müller J, Nacke H, Pašalić E, Rillig MC, Rothenwöhrer C, Schall P, Scherber C, Schulze W, Socher SA, Steckel J, Steffan-Dewenter I, Türke M, Weiner CN, Werner M, Westphal C, Wolters V, Wubet T, Gockel S, Gorke M, Hemp A, Renner SC, Schöning I, Pfeiffer S, König-Ries B, Buscot F, Linsenmair KE, Schulze E-D, Weisser WW, Fischer M (2014) Interannual variation in land-use intensity enhances grassland multidiversity. Proc Natl Acad Sci USA **111**:308–313

Baker GC, Beebee TJ, Ragan MA (1999) *Prototheca richardsi*, a pathogen of anuran larvae, is related to a clade of protistan parasites near the animal-fungal divergence. Microbiology **145(Pt 7)**:1777–1784

Barreto DP, Conrad R, Klose M, Claus P, Enrich-Prast A (2014) Distance-decay and taxa-area relationships for bacteria, archaea and methanogenic archaea in a tropical lake sediment. PLoS ONE **9**:e110128

Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R, Fierer N (2013) Global biogeography of highly diverse protistan communities in soil. ISME J **7**:652–659

Beijerinck MW (1898) Ueber ein *Contagium vivum fluidum* als Ursache der Fleckenkrankheit der Tabaksblätter. Verh Kon Akad Wetensch **VI**:3–21

Berney C, Pawlowski J (2006) A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proc Biol Sci **273**:1867–1872

Bluethgen N, Dormann CF, Prati D, Klaus VH, Kleinebecker T, Hölzel N, Alt F, Boch S, Gockel S, Hemp A, Müller J, Nieschulze J, Renner SC, Schöning I, Schumacher U, Socher SA, Wells K, Birkhofer K, Buscot F, Oelmann Y, Rothenwöhrer C, Scherber C, Tscharntke T, Weiner CN, Fischer M, Kalko EKV, Linsenmair KE, Schulze E-D, Weisser WW (2012) A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. Basic Appl Ecol **13**:207–220

Brabender M, Kiss AK, Domonell A, Nitsche F, Arndt H (2012) Phylogenetic and morphological diversity of novel soil cercomonad species with a description of two new genera (*Nucleocercomonas* and *Metabolomonas*). Protist **163**:495–528

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD, Dennett MR, Moran DM, Jones AC (2009) Defining DNA-based operational taxonomic units for microbial-eukaryotic ecology. Appl Environ Microbiol **75**:5797–5808

Chao A (1987) Estimating the population size for capturerecapture data with unequal catchability. Biometrics **43**:783–791

Core Team R (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna

de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J-M, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horák A, Jaillon O, Lima-Mendez G, Lukeš J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans Coordinators: Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iodicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E (2015) Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. Science **348**:1261605

Dopheide A, Lear G, Stott R, Lewis G (2008) Molecular characterization of ciliate diversity in stream biofilms. Appl Environ Microbiol **74**:1740–1747

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**:2194–2200

Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, Leslin C, Holder M, Taylor GT, Suarez P, Varela R, Epstein S (2011) Protistan microbial observatory in the Cariaco basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. ISME J **5**:1344–1356

Fan Y, Hu X, Gao F, Al-Farraj SA, Al-Rasheid KA (2014) Morphology, ontogenetic features and SSU rRNA gene-based phylogeny of a soil ciliate, Bistichella cystiformans spec. nov. (Protista, Ciliophora, Stichotrichia). Int J Syst Evol Microbiol **64(Pt 12)**:4049–4060

Feldmann S (2007) Studies on the occurrence of heterotrophic nanofauna in the air, MSc. Thesis. Institute for Zoology, University of Cologne

Fenchel T, Finlay BJ (2006) The diversity of microbes: resurgence of the phenotype. Philos Trans R Soc B **361**:1965–1973

Findenig BM, Chatzinotas A, Boenigk J (2010) Taxonomic and ecological characterization of stomatocysts of *Spumella*-like flagellates (Chrysophyceae). J Phycol **46**:868–881

Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. Science **296**:1061–1063

Fischer M, Bossdorf O, Gockel S, Hänsel F, Hemp A, Hessenmöller D, Korte G, Nieschulze J, Pfeiffer S, Pratia D, Renner S, Schöning I, Schumacher U, Wells K, Buscot F, Kalko EKV, Linsenmair KE, Schulze E-D, Weisser WW (2010) Implementing large-scale and long-term functional biodiversity research: The biodiversity exploratories. Basic Appl Ecol **11**:473–485

Foissner W (2006) Biogeography and dispersal of microorganisms: a review emphasizing protists. Acta Protozool **45**:111–136

Foissner W (2008) Protist diversity and distribution: some basic considerations. Biodivers Conserv **17**:235–242

Forster D, Dunthorn M, Mahé F, Dolan JR, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie JM, Decelle J, Edvardsen B, Egge E, Eikrem W, Gobet A, Kooistra WH, Logares R, Massana R, Montresor M, Not F, Ogata H, Pawlowski J, Pernice MC, Romac S, Shalchian-Tabrizi K, Simon N, Richards TA, Santini S, Sarno D, Siano R, Vaulot D, Wincker P, Zingone A, de Vargas C, Stoeck T (2016) Ben-

thic protists: the under-charted majority. FEMS Microbiol Ecol **92(8)**:fiw120

Geisen S, Laros I, Vizcaíno A, Bonkowski M, de Groot GA (2015) Not all are free-living: high-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. Mol Ecol **24**:4556–4569

Gossner MM, Lewinsohn T, Kahl T, Grassein F, Boch S, Prati D, Birkhofer K, Renner SC, Sikorski J, Arndt H, Baumgartner V, Blaser S, Blüthgen N, Börschig C, Buscot F, Diekötter T, Jorge LR, Jung K, Keyel AC, Klein A-M, Klemmer S, Krauss J, Lange M, Müller J, Overmann J, Pašalić E, Penone C, Perović D, Purschke O, Schall P, Socher SA, Sonnemann I, Tschapka M, Tscharntke T, Türke M, Venter PC, Weiner CN, Werner M, Wolters V, Wurst S, Westphal C, Wubet T, Fischer M, Weisser WW, Allan E (2016) Land-use intensification causes multitrophic homogenization of grassland communities. Nature **540**:266–269

Green J, Bohannan BJM (2006) Spatial scaling of microbial biodiversity. Trends Ecol Evol **21**:501–507

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, Del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WH, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet AL, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R (2013) The Protist ribosomal reference database (PR^2): a catalog of unicellular eukaryote small subunit rRNA sequences with curated taxonomy. Nucleic Acids Res **41**:D597–D604

Hadziavdic K, Lekanq K, Lanzen A, Jonassen I, Thompson EM, Troedsson C (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. PLoS ONE **9**:e87624

Hansen G, Botes L, De Salas M (2007) Ultrastructure and large subunit rDNA sequences of *Lepidodinium viride* reveal a close relationship to *Lepidodinium chlorophorum* comb. nov. (=*Gymnodinium chlorophorum*). Phycol Res **55**:25–41

Horner-Devine MC, Lage M, Hughes JB, Bohannan BJ (2004) A taxa-area relationship for bacteria. Nature **432**:750–753

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ Microbiol **12**:1889–1898

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol **8**:R143

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res **17**:377–386

Kanyuka K, Ward E, Adams MJ (2003) *Polymyxa graminis* and the cereal viruses it transmits: a research challenge. Mol Plant Pathol **4**:393–406

Koeppel AF, Wu M (2013) Surprisingly extensive mixed phylogenetic and ecological signals among bacterial operational taxonomic units. Nucleic Acids Res **41**:5175–5188

Lara E, Mitchell EA, Moreira D, López-Gracía P (2011) Highly diverse and seasonally dynamic protist community in a pristine peat bog. Protist **162**:14–32

Lejzerowicz F, Pawlowski J, Fraissinet-Tachet L, Marmeisse R (2010) Molecular evidence for widespread occurrence of Foraminifera in soils. Environ Microbiol **12**:2518–2526

Lentendu G, Wubet T, Chatzinotas A, Wilhelm C, Buscot F, Schlegel M (2014) Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. Mol Ecol **23**:3341–3355

Lesaulnier C, Papamichail D, McCorkle S, Ollivier B, Skiena S, Taghavi S, Zak D, van der Lelie D (2008) Elevated atmospheric CO2 affects soil microbial diversity associated with trembling aspen. Environ Microbiol **10**:926–941

Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F, Bittner L, Darzi Y, Wang J, Audic S, Berline L, Bontempi G, Cabello AM, Coppola L, Cornejo-Castillo FM, d'Ovidio F, De Meester L, Ferrera I, Garet-Delmas M-J, Guidi L, Lara E, Pesant S, Royo-Llonch M, Salazar G, Sánchez P, Sebastian M, Souffreau C, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans coordinators: Gorsky G, Not F, Ogata H, Speich S, Stemmann L, Weissenbach J, Wincker P, Acinas SG, Sunagawa S, Bork P, Sullivan MB, Karsenti E, Bowler C, de Vargas C, Raes J (2015) Ocean plankton. Determinants of community structure in the global plankton interactome. Science **348**:e1262073

Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ **3**:e1420

Mahé F, de Vargas C, Bass D, Czech L, Stamatakis A, Lara E, Singer D, Mayor J, Bunge J, Sernaker S, Siemensmeyer T, Trautmann I, Romac S, Berney C, Kozlov A, Mitchell EAD, Seppey CVW, Egge E, Lentendu G, Wirth R, Trueba G, Dunthorn M (2017) Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. Nature Ecology and Evolution **1**:0091

Massana R, Pernice M, Bunge JA, del Campo J (2011) Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean. ISME J **5**:184–195

Massana R, del Campo J, Sieracki ME, Audic S, Logares R (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. ISME J **8**:854–866

Mueller JA, Mueller WP (1970) *Colpoda cucullus*: a terrestrial aquatic. Am Midl Nat **84**:1–12

Nakamura K, Suto M (2012) Analysis of bacterivorous protozoa in the natural environment. J JSCE, Ser. G (Env Res) **68**:III_31–III_40

Nebel M, Pfabel C, Stock A, Dunthorn M, Stoeck T (2011) Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. Environ Microbiol Rep **3**:154–158

Niklas N, Pröll J, Danzer M, Stabentheiner S, Hofer K, Gabriel C (2013) Routine performance and errors of 454 HLA exon sequencing in diagnostics. MBC Bioinformatics **14**:176

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2015) Community ecology package version 2.3-1, http://cran.r-project.org, https://github.com/vegandevs/vegan [accessed 2016 Mar 29]

Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M, Fiore-Donno AM, Gile GH, Holzmann M, Jahn R, Jirků M, Keeling PJ, Kostka M, Kudryavtsev A, Lara E, Lukeš J, Mann DG, Mitchell EAD, Nitsche F, Romeralo M, Saunders GW, Simpson AGB, Smirnov AV, Spouge JL, Stern RF, Stoeck T, Zimmermann J, Schindel D, de Vargas C (2012) CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol **10**:e1001419

Pawlowski J, Lecroq B (2010) Short rDNA barcodes for species identification in foraminifera. J Eukaryot Microbiol **57**:197–205

Preheim SP, Perrotta AR, Martin-Platero AM, Gupta A, Alm EJ (2013) Distribution-based clustering: using ecology to refine the operational taxonomic unit. Appl Environ Microbiol **79**:6593–6603

Robideau GP, De Cock AW, Coffey MD, Voglmayr H, Brouwer H, Bala K, Chitty DW, Désaulniers N, Eggertson QA, Gachon CMM, Hu C-H, Küpper FC, Rintoul TL, Sarhan E, Verstappen ECP, Zhang Y, Bonants PJM, Ristaino JB, Lévesque CA (2011) DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. Mol Ecol Resour **11**:1002–1011

Rogerson A, Detwiler A (1999) Abundance of airborne heterotrophic protists in ground level air of South Dakota. Atmos Res **51**:35–44

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ **4**:e2584

Santos SS, Nielsen TK, Hansen LH, Winding A (2015) Comparison of three DNA extraction methods for recovery of soil protist DNA. J Microbiol Methods **115**:13–19

Schloss PD, Westcott SL (2011) Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol **77**:3219–3226

Simon M, López-Gracía P, Deschamps P, Moreira D, Restoux G, Bertolino P, Jardillier L (2015) Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. ISME J **9**:1941–1953

Škaloud P, Rindi F (2013) Ecological differentiation of cryptic species within an asexual protist morphospecies: a case study of filamentous green alga *Klebsormidium* (Streptophyta). J Eukaryot Microbiol **60**:350–362

Šlapeta J, Moreira D, López-García P (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. Proc Biol Sci **272**:2073–2081

Soliveres S, Manning P, Prati D, Grossner MM, Alt F, Arndt H, Baumgartner V, Binkenstein J, Birkhofer K, Blaser S, Blüthgen N, Boch S, Böhm S, Börschig C, Buscot F, Diekötter T, Heinze J, Hölzel N, Jung K, Klaus VH, Klein A-M, Kleinebecker T, Klemmer S, Krauss J, Lange M, Morris EK, Müller J, Oelmann Y, Overmann J, Pašalić E, Renner SC, Rillig MC, Schaefer HM, Schloter M, Schmitt B, Schöning I, Schrumpf M, Sikorski J, Socher SA, Solly EF, Sonnemann I, Sorkau E, Steckel J, Steffan-Dewenter I, Stempfhuber B, Tschapka M, Türke M, Venter P, Weiner CN, Weisser WW, Werner M, Westphal C, Wilcke W, Wolters V, Wubet T, Wurst S, Fischer M, Allan E (2016a) Locally rare species influence grassland ecosystem multifunctionality. Phil Trans R Soc B **371**:20150269

Soliveres S, Van der Plas F, Manning P, Prati D, Grossner MM, Renner SC, Alt F, Arndt H, Baumgartner V, Binkenstein J, Birkhofer K, Blaser S, Blüthgen N, Boch S, Böhm S, Börschig C, Buscot F, Diekötter T, Heinze J, Hölzel N, Jung K, Klaus VH, Kleinebecker T, Klemmer S, Krauss J, Lange M, Morris EK, Müller J, Oelmann Y, Overmann J, Pašalić E, Rillig MC, Schaefer HM, Schloter M, Schmitt B, Schöning I, Schrumpf M, Sikorski J, Socher SA, Solly EF, Sonnemann I, Sorkau E, Steckel J, Steffan-Dewenter I, Stempfhuber B, Tschapka M, Türke M, Venter PC, Weiner CN, Weisser WW, Werner M, Westphal C, Wilcke W, Wolters V, Wubet T, Wurst S, Fischer M, Allan E (2016b) Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. Nature **536**:456–459

Stoeck T, Bass D, Nebel M, Christian R, Jones MDM, Breiner H-W, Richards TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol **19**:21–31

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans coordinators: Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P (2015) Ocean plankton. Structure and function of the global ocean microbiome. Science **348**:e1261359

Taib N, Mangot JF, Domaizon I, Bronner G, Debroas D (2013) Phylogenetic affiliation of SSU rRNA genes generated by massively parallel sequencing: new insights into the freshwater protist diversity. PLoS ONE **8**:e58950

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6. 0. Mol Biol Evol **30**:2725–2729

Tedersoo L, Nilsson RH, Aberankov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G, Kõljalg U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. New Phytol **188**:291–301

Terrat S, Dequiedt S, Horrigue W, Lelievre M, Cruaud C, Saby NPA, Jolivet C, Arrouays D, Maron P-A, Ranjard L, Prévost-Bouré NC (2015) Improving soil bacterial taxa-area relationships assessment using DNA meta-barcoding. Heredity (Edinb.) **114**:468–475

Van Dorst J, Bissett A, Palmer AS, Brown M, Snape I, Stark JS, Raymond B, McKinlay J, Ji M, Winsley T, Ferrari BC (2014) Community fingerprinting in a sequencing world. FEMS Microbiol Ecol **89**:316–330

Wang Y, Zhang WP, Cao HL, Shek CS, Tian RM, Wong YH, Batang Z, Al-Suwailem A, Qian P-Y (2014) Diversity and distribution of eukaryotic microbes in and around a brine pool adjacent to the Thuwal cold seeps in the Red Sea. Front Microbiol **5**:1–37

Weber AA, Pawlowski J (2013) Can abundance of protists be inferred from sequence data: a case study of foraminifera. PLoS ONE **8**:e56739

Westcott SL, Schloss PD (2015) *De novo* clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. PeerJ **3**:e1487

Whittaker RH (1960) Vegetation of the Siskiyou Mountains, Oregon and California. Ecol Monogr **30**:279–338

Woodcock S, Curtis TP, Head IM, Lunn M, Sloan WT (2006) Taxa-area relationships for microbes: the unsampled and the unseen. Ecol Lett **9**:805–812

Supplementary Data

The protistan microbiome of grassland soil: distribution in the mesoscale

by Paul Christiaan Venter, Frank Nitsche, Anne Domonell, Peter Heger and Hartmut Arndt.

Table S1. Breakdown of SSU sequencing data for 150 samples. Steps follow the pipeline given in Fig. S1. The Chao richness estimate indicates the statistical pool size of operational taxonomic units (OTUs) and unique individual reads (UIRs), including standard error (SE).

	OTUs	Obtained sequences
Total number of sequences		1 249 645
Sequences after filtering		1 101 087
Unique (dereplicated)		479 108
Singletons removed		58 216
Sequences with 100% query coverage	2 764	45 686
- Fungi		23 805
- Metazoa		6 045
- Streptophyta		6 777
Unique protist only sequences		9 059
Assigned protist reads with ≥80% pairwise identity	968	8 407 UIRs
Chao's richness estimate (±SE)	978±4	8 560±16 UIRs
Assigned protist reads with ≥97% pairwise identity	695	6 401 UIRs
Chao's richness estimate (±SE)	699±3	6 485±11 UIRs
Assigned protist reads with ≥99.7% pairwise identity	298	831 UIRs
Chao's richness estimate (±SE)	300±2	835±2 UIRs

SE = Standard error (SE), UIRs = unique individual reads, OTUs = operational taxonomic units (OTUs).

Table S2. Summary of the number of operational taxonomic units (OTUs), unique individual reads (UIRs) and sequence abundance within each class taxonomic group derived from all 150 soil samples. Data derived from ≥97% reference sequence identity to reference sequences in the protist ribosomal reference (PR^2) database.

Supergroup data	Total OTUs	Schwaebische Alb (AEG)			Hainich-Duen (HEG)			Schorfheide-Chorin (SEG)		
		OTUs	UIRs	Sequences	OTUs	UIRs	Sequences	OTUs	UIRs	Sequences
Alveolata	196	138	1 195	5 306	168	1 543	8 582	170	1 654	8 234
Amoebozoa	21	14	69	249	15	93	643	20	101	434
Apusozoa	3	3	16	36	3	24	85	3	39	174
Archaeplastida	88	55	128	366	78	237	1 037	59	151	355
Hacrobia	4	1	2	3	1	2	3	3	4	13
Opisthokonta	11	8	43	205	9	26	72	8	33	88
Rhizaria	245	191	988	2 871	190	1 099	3 847	224	1 390	5 135
Stramenopiles	127	96	700	4 005	92	564	3 835	106	768	4 255
Total count	695	506	3 141	13 041	556	3 588	18 104	593	4 140	18 688

Environmental metadata	AEG	HEG	SEG
Latitude / longitude	54° 48.5'N / 35° 9.2'W	57°51.4'N / 44° 10.2'N	59° 53.2'N / 46° 13.4'W
Habitat type	Grassland	Grassland	Grassland
Sampling area size (km^2)	423	1 561	1 300
Mean annual temperature (°C)	6.5 - 8.0	6.5 - 8.0	8.0 - 8.4
Mean annual precipitation (mm)	938 - 963	500 - 800	520 - 600
Mean elevation (m)	462 - 858	285 - 550	10 - 140

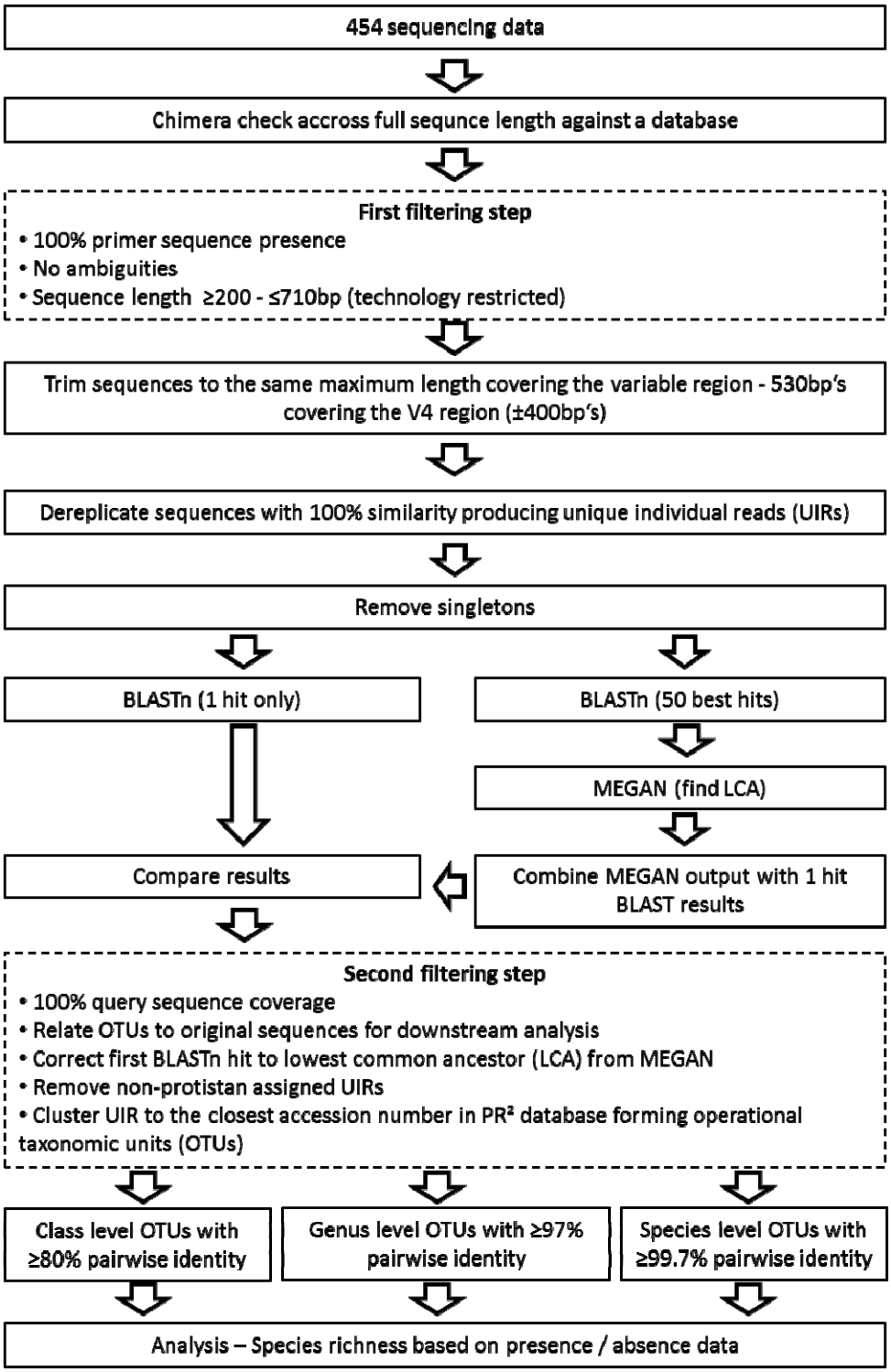


Figure S1. Bio-informatical pipeline used to convert next-generation sequencing raw data into unique individual reads (UIRs) and then functional operational taxonomic units (OTUs) for downstream analysis. Processing steps follow chronologically, relating OTUs to raw query sequences (UIRs) at all steps. Confidence of UIR assignments by basic local alignment search tool for nucleotide sequences (BLASTn) is dependent on the levels of pairwise identity, where UIRs are grouped into OTUs. The lowest common ancestor (LCA) principle in MEGAN (Huson et al. 2007) was employed to check the BLASTn hit accuracy.

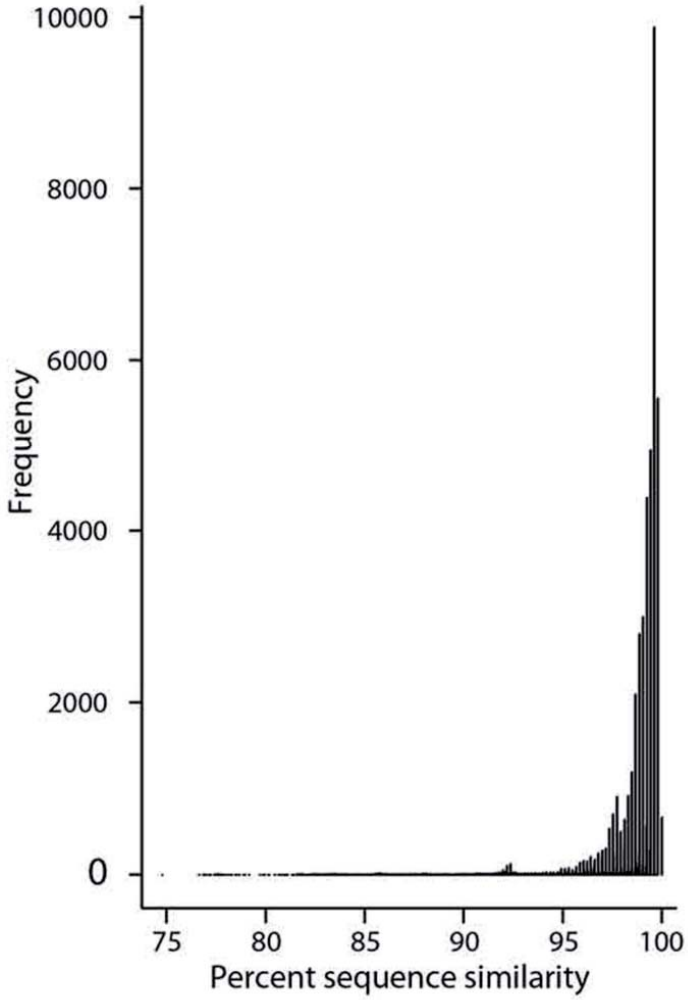


Figure S2. Similarity of raw reads to the curated Protist Ribosomal Reference Database (PR^2) database (Guillou et al, 2013) after Blast annotation of UIRs. Only 11 protist affiliated V4 SSU OTUs (UIRs associated with protist reference sequences) from 150 soil samples fell below the 80% sequence pairwise similarity cut-off for inclusion into the analyses.

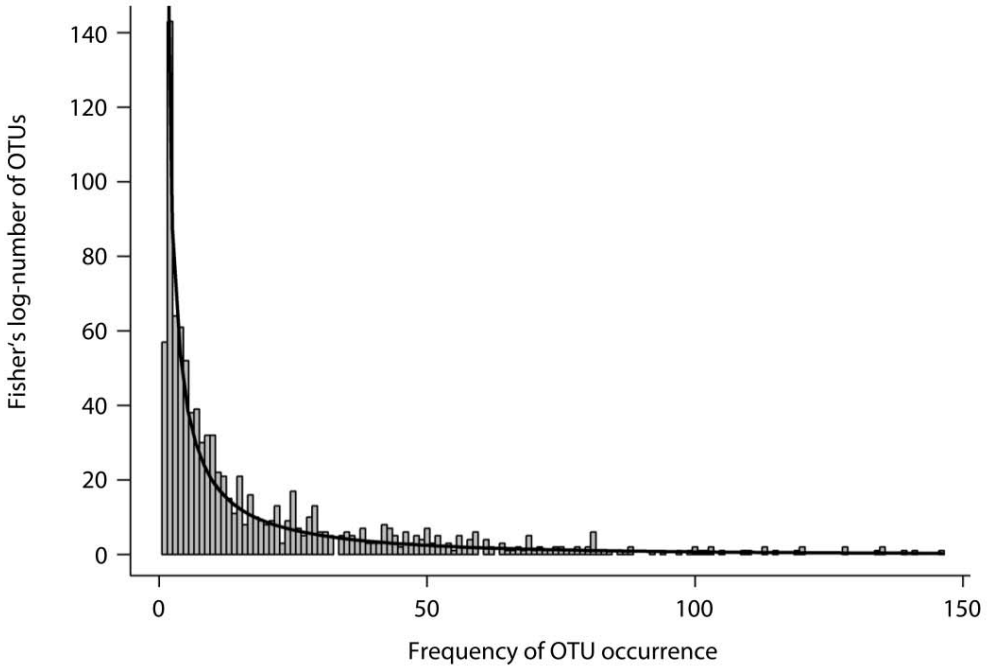


Figure S3. Fisher's log-series frequency distribution graph (Oksanen et al. 2017) with slope line fitted for operational taxonomic units (OTUs) from all 150 samples. Incidence per frequency of OTU occurrence (max=146) are indicated for all OTUs with sequence identities of ≥80% to reference sequences in the PR^2 database. Singletons (OTUs that occurred only once in the data) were removed prior to this analysis and data was binary transformed.

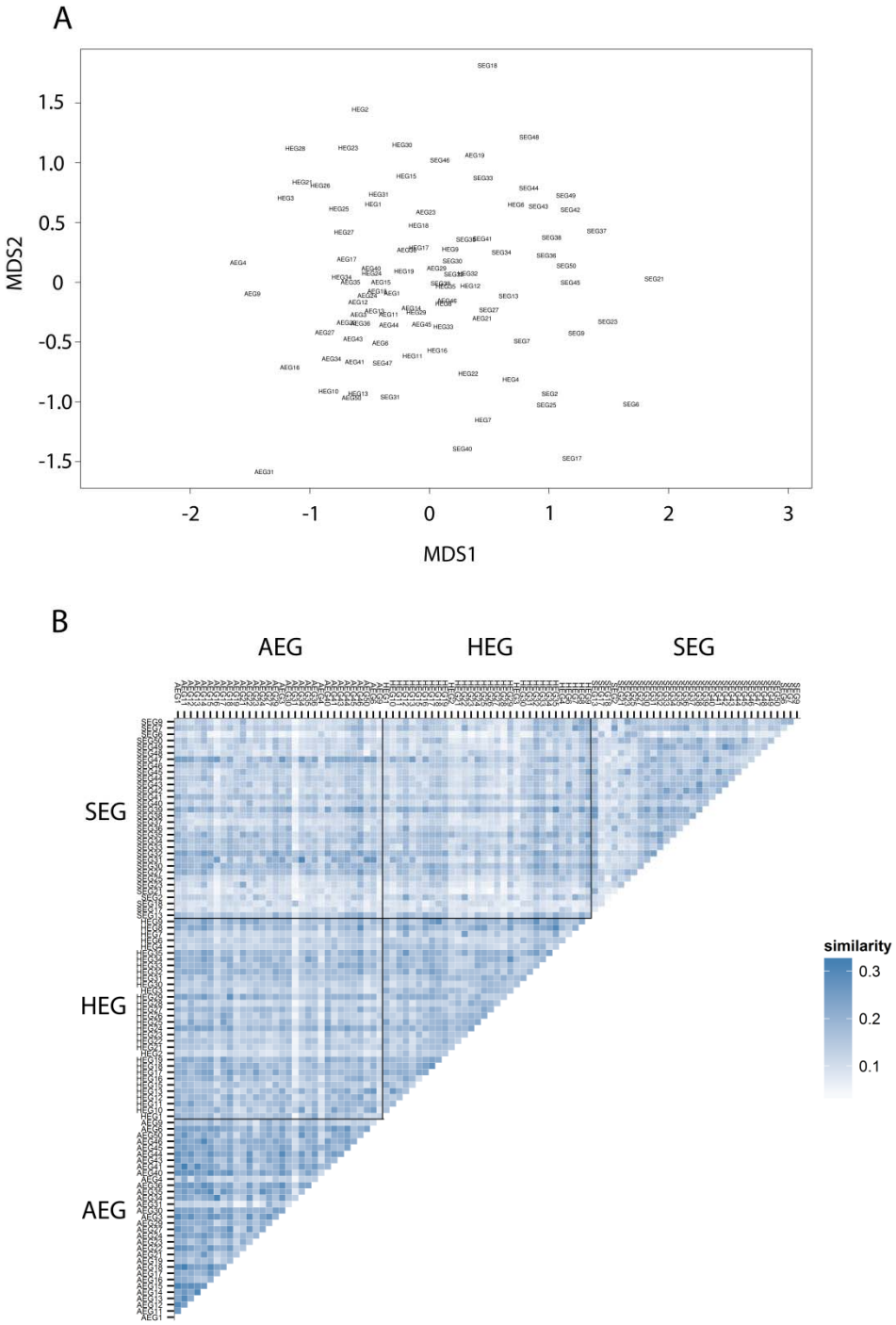


Figure S4. OTU compositional differences between soil samples from the three Biodiversity Exploratory plots Schorfheide Chorin (SEG), Hainich Dün (HEG) and Schwäbische Alb (AEG). (A.) Non-metric multidimentional scaling (MDS; Bray-Curtis method): grouping samples according to taxonomic compositional similarity, and (B.) Jaccard similarity index indicate a low value when OTU turnover between samples is high.

Chapter 2 – Land-use intensification causes multitrophic homogenization of grassland communities

LETTER

doi:10.1038/nature20575

Land-use intensification causes multitrophic homogenization of grassland communities

Martin M. Gossner[1,2,3], Thomas M. Lewinsohn[1,4], Tiemo Kahl[5,6], Fabrice Grassein[7], Steffen Boch[7], Daniel Prati[7], Klaus Birkhofer[8,9], Swen C. Renner[10,11], Johannes Sikorski[12], Tesfaye Wubet[13,14], Hartmut Arndt[15], Vanessa Baumgartner[12], Stefan Blaser[7], Nico Blüthgen[16], Carmen Börschig[17], Francois Buscot[13,14], Tim Diekötter[18,19], Leonardo Ré Jorge[4], Kirsten Jung[11], Alexander C. Keyel[20], Alexandra-Maria Klein[21], Sandra Klemmer[13,22], Jochen Krauss[17], Markus Lange[1,2,23], Jörg Müller[24], Jörg Overmann[12], Esther Pašalić[1,2], Caterina Penone[7], David J. Perović[25,26], Oliver Purschke[22,27,28], Peter Schall[29], Stephanie A. Socher[30], Ilja Sonnemann[31], Marco Tschapka[11], Teja Tscharntke[26], Manfred Türke[1,2,14,22], Paul Christiaan Venter[15], Christiane N. Weiner[16], Michael Werner[16], Volkmar Wolters[18], Susanne Wurst[31], Catrin Westphal[26], Markus Fischer[7], Wolfgang W. Weisser[1,2] & Eric Allan[7,32]

Land-use intensification is a major driver of biodiversity loss[1,2]. Alongside reductions in local species diversity, biotic homogenization at larger spatial scales is of great concern for conservation. Biotic homogenization means a decrease in β-diversity (the compositional dissimilarity between sites). Most studies have investigated losses in local (α)-diversity[1,3] and neglected biodiversity loss at larger spatial scales. Studies addressing β-diversity have focused on single or a few organism groups (for example, ref. 4), and it is thus unknown whether land-use intensification homogenizes communities at different trophic levels, above- and belowground. Here we show that even moderate increases in local land-use intensity (LUI) cause biotic homogenization across microbial, plant and animal groups, both above- and belowground, and that this is largely independent of changes in α-diversity. We analysed a unique grassland biodiversity dataset, with abundances of more than 4,000 species belonging to 12 trophic groups. LUI, and, in particular, high mowing intensity, had consistent effects on β-diversity across groups, causing a homogenization of soil microbial, fungal pathogen, plant and arthropod communities. These effects were nonlinear and the strongest declines in β-diversity occurred in the transition from extensively managed to intermediate intensity grassland. LUI tended to reduce local α-diversity in aboveground groups, whereas the α-diversity increased in belowground groups. Correlations between the β-diversity of different groups, particularly between plants and their consumers, became weaker at high LUI. This suggests a loss of specialist species and is further evidence for biotic homogenization. The consistently negative effects of LUI on landscape-scale biodiversity underscore the high value of extensively managed grasslands for conserving multitrophic biodiversity and ecosystem service provision. Indeed, biotic homogenization rather than local diversity loss could prove to be the most substantial consequence of land-use intensification.

Land-use intensification threatens biodiversity[2,4] by reducing the α-diversity of many taxa[1,3]. Similarly, β-diversity[5] may decline strongly. This biotic homogenization[6–9] might occur through either a loss of rare or specialized species (reducing differences between communities), a gain of widespread, generalist species in intensively managed systems (increasing similarity), or most likely a combination of both. Most studies have investigated loss of species richness, but global change may have larger effects on community composition than on local diversity[10,11]. To separate biotic homogenization from loss of species richness requires measures of β-diversity that distinguish pure species turnover from changes in α-diversity[5]. To predict and manage the loss of β-diversity, we also need to understand whether biotic homogenization occurs at a constant rate as land use intensifies. Land-use intensification can affect α-diversity nonlinearly[1,12], but, although environmental gradients can have nonlinear effects on β-diversity[13,14], no such effects of land use have been investigated. Here we use data from several landscapes and regions to analyse land-use effects on species turnover (β-turnover)[15,16] and on total β-diversity (also including differences in species richness) across a wide range of trophic groups.

Different types of organisms probably respond differently to land use. In grasslands, α-diversity belowground may be less affected than aboveground[1]. However, land-use intensification may homogenize species composition belowground and reduce β-diversity without reducing

[1]Terrestrial Ecology Research Group, Department of Ecology and Ecosystem Management, School of Life Sciences Weihenstephan, Technical University of Munich, Hans-Carl-von-Carlowitz-Platz 2, Freising D-85354, Germany. [2]Institute of Ecology, Friedrich-Schiller-University Jena, Dornburger Str 159, Jena D-07743, Germany. [3]Swiss Federal Research Institute WSL, Birmensdorf CH-8903, Switzerland. [4]Department of Animal Biology, IB, UNICAMP—University of Campinas, Campinas, Sao Paulo, CEP, 13083-970, Brazil. [5]Chair of Silviculture, Faculty of Environment and Natural Resources, University of Freiburg, Tennenbacherstraße 4, Freiburg im Breisgau D-79106, Germany. [6]Biosphere Reserve Vessertal-Thuringian Forest, Brunnenstr 1, Schmiedefeld am Rennsteig D-98711, Germany. [7]Institute of Plant Sciences, University of Bern, Altenbergrain 21, Bern CH-3013, Switzerland. [8]Department of Biology, Biodiversity and Conservation Science, Lund University, Sölvegatan 37, Lund S-22362, Sweden. [9]Chair of Ecology, Faculty Environment and Natural Sciences, BTU Cottbus-Senftenberg, Großenhainer Str 57, Senftenberg D-01968, Germany. [10]Institute of Zoology, University of Natural Resources and Life Sciences, Wien A-1180, Austria. [11]Institute of Evolutionary Ecology and Conservation Genomics, University of Ulm, Ulm D-89069, Germany. [12]Leibniz-Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7B, Braunschweig D-38302, Germany. [13]Department of Soil Ecology, UFZ-Helmholtz Centre for Environmental Research, Halle-Saale D-06120, Germany. [14]Institute of Biology, Leipzig University, Johannisallee 21, Leipzig D-04103, Germany. [15]Biocentre, Institute for Zoology, General Ecology, University of Cologne, Zuelpicher Str 47b, Cologne (Köln) D-50674, Germany. [16]Department of Biology, Ecological Networks, Technische Universität Darmstadt, Schnittspahnstraße 3, Darmstadt D-64287, Germany. [17]Department of Animal Ecology and Tropical Biology, Biocentre, University of Würzburg, Am Hubland, Würzburg D-97074, Germany. [18]Animal Ecology, Justus-Liebig-University, Heinrich-Buff-Ring 26-32, Giessen D-35392, Germany. [19]Landscape Ecology, Institute for Natural Resource Conservation, Kiel University, Olshausenstr 75, Kiel D-24118, Germany. [20]Department of Ecosystem Modelling, University of Göttingen, Büsgenweg 4, Göttingen D-37077, Germany. [21]Chair of Nature Conservation and Landscape Ecology, Faculty of Environment and Natural Resources, University of Freiburg, Tennenbacherstraße 4, Freiburg im Breisgau D-79106, Germany. [22]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, Leipzig D-04103, Germany. [23]Max Planck Institute for Biogeochemistry, Hans-Knoell-Str 10, Jena D-07745, Germany. [24]Institute of Biochemistry and Biology, University of Potsdam, Maulbeerallee 1, Potsdam D-14469, Germany. [25]Institute of Applied Ecology, Fujian Agriculture and Forestry University, Fuzhou, China. [26]Agroecology, Department of Crop Sciences, Georg-August-University Göttingen, Göttingen D-37077, Germany. [27]Department of Computer Science, Martin Luther University, Halle-Wittenberg, Halle (Saale) D-06120, Germany. [28]Geobotany and Botanical Garden, Institute of Biology, Martin Luther University, Halle-Wittenberg, Halle (Saale) D-06108, Germany. [29]Department Silviculture and Forest Ecology of the Temperate Zones, University of Göttingen, Göttingen D-37077, Germany. [30]Department of Ecology and Evolution, Botanical Garden, University of Salzburg, Hellbrunnerstrasse 34, Salzburg 5020, Austria. [31]Functional Biodiversity, Institute of Biology, Freie Universität Berlin, Königin-Luise-Str. 1-3, Berlin D-14195, Germany. [32]Centre for Development and Environment, University of Bern, Hallerstrasse, 10, Bern CH-3012, Switzerland.

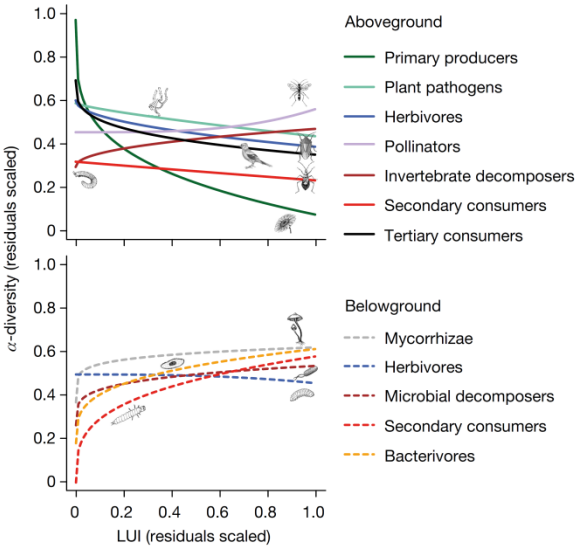


Figure 1 | The effect of LUI on α-diversity above- and belowground. The partial effect of local LUI was generated with a power law model fitted to the species richness of the seven aboveground (solid lines) and the five belowground trophic groups (dashed lines) ($n = 105$ plots; for details, see Methods). Species richness and LUI were corrected for differences due to region, pH, soil nutrients, sdLUI (standard deviation of LUI across five years) and isolation, by taking residuals, and were then scaled between 0 and 1.

α-diversity, as shown for tropical soil bacteria[17]. Biotic homogenization might therefore be widespread, but at present we lack a broader view of how land-use intensification alters α- and β-diversity of distinct functional or trophic groups. The loss of specialist species, which generally accompanies biotic homogenization, could also reduce correlations between the β-diversities of different groups[18], indicating changes in trophic structure.

We compiled a unique set of biodiversity data, including more than 4,000 species of plants, arthropods, foliar fungal pathogens, mycorrhizal fungi, bacteria, protists, bats and birds, measured on 105 grasslands varying in local LUI (a compound index of grazing, mowing and fertilization intensity)[19]. We divided the species into 12 groups on the basis of trophic level and whether they lived above- or belowground. We modelled the effect of LUI on α- and β-diversity of each group, accounting for potential environmental and spatial effects. We assessed biotic homogenization in three ways, by testing: (1) for an overall negative effect of land use on β-diversity; (2) where along the land-use gradient the greatest change in β-diversity occurred; and (3) whether correlations among β-diversity of different trophic groups were reduced, which would indicate a loss of specialist species.

High LUI reduced α-diversity for most aboveground groups, but had neutral or positive effects on belowground organisms (Fig. 1). These results were consistent regardless of the weight given to common species, that is, whether α-diversity was measured as species richness, Shannon or Simpson diversity (Extended Data Fig. 1). Land-use effects were not driven by co-varying environmental factors because we adjusted for soil pH, nutrients and geography (see Methods). Differences between above- and belowground communities may occur because they respond at different spatial scales[20] and belowground groups are better protected from disturbance[21]. Alternatively, a shift towards bacterial-dominated communities in more intensively managed grasslands[22] may have cascaded up to increase the diversity of higher trophic levels. Nevertheless, these opposing above- and belowground responses were unanticipated and have not previously been shown in a multitrophic dataset.

We then analysed the effects of land-use on β-diversity and found widespread evidence for biotic homogenization both above and below ground. We modelled β-diversity between all possible plot pairs using linear models, again correcting for other environmental and geographic drivers, and including two descriptors of LUI: the mean and the difference in LUI (ΔLUI) between them. The mean LUI represents overall intensity; any negative effects on β-diversity indicate biotic homogenization. The linear ΔLUI term represents land-use heterogeneity and positive effects indicate that mixing grasslands of low and high LUI increases β-diversity. Increasing land-use intensity (mean LUI) had strong negative effects on the β-turnover of many above- (4 out of 7) and belowground (2 out of 5) groups (Fig. 2a, Extended Data Figs 2 and 3), indicating biotic homogenization both above- and belowground, in contrast to the opposing responses of α-diversity. Belowground groups, especially mycorrhizae and bacterivores, were therefore homogenized at high LUI even though their α-diversity increased. These stronger effects of LUI on belowground β-turnover extend findings from Amazonian bacterial communities, responding to marked changes in land use[17], to a much larger number of groups. For many groups, increasing LUI had an even larger effect on total β-diversity, which includes changes in species richness and turnover (Extended Data Fig. 4). This was particularly evident for plants, which suffered substantial species loss. In general, ΔLUI effects were smaller than mean LUI effects, showing that increasing land-use heterogeneity has limited potential to offset negative effects of intensification. Land-use heterogeneity might be even less beneficial in cases where high LUI grasslands are dominated by exotic species of low conservation value. Our multitrophic results suggest that, despite their differences in dispersal rates and body sizes[23,24], large-scale spatial dynamics are similar in below- and aboveground groups.

We next investigated whether the rate of biotic homogenization was constant over the LUI gradient, and found that it peaked in the transition from low to intermediate LUI. Using generalized dissimilarity modelling (GDM) we fitted nonlinear effects of ΔLUI on β-diversity along the LUI gradient[13]. LUI was a major driver of β-turnover and total β-diversity, even compared to the large spatial and nutrient differences between the grasslands (deviance explained in Fig. 2b and relative effects in Extended Data Figs 5–7; Supplementary Information Section 5). There was a general trend for saturating responses in the β-diversity of aboveground (plants, herbivores and pollinators) and belowground groups (bacterivores and mycorrhizae) (Fig. 2b), which was parallel to the α-diversity response aboveground[1]. Differences in LUI between grasslands therefore drive turnover only at low overall LUI and increasing land-use heterogeneity beyond a certain point will not increase β-diversity, supporting the conclusion that minimizing LUI across the landscape most effectively enhances β-diversity. Some other groups, plant pathogens and secondary consumers, showed accelerating responses in β-diversity, which indicates strong homogenization in the most intensively managed grasslands. When we analysed the effects of the LUI components (grazing, mowing and fertilization) separately in the GDMs, mowing intensity was the main driver of biotic homogenization for most groups (Extended Data Fig. 8). Frequent mowing creates a homogenous sward, reduces flowering and seed set, causes high insect mortality and may lead to soil compaction, all of which may cause extinctions of rare species and favour a small set of disturbance-tolerant species both above- and belowground. In a global analysis, elevated nutrient input proved to be a main driver of soil microbial community composition[25]. In our study, fertilization had comparatively minor effects: increased homogeneity in soil nutrient levels at high LUI seemed to reduce β-diversity (see Supplementary Table 5 for LUI results without soil nutrients) less than homogenization of disturbance regimes.

Effects of LUI on total β-diversity were generally larger when measurements were weighted by species abundances (Extended Data Fig. 4). Intensively managed grasslands may be dominated by the same common species, even if they differ in their rare species. Indeed, a common

RESEARCH LETTER

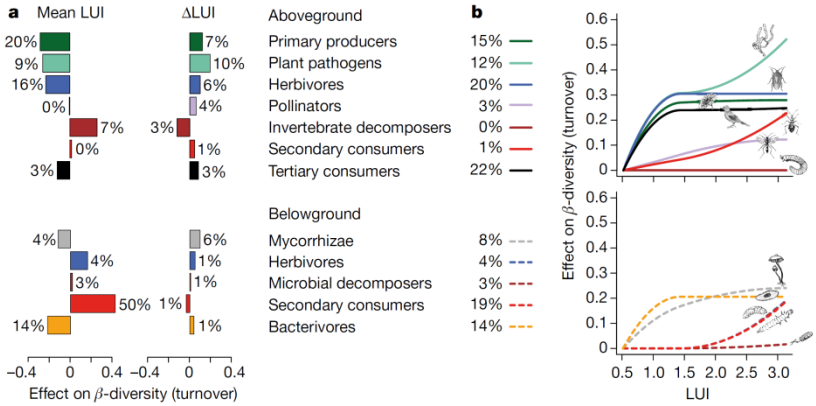


Figure 2 | Effects of LUI on β-diversity above- and belowground. **a**, Bars show partial effects of mean LUI and differences in LUI (ΔLUI) between plot pairs (105 plots), on species turnover (β_{sim}), from linear models. Numbers adjoining bars are the proportion of explained variance uniquely explained by mean LUI or ΔLUI. **b**, Results from the generalized dissimilarity models (GDMs) showing the effect of ΔLUI on β_{sim} along the LUI gradient, with percentages of deviance uniquely explained by LUI. Higher maximum curves indicate larger effects. All effects are corrected for environmental covariates and explanatory variables are scaled to allow comparisons across trophic levels.

practice in grassland management is to seed intensively used plots with a few fast-growing species of high nutritional value, which reinforces the homogenization of plant communities under high LUI. Increased abundance of common, generalist species might also drive biotic homogenization in other trophic groups. Under high LUI, dominance increased in most aboveground groups, although not in belowground groups (Extended Data Fig. 1). Increased dominance by a small set of common species, across a wide array of trophic groups, might threaten the delivery of critical ecosystem services in intensively managed landscapes[26].

Despite the overall consistency of land-use effects, some exceptions are worth noting. Bacteria responded weakly and had very low β-diversity, perhaps because their taxonomic resolution was coarser than for other groups (Methods). Responses of pollinators were also weak, possibly because their β-diversity responds more to land use at the landscape scale[27], as shown by the strong response to grassland isolation (Extended Data Fig. 5, Supplementary Information Section 5). In most other groups, isolation was a much less important driver of β-diversity. Only in three invertebrate groups did β-diversity increase with LUI; however, these groups were species-poor and had a lower sample coverage (see Supplementary Information Section 2 and Extended Data Fig. 9).The relative importance of LUI as a driver of β-diversity therefore varied between trophic groups (Extended Data Fig. 7), but it affected key ecosystem service providers such as plants and herbivores, as well as rare birds, which have a high conservation value.

High LUI also homogenized trophic structure and disrupted correlations between β-diversity of adjacent trophic levels. We calculated correlations between β-diversities for sets of plots with low, versus high,

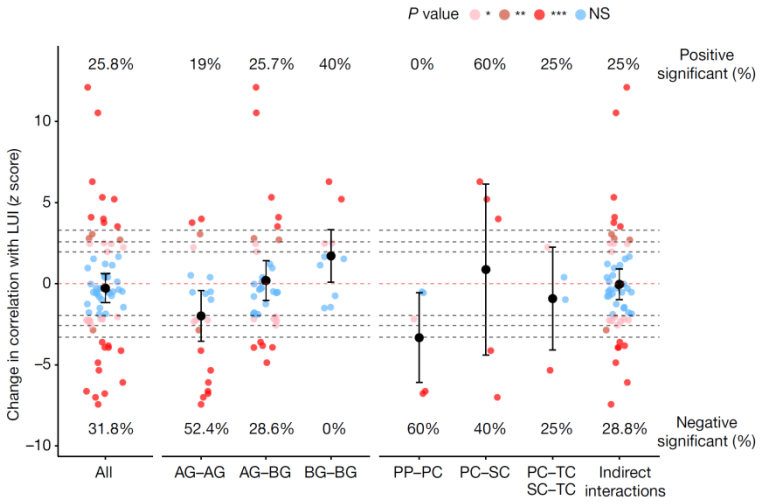


Figure 3 | Effect of LUI on correlations between the β-diversities (β_{sim}) of different trophic groups. z scores (positive z scores indicate that correlations are higher at high LUI, and negative z scores indicate that correlations are lower at high LUI) and P values (dashed lines separate P levels; $^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, NS, not significant) show the change in correlation strength (R^2 values from matrix regressions, corrected for effects of differences in LUI) between low (52 plots) and high LUI (53), comparing observed values to random values (see Methods). All correlations were grouped into categories: AG, aboveground; BG, belowground; PP, primary producers; PC, primary consumers; SC, secondary consumers; TC, tertiary consumers. Each coloured dot represents one correlation, black dots represent the mean and black bars the 95% confidence intervals. For statistical details see Supplementary Information Section 5.

LUI and expected a drop in correlations at high LUI, which would indicate biotic homogenization. Correlation strength dropped by more than 50% on average (for $R^2 > 0.1$ at low LUI) at high LUI and correlations between aboveground groups and between producers and primary consumers (plants and herbivores or pathogens) declined substantially (Fig. 3 and Extended Data Fig. 10), potentially reflecting a loss of host specialists. Some correlations increased in strength but these mainly involved the species-poor invertebrate groups whose β-diversity increased with LUI (see Supplementary Table 6). A previous study[28] showed that land-use intensification disrupted correlations in α-diversity and we extend this finding to show spatial decoupling for a wider range of trophic groups.

By analysing a uniquely comprehensive biodiversity dataset, we showed that LUI substantially reduces β-diversity across many different trophic groups. This threatens biodiversity by homogenizing communities within and across agricultural landscapes. The consequences of biotic homogenization for landscape-scale ecosystem service provisioning remain uncertain, but are likely to be severe[26]. Moreover, our results show that measures to reduce management intensity should be most effective at intermediate LUI[12]; they also underscore the high value of extensively managed grassland for conserving multitrophic diversity by showing that reduced intensity across the landscape effectively promotes large-scale diversity. This could be achieved by increasing the area of extensively managed grasslands in general, and especially by reducing the intensity of mowing. Conservation strategies and agricultural policies will increase in effectiveness if they aim to maintain the heterogeneity of biotic communities at the landscape scale, for instance by coupling subsidies to landscape-scale measures of diversity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 June; accepted 25 October 2016.
Published online 30 November 2016.

1. Allan, E. *et al.* Interannual variation in land-use intensity enhances grassland multidiversity. *Proc. Natl Acad. Sci. USA* **111,** 308–313 (2014).
2. Sala, O. E. *et al.* Global biodiversity scenarios for the year 2100. *Science* **287,** 1770–1774 (2000).
3. Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity. *Nature* **520,** 45–50 (2015).
4. Foley, J. A. *et al.* Global consequences of land use. *Science* **309,** 570–574 (2005).
5. Anderson, M. J. *et al.* Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecol. Lett.* **14,** 19–28 (2011).
6. Gámez-Virués, S. *et al.* Landscape simplification filters species traits and drives biotic homogenization. *Nat. Commun.* **6,** 8568 (2015).
7. Karp, D. S. *et al.* Intensive agriculture erodes β-diversity at large scales. *Ecol. Lett.* **15,** 963–970 (2012).
8. McKinney, M. L. & Lockwood, J. L. Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends Ecol. Evol.* **14,** 450–453 (1999).
9. Smart, S. M. *et al.* Biotic homogenization and changes in species diversity across human-modified ecosystems. *Proc. R. Soc. B* **273,** 2659–2665 (2006).
10. Dornelas, M. *et al.* Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344,** 296–299 (2014).
11. Solar, R. R. C. *et al.* How pervasive is biotic homogenization in human-modified tropical forest landscapes? *Ecol. Lett.* **18,** 1108–1118 (2015).
12. Kleijn, D. *et al.* On the relationship between farmland biodiversity and land-use intensity in *Europe. Proc. R. Soc. B* **276,** 903–909 (2009).
13. Ferrier, S., Manion, G., Elith, J. & Richardson, K. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers. Distrib.* **13,** 252–264 (2007).
14. Fitzpatrick, M. C. *et al.* Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. *Proc. R. Soc. B* **280,** 20131201 (2013).
15. Carvalho, J. C., Cardoso, P., Borges, P. A. V., Schmera, D & Podani, J. Measuring fractions of beta diversity and their relationships to nestedness: a theoretical and empirical comparison of novel approaches. *Oikos* **122,** 825–834 (2013).
16. Baselga, A. & Leprieur, F. Comparing methods to separate components of beta diversity. *Methods Ecol. Evol.* **6,** 1069–1079 (2015).
17. Rodrigues, J. L. M. *et al.* Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proc. Natl Acad. Sci. USA* **110,** 988–993 (2013).
18. Pellissier, L. *et al.* Turnover of plant lineages shapes herbivore phylogenetic beta diversity along ecological gradients. *Ecol. Lett.* **16,** 600–608 (2013).
19. Blüthgen, N. *et al.* A quantitative index of land-use intensity in grasslands: integrating mowing, grazing and fertilization. *Basic Appl. Ecol.* **13,** 207–220 (2012).
20. De Deyn, G. B. & Van der Putten, W. H. Linking aboveground and belowground diversity. *Trends Ecol. Evol.* **20,** 625–633 (2005).
21. Haimi, J., Fritze, H. & Moilanen, P. Responses of soil decomposer animals to wood-ash fertilisation and burning in a coniferous forest stand. *For. Ecol. Manage.* **129,** 53–61 (2000).
22. Bardgett, R., Hobbs, P. & Frostegård, Å. Changes in soil fungal:bacterial biomass ratios following reductions in the intensity of management of an upland grassland. *Biol. Fertil. Soils* **22,** 261–264 (1996).
23. Fenchel, T. & Finlay, B. J. The ubiquity of small species: patterns of local and global diversity. *Bioscience* **54,** 777–784 (2004).
24. Finlay, B. J. Global dispersal of free-living microbial eukaryote species. *Science* **296,** 1061–1063 (2002).
25. Leff, J. W. *et al.* Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc. Natl Acad. Sci. USA* **112,** 10967–10972 (2015).
26. van der Plas, F. *et al.* Biotic homogenization can decrease landscape-scale forest multifunctionality. *Proc. Natl Acad. Sci. USA* **113,** 3557–3562 (2016).
27. Perović, D. *et al.* Configurational landscape heterogeneity shapes functional community composition of grassland butterflies. *J. Appl. Ecol.* **52,** 505–513 (2015).
28. Manning, P. *et al.* Grassland management intensification weakens the associations among the diversities of multiple plant and animal taxa. *Ecology* **96,** 1492–1501 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to J. Chase and M. Fitzpatrick for their comments and suggestions on a previous version of the manuscript; B. Büche, R. Achtziger, T. Wagner, F. Köhler, T. Blick and M.-A. Fritze for arthropod species identification and U. Kern for creating the small icons of the 12 trophic groups used in the figures. We thank the managers of the three Exploratories, K. Hartwich, S. Gockel, K. Wiesner and M. Gorke for their work in maintaining the plot and project infrastructure; C. Fischer and S. Pfeiffer for giving support through the central office, M. Owonibi for managing the central data base; and E. Linsenmair, D. Hessenmöller, J. Nieschulze, I. Schöning and the late E. Kalko for their role in setting up the Biodiversity Exploratories project. We are also grateful to E. Kalko for her invaluable inspiration and for launching the studies on bats and birds. The work has been funded by the DFG Priority Program 1374 'Infrastructure-Biodiversity-Exploratories'. Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen and Brandenburg (according to §72 BbgNatSchG).

Author Contributions M.M.G. and E.A. conceived the idea for the manuscript, and defined the final analysis, M.M.G., E.A., C.P., T.M.L. and T.K. analysed the data, M.M.G. and E.A. wrote the first manuscript draft and finalized the manuscript. A.M.K., C.B., C.N.W., C.W., D.J.P., D.P., E.P., F.B., H.A., I.S., J.K., J.M., J.S., J.O., K.J., K.B., M.Tü., M.Ts., M.F., M.L., M.M.G., M.W., N.B., P.C.V., S.BI., S.Bo., S.A.S., S.C.R, S.K., S.W., T.D., T.W., V.B., V.W., and W.W.W. contributed data. T.M.L., F.G., S.Bo., D.P., L.R.J., K.B., S.C.R., A.C.K., O.P., P.S., T.T., W.W.W. and J.S. contributed substantially to revisions. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M.G. (martin.gossner@wsl.ch).

Reviewer Information *Nature* thanks P. Barton, S. Prober and the other anonymous reviewer(s) for their contribution to the peer review of this work.

RESEARCH LETTER

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Study system. The study was conducted as part of the Biodiversity Exploratories project (www.biodiversity-exploratories.de) in three German regions: (1) The UNESCO Biosphere area Schwäbische Alb in the low mountain range in south-western Germany (420 km^2, 460–860 m above sea level (a.s.l.)); (2) the Hainich National Park and its surrounding areas in the hilly lands of central Germany (1560 km^2, 285–550 m a.s.l.); and (3) the UNESCO Biosphere Reserve Schorfheide-Chorin in the glacially formed lowlands of northeastern Germany (1300 km^2 in size, 3–140 m a.s.l.). The three regions differ in climate, geology and topography, but each is characterized by a gradient of LUIs typical for large parts of temperate Europe[29]. Land-use gradients in this study range from semi-natural to intensively managed grasslands, because natural grasslands, that do not require management to prevent succession to forest, are almost absent from western and central Europe. In each region, 50 grassland plots were chosen from a total of 500 candidate plots, on which initial vegetation and land-use surveys were conducted, by stratified random sampling. This ensured that the plots covered the whole range of LUIs and management types, while minimizing confounding factors such as spatial position or soil type[29]. Thereby we avoided, for instance, sampling low-intensity grasslands only in the low productive parts of the landscape. All grasslands have a long history of broadly similar LUI (that is, low intensity grasslands have not been recently converted from high intensity grasslands and vice versa and all had been grasslands for at least 10 years before the start of project), although we are aware that temporal variation in land use is substantial[1]. In this study, we analysed a subset of 105 plots (Schwäbische Alb: 32, Hainich-Dün: 37, Schorfheide-Chorin: 36) for which data on all taxa (see below) were available.

Study design and land-use measures. All plots were continuously managed by farmers. Information on management practices, including the level of fertilization (kg N ha^{-1} yr^{-1}), grazing (number of livestock units ha^{-1} yr^{-1}) and mowing (number of cuts per year), was assessed annually by standardized interviews with the land owners. LUI at the local scale was then quantified as a compound index on the basis of summing the standardized intensities of these three components[19]. We decided to employ a compound index of LUI because the individual components are correlated with each other (fertilization and mowing are positively correlated, and grazing and mowing negatively correlated[19]) and the distribution of overall intensity is more even (each individual component has many 0 values). Each component was divided by the global mean value for each year to standardize the components[19]. We then calculated the mean LUI for each plot over five years (2006–2010) because this reflects the average LUI around the years when most of the data was assessed (2008, 2009 and 2011). At the low end of land-use intensity, with a LUI of 0.5, grasslands are typically unfertilized, not mown, and grazed by 40–50 sheep per hectare for about 10 days (or more rarely by 1–3 cattle per hectare for 20 days). At an intermediate LUI of 1.5, grasslands are usually unfertilized (or fertilized with less than 30 kg N ha^{-1} y^{-1}), and are either mown twice a year or grazed by four cattle per hectare for about 50 days. At the high end of land-use intensity with a LUI of 3, grasslands are typically fertilized at a rate of 60–120 kg N ha^{-1} y^{-1}, are mown 2–3 times a year or grazed by 5–10 cattle for 100–150 days, or are managed by a combination of grazing and mowing. In addition to using the LUI index (that is, where all three types of land use are given equal weight), we tested the individual standardized land-use components in our models to separate nutrient and disturbance effects.

Covariates. We corrected for a series of other variables that might affect diversity. In addition to the mean land use across time, we quantified two other measures of land use: (1) the temporal variation of LUI, as its standard deviation (sdLUI) across five years (2006–2010) because this has been shown to increase α-diversity[1]; and (2) to measure effects of land use at the landscape scale, which may also be important drivers of diversity[6], we calculated one minus the proportion of grasslands (including managed grasslands and semi-natural vegetation and thus all potential habitat for grassland species) within 500 m of each plot as a proxy for isolation. Plot neighbourhood cover was mapped in 2009 using high resolution (40 cm) aerial photographs from 2008, including the following land-cover types besides grasslands: arable land, forest, roads, trees (woodlots smaller than 1 ha), urban areas and water bodies.

We also corrected for soil nutrients and pH. We sampled the upper 10 cm of the mineral soil to assess soil variables. This was conducted in May 2011 with 14 locations per plot along two 18-m transects with distances of 3 m between sampling points, for details see ref. 30 and Supplementary Information Section 1. Soil nutrients were quantified by the first axis of a principal component analysis (PCA, ade4 package[31]) of nitrogen stock, total soil nitrogen, organic carbon stock, soil inorganic carbon, soil organic carbon, soil C:N ratio (Supplementary Table 1). Note that our measure of soil nutrients partly includes, in addition to natural variation among soils, the effect of fertilization, which in turn is also part of the LUI. Therefore, our estimates on the effects of LUI are on the conservative side and fitting models without the soil nutrient variable slightly increases the strength of the LUI effects in some cases, see Supplementary Information Sections 4, 5. Finally, in the analysis of β-diversity, we also corrected for spatial distance between all sites. To consider spatial effects on β-diversity, we calculated geographic distance between each pair of plots on the basis of geographic coordinates using Euclidean distances in the R package vegan[32].

Biodiversity assessment. We assessed biodiversity across a broad range of organisms from bacteria to vertebrates using molecular (bacteria, arbuscular mycorrhizal fungi and protists) as well as morphological (plants, arthropods and pathogenic fungi) or acoustic characteristics (bird and bat calls).

Vascular plants, bryophytes and lichens. We sampled vascular plants between mid-May and mid-June 2009, bryophytes (2007–2008) and lichens (2007–2008) in an area of 4 m × 4 m on each plot, and estimated the percentage cover of each occurring species. For details see refs 33, 34 for vascular plants, ref. 35 for lichens and ref. 36 for bryophytes).

Arthropods. All arthropods of the herb layer were sampled in 2008 by different methods. For sampling Araneae, Coleoptera, Hemiptera: Heteroptera and Auchenorrhyncha, Hymenoptera: Symphyta, Neuroptera, Orthoptera, Dermaptera and Dictyoptera we used biannual (June and August) sweep netting by conducting 60 double sweeps along three 50-m long plot border transects[37,38]. Additionally, Diptera and Hymenoptera were hand-collected during their visits on flowers, identified and individuals counted[39]. This survey involved a transect of 200 m × 3 m along the edge of the plot, for which three transect walks were performed on a single day (total, 6 h). In some cases, plots were measured several times; these were averaged in less than one month apart or, if repeated over one month later, the earlier measure was used[39]. We conducted surveys of butterflies and day-active moths (hereafter termed as Lepidoptera) from beginning of May to mid August[40,41]. We sampled Lepidoptera on fixed transects in the three regions repeating the sampling three times in a randomized sequence within each region. Each transect had a length of 300 m and we recorded all Lepidoptera within 30 min per site within a 5 m corridor.

Soil arthropods (Myriapoda, soil living larvae) were sampled in spring 2011 (within ten days in April) by collecting two soil cores (diameter, 20 cm, depth, 10 cm) from each plot. Soil fauna was extracted from the first core using a modified heat extraction system[42] over a period of eight days and the second soil core was hand-sorted for soil macrofauna.

All arthropod species were assigned to one of four trophic groups (herbivores, pollinators, predators and decomposers) on the basis of their known main food resource as adults.

Pathogenic fungi. From July to August 2011, we sampled pathogenic fungi including rust, powdery mildew, downy mildew and smut fungi in four transects of 25 m × 1 m per plot. We inspected all occurring vascular plant species for infested individuals, sampled them and later identified the pathogenic fungi to the species level.

Birds. Birds were sampled by audio-visual point counts[43] covering the area of the respective grassland plot (50 m × 50 m). We noted all individuals of each bird species during the five-minute interval. In each year, from 2008 to 2012, we visited each plot five times between 15 March and 15 June (1st surveying period, 15–30 March; 2nd, 15–30 April; 3rd, 1–15 May; 4th, 16–31 May; 5th, 1–15 June). A maximum of 15 plots was surveyed per day from sunrise to 11:00; occasionally the evening chorus was surveyed after 17:00 (<20 times out of 750 events per year). The sequence in which plots were visited was randomized. The maximum number of birds displaying per site per year (that is, the maximum number of record individuals per species over the five rounds within a surveying year) was used as a measure of the relative abundance of birds. We considered a species as present in the particular plot if it was recorded at least once during a survey within each year. Aerial species (swifts and swallows) were excluded from analysis, as they had only been surveyed irregularly and without standardization. For this study, we combined species richness and relative abundance data across the five sampling periods.

Bats. We surveyed bats from June to September during the years 2008, 2009 and 2010. Plot sampling was conducted along a 24 min point-stop transect of 200 m at the borders of each grassland plot. Sampling started 30 min after local sunset and was limited to the first half of the night (01:00) to account for the first peak in bat activity[44]. We randomly sampled 4–6 plots per night. Each plot was surveyed twice during each year with a minimum time interval between repeated sampling of five weeks. Acoustic recordings of bats were taken in real time (sample rate: 384 kHz, 16 bit) with a Pettersson-D1000x bat detector (Pettersson Electronic AG, Uppsala, Sweden) and triggered manually by an observer listening through headphones to the output of the heterodyne system while continuously scanning the frequency

range between 20 and 80 kHz. Bat species identification was conducted using Avisoft SAS Laboratory Pro, Version 5.0.24 (R. Specht, Avisoft Bioacoustics, Berlin, Germany, Hamming window, 1024 FFT, 96% overlap) following various references on echolocation call parameters, for example refs 45–47. For details on species identification see ref. 48. In addition, we evaluated the number of bat passes that were defined as a minimum of two consecutive echolocation calls[49]. Successive passes within one recording were discriminated if the time interval between calls was larger than three times the regular pulse interval of the particular species[48,50].

Belowground microorganisms. At each grassland plot fourteen soil cores (diameter, 8.3 cm) were taken from a 20 m × 20 m subarea and soil from the upper 10 cm of the A horizon was homogenized after removal of root material. The bulk sample was split into subsamples for the analyses of bacteria, protists and arbuscular mycorrhizal fungi.

For bacteria, 10 g of the homogenized soil was put immediately on liquid nitrogen and stored until RNA extraction[51]. Briefly, total RNA was isolated from soils and reverse transcribed into cDNA. Amplicons of the V3 region of the 16S rRNA gene were sequenced on an Illumina Hiseq platform using universal bacterial primers as described in ref. 52.

For the analysis of protists, 1 g of the bulk soil sample was used for DNA extraction and the analyses of the V4 region of the 18S rRNA gene amplified using eukaryotic specific primers. Sequences were filtered for (1) 100% forward primer match; (2) length ≥ 200–710 bp and (3) ambiguities (*N*). Traces were scanned for chimaeras, trimmed to 530 bp, dereplicated to group 100% identical amplicons, and singletons removed. Remaining sequences were treated as operational taxonomic units (OTUs) and aligned to the PR^2 database using BLASTn (default parameters). One hit per sequence was retained. Only OTUs with 100% coverage and protist taxa (excluding Metazoa, Fungi and Streptophyta) were retained for analysis.

For the study of arbuscular mycorrhizal fungi, total microbial DNA was isolated from the bulk soil sample using a MoBioPowerSoil DNA Isolation Kit. The NS31-a.m.1 fragment of the fungal 18S rDNA was amplified using arbuscular mycorrhizal fungal specific primers[53] and sequenced using a Genome Sequencer FLX+ 454 System. The reads were quality filtered using Mothur[54] and classified using the MaarjaM AMF reference database[55]. A total of 825 arbuscular mycorrhizal fungal OTUs were detected.

Detailed description of the data processing of bacteria, protists and arbuscular mycorrhizal fungi is presented in Supplementary Information Section 2.

Statistical analyses. All analyses were conducted in R 3.0.2 (ref. 56).

Sample completeness. To test for sample completeness, we used a previously published approach of sample coverage[57,58]. Coverage is defined as the proportion of the total number of individuals in an assemblage that belong to species represented in the sample. We used two approaches to estimate sample coverage. First, we estimated sample coverage for low (52) and high (53) LUI plots on the basis of species incidences. Second, we estimated sample coverage for each plot on the basis of species abundances. Sample coverage did not differ significantly along the LUI gradient and was estimated to be higher than 90% in all trophic groups, except aboveground invertebrate decomposers and secondary consumers (Extended Data Fig. 10 and Supplementary Table 2-2). Therefore, the results for secondary consumers and invertebrate decomposers should be treated with caution but there is no evidence that any undersampling is biased along the LUI gradient. A further line of evidence for the robustness of our findings to issues of sample completeness is provided by the results for different q-levels. When increasing q-levels, rare species are less strongly weighted in the calculation of β-diversity. As we found similar or even stronger responses of β-diversity to LUI at higher q-levels (see below) this further indicates that undersampling of rare species is unlikely to affect the conclusions of our study. Analyses were conducted using the iNEXT function in the iNEXT library[59].

Diversity measures. We calculated several measures of α- and β-diversity for each of the 12 trophic groups (Fig. 1). We used previously published q-metrics[60,61] to incorporate different weightings for species abundance in α- and β-diversity. These are based on Hill numbers[62], and allow the calculation of diversity measures in which increasing weight is given to species abundances. At $q = 0$, rare and abundant species are weighted equally, which corresponds to species richness for α-diversity and the Sørensen index of dissimilarity for β-diversity. At $q = 1$, species are weighted in proportion to their frequency in the sampled community, which corresponds to the exponential of Shannon entropy (or effective number of species) for α-diversity and Horn's index of dissimilarity for β-diversity. Finally, at $q = 2$, abundant species receive more weight relative to their frequency and this corresponds to the inverse Simpson index for β-diversity and Morisita–Horn index for β-diversity[63,64]. Analyses were conducted using multipart function in the vegan library[65].

Spatial turnover in composition between locations involves two main processes: a replacement of species (pure turnover) and changes in species richness[15,16]. To test for effects on β-diversity that are independent of species richness differences, we used the Simpson dissimilarity β_{sim}, which is the turnover component of Sørensen dissimilarity, see also ref. 16. Details on β- (diversity across the three study regions) as well as β-diversity partitioning is given in Supplementary Information Section 3.

α-diversity analysis. To analyse the response of α-diversity to land use, we used power law models that allow different shapes of responses to be fitted. We modelled the response of α-diversity for each of the 12 trophic groups, calculated with $q = 0$, 1 or 2. The explanatory variable was LUI, the model formula was $y = a + (b \times \mathrm{LUI}^c)$, where a (intercept), b (slope) and c (degree of curvation) are parameters estimated by the model. In order to correct for confounding environmental effects we analysed residuals in the power law models. We calculated residuals from linear models with diversity or land use (LUI) as the response variable and region, soil nutrients, pH, variation in LUI (sdLUI) and isolation (1 – proportion of grasslands in the plot surrounding) as explanatory variables. We calculated residuals because incorporating many explanatory variables in the power law models would have led to extremely complex models. After taking residuals, we then scaled all explanatory and response variables between 0 and 1, to allow comparison of effects and responses. Models were fitted using the gnls function in the nlme library[66].

β-diversity analysis. *Linear models.* To analyse the response of β-diversity to land-use intensification, we first fitted linear models. These were fitted to values of turnover (β_{sim}), total β-diversity (Sørensen, $q = 0$) and abundance weighted β-diversity ($q = 1$ or 2) for each of the 12 trophic groups. Explanatory variables in these models were: the mean LUI, sdLUI, isolation, soil nutrients and pH between each pair of plots. The effects of mean LUI provide a test of biotic homogenization: a negative effect indicates that land-use intensification reduces turnover. In this case, reducing LUI across the landscape would promote β-diversity. However, as plot pairs with the same mean LUI can either be very similar in LUI or come from different ends of the gradient, we also fitted differences in LUI (ΔLUI) between all plot pairs. This term tests for the effect of land-use heterogeneity, that is, whether β-diversity is higher between plots of different intensities. A positive effect would suggest that maximizing land-use heterogeneity across the landscape would increase β-diversity. The terms mean LUI and ΔLUI are not correlated with each other although ΔLUI is constrained to zero at maximum and minimum mean LUI. We additionally fitted differences in all other variables (sdLUI, isolation, soil nutrients and pH) together with the spatial distance between all plot pairs in the models. To compare the effects of the different predictors, we scaled all predictors to between 0 and 1. We then calculated the variance explained uniquely by mean LUI or ΔLUI by comparing the variance explained by the full model with that explained by models containing all terms except mean LUI or ΔLUI. The unique variance is expressed as a proportion of the total explained variance. We calculated the significance of all terms in the linear models using a permutation procedure, implemented with the lmp function in the lmPerm library[67], using 100,000 iterations.

We also ran a second series of linear models in which we replaced mean LUI with the mean grazing, mowing and fertilization intensity between plot pairs (all scaled to the maximum across plots) and replaced ΔLUI with differences in grazing, mowing and fertilization between plots. Finally, we ran linear models without soil nutrient levels, to test whether effects of LUI were mediated by its effects on soil nutrients.

Generalized dissimilarity modelling. To analyse the nonlinear effects of differences in LUI, we used GDM[13]. This is a matrix regression technique for modelling turnover in species composition between sites as a function of the spatial and/or environmental distance between them. The advantages of GDM are that it can incorporate variation in the rate of compositional turnover along an environmental or spatial gradient (non-stationarity) and that it allows the relationships between dissimilarity and distance to be nonlinear. The only constraint is that compositional turnover is assumed to always increase with distance between sites (monotonicity). For more details on GDM see ref. 13. All GDMs were fitted using the gdm function in the gdm library[68]. We plot the partial effect of each predictor, that is, while holding all other predictors constant, against the level of a given predictors to visualize the results of the GDM (Extended Data Fig. 5 and Supplementary Table 5-1). The height of the line shows how large the effect of LUI is relative to all other predictors in the model. Variation in compositional turnover along an environmental or spatial gradient can be seen from the shape of the line, which shows how the effect of a given predictor on compositional turnover varies with the mean level of that predictor. For instance, for LUI, the shape of the line shows how the effect of heterogeneity in LUI varies with mean LUI. We also calculated a bootstrapped P value for each term in the full GDM, using the gdm.varImp function in the gdm library (Supplementary Table 5-2). Additionally we estimated uncertainty for the GDM plots by using 100 bootstraps for each model, each time removing 30% of the plot pairs and then fitting a GDM and extracting the predictions. We then calculated the s.d. of the predictions and added this (±) to the fitted line (Extended Data Fig. 6). This is based on the function plotUncertainty in the gdm library.

We fitted GDMs to four measures of β-diversity for each group: turnover (β_{sim}), total β-diversity (Sorensen, $q=0$) and abundance weighted β-diversity ($q=1$ and 2). In each case differences in LUI were fitted as an explanatory variable. To correct for spatial, environmental and other land-use distances, we additionally fitted the spatial distance between plots, differences in pH, differences in nutrients, differences in sdLUI and differences in isolation in the model. We also ran GDMs with individual land-use components, that is, with grazing, mowing and fertilization, instead of LUI. These models had the same covariates as the LUI models. For the linear models, we ran the LUI GDMs without soil nutrient levels to test whether some effects of LUI were driven by soil nutrients. In both cases, the effects of LUI were very similar regardless of whether soil nutrients were included or not; this indicates that LUI effects on β-diversity are mostly not caused by LUI homogenizing the soil abiotic environment.

Furthermore, it is possible to estimate the amount of deviance in compositional turnover explained by the GDM. Note that the GDM optimizes the fit between predictors and response variables, so different models can have different response spline shapes. We also determined the proportion of deviance uniquely attributable to land use. We did this by comparing the deviance explained by a GDM containing all of the variables and a GDM with all variables except the difference in LUI between plots. We calculated the unique deviance explained by LUI as the difference in deviance explained between these models. We then converted this to a percentage by dividing by the deviance explained by the full GDM.

Correlation in β-diversity between trophic groups. To test for possible effects of land use on the correlation of β-diversities between trophic groups, we used partial multivariate correlograms and multiple regressions (pmgram and MRM functions in the ecodist library[69]). We correlated β-diversity of different groups (β_{sim}, $q=0$, $q=1$, and $q=2$; see above), and corrected for LUI distances between plots. We corrected for LUI to account for potential shared responses to common environmental drivers. We did this by using the residuals for the matrix correlations between trophic levels. The multiple regressions use permutation tests (999 permutations) of significance for the regression coefficients and for the R^2 values.

To test whether the strength of correlations differed between low and high LUIs, we divided the 105 plots into 52 low (less than median LUI) and 53 high (greater than median LUI) intensity plots and calculated the R^2-value differences between high and low LUI (R^2_{high}–R^2_{low}). We then compared these values to a distribution of simulated R^2-values differences ($n=1{,}899$) where we randomized the LUI differences between plots. On the basis of this random distribution, we calculated Z scores (standardized effect sizes (SES)) and P values. Significant values thus indicate stronger trophic interactions at lower (or higher) LUI than expected by chance.

Data availability. The data will become publicly available according to the Rules of Procedure of the German Science Foundation (DFG)-funded Biodiversity Exploratories, that is, five years after completion of the datasets.

29. Fischer, M. *et al.* Implementing large-scale and long-term functional biodiversity research: the Biodiversity Exploratories. *Basic Appl. Ecol.* **11,** 473–485 (2010).
30. Birkhofer, K. *et al.* General relationships between abiotic soil properties and soil biota across spatial scales and different land-use types. *PLoS One* **7,** e43292 (2012).
31. Dray, S. & Dufour, A. B. The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* **22,** 1–20 (2007).
32. *vegan: Community Ecology Package.* R package version 2.3-3. https://CRAN.R-project.org/package=vegan (2016).
33. Socher, S. A. *et al.* Interacting effects of fertilization, mowing and grazing on plant species diversity of 1500 grasslands in Germany differ between regions. *Basic Appl. Ecol.* **14,** 126–136 (2013).
34. Socher, S. A. *et al.* Direct and productivity-mediated indirect effects of fertilization, mowing and grazing on grassland species richness. *J. Ecol.* **100,** 1391–1399 (2012).
35. Boch, S., Prati, D., Schöning, I. & Fischer, M. Lichen species richness is highest in non-intensively used grasslands promoting suitable microhabitats and low vascular plant competition. *Biodivers. Conserv.* **25,** 225–238 (2016).
36. Müller, J. *et al.* Impact of land-use intensity and productivity on bryophyte diversity in agricultural grasslands. *PLoS One* **7,** e51520 (2012).
37. Simons, N. K. *et al.* Resource-mediated indirect effects of grassland management on arthropod diversity. *PLoS One* **9,** e107033 (2014).
38. Simons, N. K. *et al.* Effects of land-use intensity on arthropod species abundance distributions in grasslands. *J. Anim. Ecol.* **84,** 143–154 (2015).
39. Weiner, C. N., Werner, M., Linsenmair, K. E. & Blüthgen, N. Land-use impacts on plant–pollinator networks: interaction strength and specialization predict pollinator declines. *Ecology* **95,** 466–474 (2014).
40. Börschig, C. *Effects of land-use intensity in grasslands on diversity, life-history traits and multitrophic interactions* Dr. rer. nat. thesis, Georg-August-Universität (2012).
41. Börschig, C., Klein, A. M., von Wehrden, H. & Krauss, J. Traits of butterfly communities change from specialist to generalist characteristics with increasing land-use intensity. *Basic Appl. Ecol.* **14,** 547–554 (2013). 10.1016/j.baae.2013.09.002
42. Kempson, D., Lloyd, M. & Ghelardi, R. A new extractor for woodland litter. *Pedobiologia* **3,** 1–21 (1963).
43. Renner, S. C. *et al.* Temporal changes in randomness of bird communities across central Europe. *PLoS One* **9,** e112347 (2014).
44. Rydell, J., Entwistle, A. & Racey, P. A. Timing of foraging flights of three species of bats in relation to insect activity and predation risk. *Oikos* **76,** 243–252 (1996).
45. Denzinger, A., Siemers, B. M., Schaub, A. & Schnitzler, H.-U. Echolocation by the barbastelle bat, *Barbastella barbastellus*. *J. Comp. Physiol. A* **187,** 521–528 (2001).
46. Russo, D. & Jones, G. Identification of twenty-two bat species (Mammalia: Chiroptera) from Italy by analysis of time-expanded recordings of echolocation calls. *J. Zool.* **258,** 91–103 (2002).
47. Obrist, M. K., Boesch, R. & Flückiger, P. F. Variability in echolocation call design of 26 Swiss bat species: consequences, limits and options for automated field identification with a synergetic pattern recognition approach. *Mammalia* **68,** 307–322 (2004).
48. Jung, K., Kaiser, S., Böhm, S., Nieschulze, J. & Kalko, E. K. V. Moving in three dimensions: effects of structural complexity on occurrence and activity of insectivorous bats in managed forest stands. *J. Appl. Ecol.* **49,** 523–531 (2012). 10.1111/j.1365-2664.2012.02116.x
49. Fenton, M. B. in *Bat Echolocation Research: Tools, Techniques and Analysis* (eds Brigham, M. *et al.*) 133–140 (Bat Conservation International, 2004).
50. Estrada-Villegas, S., Meyer, C. F. J. & Kalko, E. K. V. Effects of tropical forest fragmentation on aerial insectivorous bats in a land-bridge island system. *Biol. Conserv.* **143,** 597–608 (2010).
51. Lueders, T., Manefield, M. & Friedrich, M. W. Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. *Environ. Microbiol.* **6,** 73–78 (2004).
52. Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl. Environ. Microbiol.* **77,** 3846–3852 (2011).
53. Morris, E. K. *et al.* Land use and host neighbor identity effects on arbuscular mycorrhizal fungal community composition in focal plant rhizosphere. *Biodivers. Conserv.* **22,** 2193–2205 (2013).
54. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).
55. Öpik, M. *et al.* The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytol.* **188,** 223–241 (2010).
56. *R: A language and environment for statistical computing v. 3.2.2.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2015).
57. Chao, A. *et al.* Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84,** 45–67 (2014).
58. Chao, A. & Jost, L. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* **93,** 2533–2547 (2012).
59. *iNEXT: iNterpolation and EXTrapolation for species diversity.* R package version 2.0, http://chao.stat.nthu.edu.tw/blog/software-download (2014).
60. Jost, L. Entropy and diversity. *Oikos* **113,** 363–375 (2006).
61. Jost, L. Partitioning diversity into independent alpha and beta components. *Ecology* **88,** 2427–2439 (2007).
62. Hill, M. O. Diversity and eveness: unifying notations and its consequences. *Ecology* **54,** 427–432 (1973).
63. Maurer, B. A. & McGill, B. J. in *Biological Diversity: Frontiers in Measurement and Assessment* Vol. 12 (eds Magurran, A. E. & McGill, B. J.) 55–65 (Oxford Univ. Press, 2011).
64. Jost, L., Chao, A. & Chazdon, R. in *Biological Diversity: Frontiers in Measurement and Assessment* Vol. 12 (eds Magurran, A. E. & McGill, B. J.) 66–84 (Oxford Univ. Press, 2011).
65. *vegan: Community Ecology Package.* R package version 2.2-1. (2015).
66. *nlme: Linear and Nonlinear Mixed Effects Models.* R package version 3.1-125, http://CRAN.R-project.org/package=nlme (2016).
67. *lmPerm: Permutation Tests for Linear Models.* R package version 2.1.0. https://CRAN.R-project.org/package=lmPerm (2016).
68. *gdm: Functions for Generalized Dissimilarity Modeling v.* R-package version 1.1-7 (2016).
69. Goslee, S. C. & Urban, D. L. The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* **22,** 1–19 (2007).

RESEARCH LETTER

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Study system. The study was conducted as part of the Biodiversity Exploratories project (www.biodiversity-exploratories.de) in three German regions: (1) The UNESCO Biosphere area Schwäbische Alb in the low mountain range in south-western Germany (420 km^2, 460–860 m above sea level (a.s.l.)); (2) the Hainich National Park and its surrounding areas in the hilly lands of central Germany (1560 km^2, 285–550 m a.s.l.); and (3) the UNESCO Biosphere Reserve Schorfheide-Chorin in the glacially formed lowlands of northeastern Germany (1300 km^2 in size, 3–140 m a.s.l.). The three regions differ in climate, geology and topography, but each is characterized by a gradient of LUIs typical for large parts of temperate Europe[29]. Land-use gradients in this study range from semi-natural to intensively managed grasslands, because natural grasslands, that do not require management to prevent succession to forest, are almost absent from western and central Europe. In each region, 50 grassland plots were chosen from a total of 500 candidate plots, on which initial vegetation and land-use surveys were conducted, by stratified random sampling. This ensured that the plots covered the whole range of LUIs and management types, while minimizing confounding factors such as spatial position or soil type[29]. Thereby we avoided, for instance, sampling low-intensity grasslands only in the low productive parts of the landscape. All grasslands have a long history of broadly similar LUI (that is, low intensity grasslands have not been recently converted from high intensity grasslands and vice versa and all had been grasslands for at least 10 years before the start of project), although we are aware that temporal variation in land use is substantial[1]. In this study, we analysed a subset of 105 plots (Schwäbische Alb: 32, Hainich-Dün: 37, Schorfheide-Chorin: 36) for which data on all taxa (see below) were available.

Study design and land-use measures. All plots were continuously managed by farmers. Information on management practices, including the level of fertilization (kg N ha^{-1} yr^{-1}), grazing (number of livestock units ha^{-1} yr^{-1}) and mowing (number of cuts per year), was assessed annually by standardized interviews with the land owners. LUI at the local scale was then quantified as a compound index on the basis of summing the standardized intensities of these three components[19]. We decided to employ a compound index of LUI because the individual components are correlated with each other (fertilization and mowing are positively correlated, and grazing and mowing negatively correlated[19]) and the distribution of overall intensity is more even (each individual component has many 0 values). Each component was divided by the global mean value for each year to standardize the components[19]. We then calculated the mean LUI for each plot over five years (2006–2010) because this reflects the average LUI around the years when most of the data was assessed (2008, 2009 and 2011). At the low end of land-use intensity, with a LUI of 0.5, grasslands are typically unfertilized, not mown, and grazed by 40–50 sheep per hectare for about 10 days (or more rarely by 1–3 cattle per hectare for 20 days). At an intermediate LUI of 1.5, grasslands are usually unfertilized (or fertilized with less than 30 kg N ha^{-1} y^{-1}), and are either mown twice a year or grazed by four cattle per hectare for about 50 days. At the high end of land-use intensity with a LUI of 3, grasslands are typically fertilized at a rate of 60–120 kg N ha^{-1} y^{-1}, are mown 2–3 times a year or grazed by 5–10 cattle for 100–150 days, or are managed by a combination of grazing and mowing. In addition to using the LUI index (that is, where all three types of land use are given equal weight), we tested the individual standardized land-use components in our models to separate nutrient and disturbance effects.

Covariates. We corrected for a series of other variables that might affect diversity. In addition to the mean land use across time, we quantified two other measures of land use: (1) the temporal variation of LUI, as its standard deviation (sdLUI) across five years (2006–2010) because this has been shown to increase α-diversity[1]; and (2) to measure effects of land use at the landscape scale, which may also be important drivers of diversity[6], we calculated one minus the proportion of grasslands (including managed grasslands and semi-natural vegetation and thus all potential habitat for grassland species) within 500 m of each plot as a proxy for isolation. Plot neighbourhood cover was mapped in 2009 using high resolution (40 cm) aerial photographs from 2008, including the following land-cover types besides grasslands: arable land, forest, roads, trees (woodlots smaller than 1 ha), urban areas and water bodies.

We also corrected for soil nutrients and pH. We sampled the upper 10 cm of the mineral soil to assess soil variables. This was conducted in May 2011 with 14 locations per plot along two 18-m transects with distances of 3 m between sampling points, for details see ref. 30 and Supplementary Information Section 1. Soil nutrients were quantified by the first axis of a principal component analysis (PCA, ade4 package[31]) of nitrogen stock, total soil nitrogen, organic carbon stock, soil inorganic carbon, soil organic carbon, soil C:N ratio (Supplementary Table 1). Note that our measure of soil nutrients partly includes, in addition to natural variation among soils, the effect of fertilization, which in turn is also part of the LUI. Therefore, our estimates on the effects of LUI are on the conservative side and fitting models without the soil nutrient variable slightly increases the strength of the LUI effects in some cases, see Supplementary Information Sections 4, 5. Finally, in the analysis of β-diversity, we also corrected for spatial distance between all sites. To consider spatial effects on β-diversity, we calculated geographic distance between each pair of plots on the basis of geographic coordinates using Euclidean distances in the R package vegan[32].

Biodiversity assessment. We assessed biodiversity across a broad range of organisms from bacteria to vertebrates using molecular (bacteria, arbuscular mycorrhizal fungi and protists) as well as morphological (plants, arthropods and pathogenic fungi) or acoustic characteristics (bird and bat calls).

Vascular plants, bryophytes and lichens. We sampled vascular plants between mid-May and mid-June 2009, bryophytes (2007–2008) and lichens (2007–2008) in an area of 4 m × 4 m on each plot, and estimated the percentage cover of each occurring species. For details see refs 33, 34 for vascular plants, ref. 35 for lichens and ref. 36 for bryophytes).

Arthropods. All arthropods of the herb layer were sampled in 2008 by different methods. For sampling Araneae, Coleoptera, Hemiptera: Heteroptera and Auchenorrhyncha, Hymenoptera: Symphyta, Neuroptera, Orthoptera, Dermaptera and Dictyoptera we used biannual (June and August) sweep netting by conducting 60 double sweeps along three 50-m long plot border transects[37,38]. Additionally, Diptera and Hymenoptera were hand-collected during their visits on flowers, identified and individuals counted[39]. This survey involved a transect of 200 m × 3 m along the edge of the plot, for which three transect walks were performed on a single day (total, 6 h). In some cases, plots were measured several times; these were averaged in less than one month apart or, if repeated over one month later, the earlier measure was used[39]. We conducted surveys of butterflies and day-active moths (hereafter termed as Lepidoptera) from beginning of May to mid August[40,41]. We sampled Lepidoptera on fixed transects in the three regions repeating the sampling three times in a randomized sequence within each region. Each transect had a length of 300 m and we recorded all Lepidoptera within 30 min per site within a 5 m corridor.

Soil arthropods (Myriapoda, soil living larvae) were sampled in spring 2011 (within ten days in April) by collecting two soil cores (diameter, 20 cm, depth, 10 cm) from each plot. Soil fauna was extracted from the first core using a modified heat extraction system[42] over a period of eight days and the second soil core was hand-sorted for soil macrofauna.

All arthropod species were assigned to one of four trophic groups (herbivores, pollinators, predators and decomposers) on the basis of their known main food resource as adults.

Pathogenic fungi. From July to August 2011, we sampled pathogenic fungi including rust, powdery mildew, downy mildew and smut fungi in four transects of 25 m × 1 m per plot. We inspected all occurring vascular plant species for infested individuals, sampled them and later identified the pathogenic fungi to the species level.

Birds. Birds were sampled by audio-visual point counts[43] covering the area of the respective grassland plot (50 m × 50 m). We noted all individuals of each bird species during the five-minute interval. In each year, from 2008 to 2012, we visited each plot five times between 15 March and 15 June (1st surveying period, 15–30 March; 2nd, 15–30 April; 3rd, 1–15 May; 4th, 16–31 May; 5th, 1–15 June). A maximum of 15 plots was surveyed per day from sunrise to 11:00; occasionally the evening chorus was surveyed after 17:00 (<20 times out of 750 events per year). The sequence in which plots were visited was randomized. The maximum number of birds displaying per site per year (that is, the maximum number of record individuals per species over the five rounds within a surveying year) was used as a measure of the relative abundance of birds. We considered a species as present in the particular plot if it was recorded at least once during a survey within each year. Aerial species (swifts and swallows) were excluded from analysis, as they had only been surveyed irregularly and without standardization. For this study, we combined species richness and relative abundance data across the five sampling periods.

Bats. We surveyed bats from June to September during the years 2008, 2009 and 2010. Plot sampling was conducted along a 24 min point-stop transect of 200 m at the borders of each grassland plot. Sampling started 30 min after local sunset and was limited to the first half of the night (01:00) to account for the first peak in bat activity[44]. We randomly sampled 4–6 plots per night. Each plot was surveyed twice during each year with a minimum time interval between repeated sampling of five weeks. Acoustic recordings of bats were taken in real time (sample rate: 384 kHz, 16 bit) with a Pettersson-D1000x bat detector (Pettersson Electronic AG, Uppsala, Sweden) and triggered manually by an observer listening through headphones to the output of the heterodyne system while continuously scanning the frequency

range between 20 and 80 kHz. Bat species identification was conducted using Avisoft SAS Laboratory Pro, Version 5.0.24 (R. Specht, Avisoft Bioacoustics, Berlin, Germany, Hamming window, 1024 FFT, 96% overlap) following various references on echolocation call parameters, for example refs 45–47. For details on species identification see ref. 48. In addition, we evaluated the number of bat passes that were defined as a minimum of two consecutive echolocation calls[49]. Successive passes within one recording were discriminated if the time interval between calls was larger than three times the regular pulse interval of the particular species[48,50].

Belowground microorganisms. At each grassland plot fourteen soil cores (diameter, 8.3 cm) were taken from a 20 m × 20 m subarea and soil from the upper 10 cm of the A horizon was homogenized after removal of root material. The bulk sample was split into subsamples for the analyses of bacteria, protists and arbuscular mycorrhizal fungi.

For bacteria, 10 g of the homogenized soil was put immediately on liquid nitrogen and stored until RNA extraction[51]. Briefly, total RNA was isolated from soils and reverse transcribed into cDNA. Amplicons of the V3 region of the 16S rRNA gene were sequenced on an Illumina Hiseq platform using universal bacterial primers as described in ref. 52.

For the analysis of protists, 1 g of the bulk soil sample was used for DNA extraction and the analyses of the V4 region of the 18S rRNA gene amplified using eukaryotic specific primers. Sequences were filtered for (1) 100% forward primer match; (2) length ≥ 200–710 bp and (3) ambiguities (*N*). Traces were scanned for chimaeras, trimmed to 530 bp, dereplicated to group 100% identical amplicons, and singletons removed. Remaining sequences were treated as operational taxonomic units (OTUs) and aligned to the PR^2 database using BLASTn (default parameters). One hit per sequence was retained. Only OTUs with 100% coverage and protist taxa (excluding Metazoa, Fungi and Streptophyta) were retained for analysis.

For the study of arbuscular mycorrhizal fungi, total microbial DNA was isolated from the bulk soil sample using a MoBioPowerSoil DNA Isolation Kit. The NS31-a.m.1 fragment of the fungal 18S rDNA was amplified using arbuscular mycorrhizal fungal specific primers[53] and sequenced using a Genome Sequencer FLX+ 454 System. The reads were quality filtered using Mothur[54] and classified using the MaarjaM AMF reference database[55]. A total of 825 arbuscular mycorrhizal fungal OTUs were detected.

Detailed description of the data processing of bacteria, protists and arbuscular mycorrhizal fungi is presented in Supplementary Information Section 2.

Statistical analyses. All analyses were conducted in R 3.0.2 (ref. 56).

Sample completeness. To test for sample completeness, we used a previously published approach of sample coverage[57,58]. Coverage is defined as the proportion of the total number of individuals in an assemblage that belong to species represented in the sample. We used two approaches to estimate sample coverage. First, we estimated sample coverage for low (52) and high (53) LUI plots on the basis of species incidences. Second, we estimated sample coverage for each plot on the basis of species abundances. Sample coverage did not differ significantly along the LUI gradient and was estimated to be higher than 90% in all trophic groups, except aboveground invertebrate decomposers and secondary consumers (Extended Data Fig. 10 and Supplementary Table 2-2). Therefore, the results for secondary consumers and invertebrate decomposers should be treated with caution but there is no evidence that any undersampling is biased along the LUI gradient. A further line of evidence for the robustness of our findings to issues of sample completeness is provided by the results for different q-levels. When increasing q-levels, rare species are less strongly weighted in the calculation of β-diversity. As we found similar or even stronger responses of β-diversity to LUI at higher q-levels (see below) this further indicates that undersampling of rare species is unlikely to affect the conclusions of our study. Analyses were conducted using the iNEXT function in the iNEXT library[59].

Diversity measures. We calculated several measures of α- and β-diversity for each of the 12 trophic groups (Fig. 1). We used previously published q-metrics[60,61] to incorporate different weightings for species abundance in α- and β-diversity. These are based on Hill numbers[62], and allow the calculation of diversity measures in which increasing weight is given to species abundances. At $q = 0$, rare and abundant species are weighted equally, which corresponds to species richness for α-diversity and the Sørensen index of dissimilarity for β-diversity. At $q = 1$, species are weighted in proportion to their frequency in the sampled community, which corresponds to the exponential of Shannon entropy (or effective number of species) for α-diversity and Horn's index of dissimilarity for β-diversity. Finally, at $q = 2$, abundant species receive more weight relative to their frequency and this corresponds to the inverse Simpson index for β-diversity and Morisita–Horn index for β-diversity[63,64]. Analyses were conducted using multipart function in the vegan library[65].

Spatial turnover in composition between locations involves two main processes: a replacement of species (pure turnover) and changes in species richness[15,16]. To test for effects on β-diversity that are independent of species richness differences, we used the Simpson dissimilarity β_{sim}, which is the turnover component of Sørensen dissimilarity, see also ref. 16. Details on β- (diversity across the three study regions) as well as β-diversity partitioning is given in Supplementary Information Section 3.

α-diversity analysis. To analyse the response of α-diversity to land use, we used power law models that allow different shapes of responses to be fitted. We modelled the response of α-diversity for each of the 12 trophic groups, calculated with $q = 0$, 1 or 2. The explanatory variable was LUI, the model formula was $y = a + (b \times \mathrm{LUI}^c)$, where a (intercept), b (slope) and c (degree of curvation) are parameters estimated by the model. In order to correct for confounding environmental effects we analysed residuals in the power law models. We calculated residuals from linear models with diversity or land use (LUI) as the response variable and region, soil nutrients, pH, variation in LUI (sdLUI) and isolation (1 – proportion of grasslands in the plot surrounding) as explanatory variables. We calculated residuals because incorporating many explanatory variables in the power law models would have led to extremely complex models. After taking residuals, we then scaled all explanatory and response variables between 0 and 1, to allow comparison of effects and responses. Models were fitted using the gnls function in the nlme library[66].

β-diversity analysis. *Linear models.* To analyse the response of β-diversity to land-use intensification, we first fitted linear models. These were fitted to values of turnover (β_{sim}), total β-diversity (Sørensen, $q = 0$) and abundance weighted β-diversity ($q = 1$ or 2) for each of the 12 trophic groups. Explanatory variables in these models were: the mean LUI, sdLUI, isolation, soil nutrients and pH between each pair of plots. The effects of mean LUI provide a test of biotic homogenization: a negative effect indicates that land-use intensification reduces turnover. In this case, reducing LUI across the landscape would promote β-diversity. However, as plot pairs with the same mean LUI can either be very similar in LUI or come from different ends of the gradient, we also fitted differences in LUI (ΔLUI) between all plot pairs. This term tests for the effect of land-use heterogeneity, that is, whether β-diversity is higher between plots of different intensities. A positive effect would suggest that maximizing land-use heterogeneity across the landscape would increase β-diversity. The terms mean LUI and ΔLUI are not correlated with each other although ΔLUI is constrained to zero at maximum and minimum mean LUI. We additionally fitted differences in all other variables (sdLUI, isolation, soil nutrients and pH) together with the spatial distance between all plot pairs in the models. To compare the effects of the different predictors, we scaled all predictors to between 0 and 1. We then calculated the variance explained uniquely by mean LUI or ΔLUI by comparing the variance explained by the full model with that explained by models containing all terms except mean LUI or ΔLUI. The unique variance is expressed as a proportion of the total explained variance. We calculated the significance of all terms in the linear models using a permutation procedure, implemented with the lmp function in the lmPerm library[67], using 100,000 iterations.

We also ran a second series of linear models in which we replaced mean LUI with the mean grazing, mowing and fertilization intensity between plot pairs (all scaled to the maximum across plots) and replaced ΔLUI with differences in grazing, mowing and fertilization between plots. Finally, we ran linear models without soil nutrient levels, to test whether effects of LUI were mediated by its effects on soil nutrients.

Generalized dissimilarity modelling. To analyse the nonlinear effects of differences in LUI, we used GDM[13]. This is a matrix regression technique for modelling turnover in species composition between sites as a function of the spatial and/or environmental distance between them. The advantages of GDM are that it can incorporate variation in the rate of compositional turnover along an environmental or spatial gradient (non-stationarity) and that it allows the relationships between dissimilarity and distance to be nonlinear. The only constraint is that compositional turnover is assumed to always increase with distance between sites (monotonicity). For more details on GDM see ref. 13. All GDMs were fitted using the gdm function in the gdm library[68]. We plot the partial effect of each predictor, that is, while holding all other predictors constant, against the level of a given predictors to visualize the results of the GDM (Extended Data Fig. 5 and Supplementary Table 5-1). The height of the line shows how large the effect of LUI is relative to all other predictors in the model. Variation in compositional turnover along an environmental or spatial gradient can be seen from the shape of the line, which shows how the effect of a given predictor on compositional turnover varies with the mean level of that predictor. For instance, for LUI, the shape of the line shows how the effect of heterogeneity in LUI varies with mean LUI. We also calculated a bootstrapped P value for each term in the full GDM, using the gdm.varImp function in the gdm library (Supplementary Table 5-2). Additionally we estimated uncertainty for the GDM plots by using 100 bootstraps for each model, each time removing 30% of the plot pairs and then fitting a GDM and extracting the predictions. We then calculated the s.d. of the predictions and added this (±) to the fitted line (Extended Data Fig. 6). This is based on the function plotUncertainty in the gdm library.

We fitted GDMs to four measures of β-diversity for each group: turnover (β_{sim}), total β-diversity (Sorensen, $q=0$) and abundance weighted β-diversity ($q=1$ and 2). In each case differences in LUI were fitted as an explanatory variable. To correct for spatial, environmental and other land-use distances, we additionally fitted the spatial distance between plots, differences in pH, differences in nutrients, differences in sdLUI and differences in isolation in the model. We also ran GDMs with individual land-use components, that is, with grazing, mowing and fertilization, instead of LUI. These models had the same covariates as the LUI models. For the linear models, we ran the LUI GDMs without soil nutrient levels to test whether some effects of LUI were driven by soil nutrients. In both cases, the effects of LUI were very similar regardless of whether soil nutrients were included or not; this indicates that LUI effects on β-diversity are mostly not caused by LUI homogenizing the soil abiotic environment.

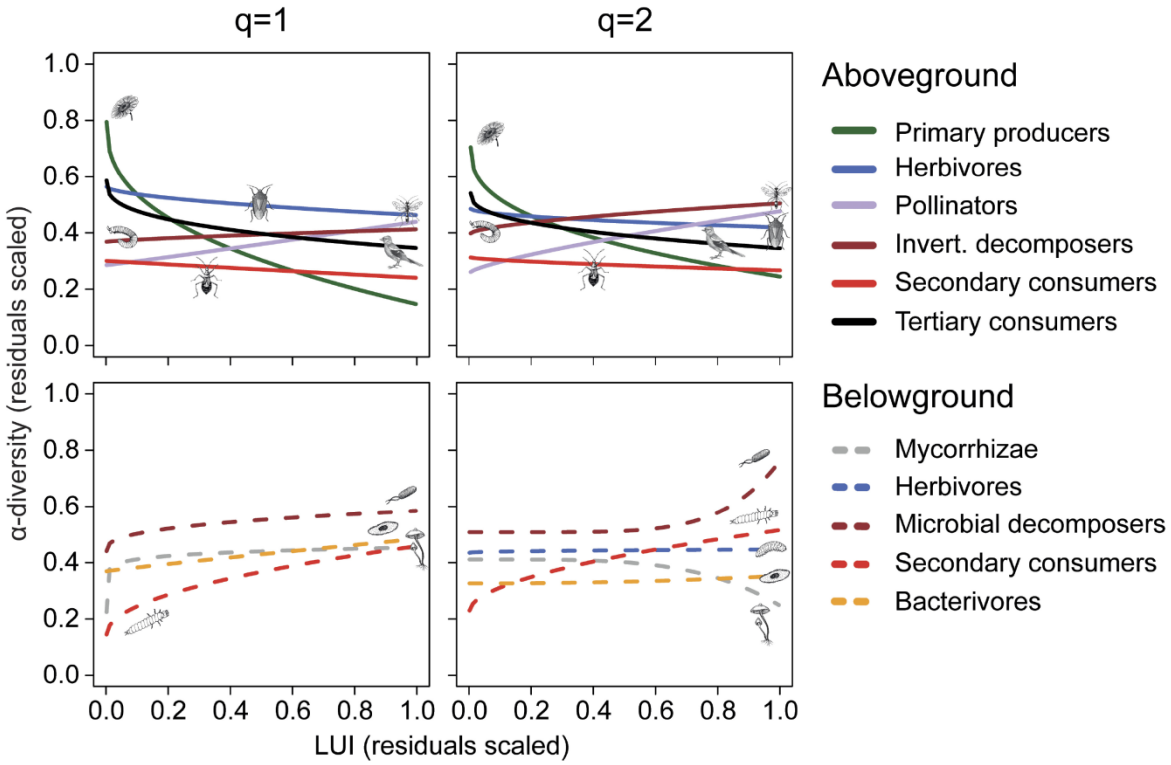
Furthermore, it is possible to estimate the amount of deviance in compositional turnover explained by the GDM. Note that the GDM optimizes the fit between predictors and response variables, so different models can have different response spline shapes. We also determined the proportion of deviance uniquely attributable to land use. We did this by comparing the deviance explained by a GDM containing all of the variables and a GDM with all variables except the difference in LUI between plots. We calculated the unique deviance explained by LUI as the difference in deviance explained between these models. We then converted this to a percentage by dividing by the deviance explained by the full GDM.

Correlation in β-diversity between trophic groups. To test for possible effects of land use on the correlation of β-diversities between trophic groups, we used partial multivariate correlograms and multiple regressions (pmgram and MRM functions in the ecodist library[69]). We correlated β-diversity of different groups (β_{sim}, $q=0$, $q=1$, and $q=2$; see above), and corrected for LUI distances between plots. We corrected for LUI to account for potential shared responses to common environmental drivers. We did this by using the residuals for the matrix correlations between trophic levels. The multiple regressions use permutation tests (999 permutations) of significance for the regression coefficients and for the R^2 values.

To test whether the strength of correlations differed between low and high LUIs, we divided the 105 plots into 52 low (less than median LUI) and 53 high (greater than median LUI) intensity plots and calculated the R^2-value differences between high and low LUI (R^2_{high}–R^2_{low}). We then compared these values to a distribution of simulated R^2-values differences ($n=1{,}899$) where we randomized the LUI differences between plots. On the basis of this random distribution, we calculated Z scores (standardized effect sizes (SES)) and P values. Significant values thus indicate stronger trophic interactions at lower (or higher) LUI than expected by chance.

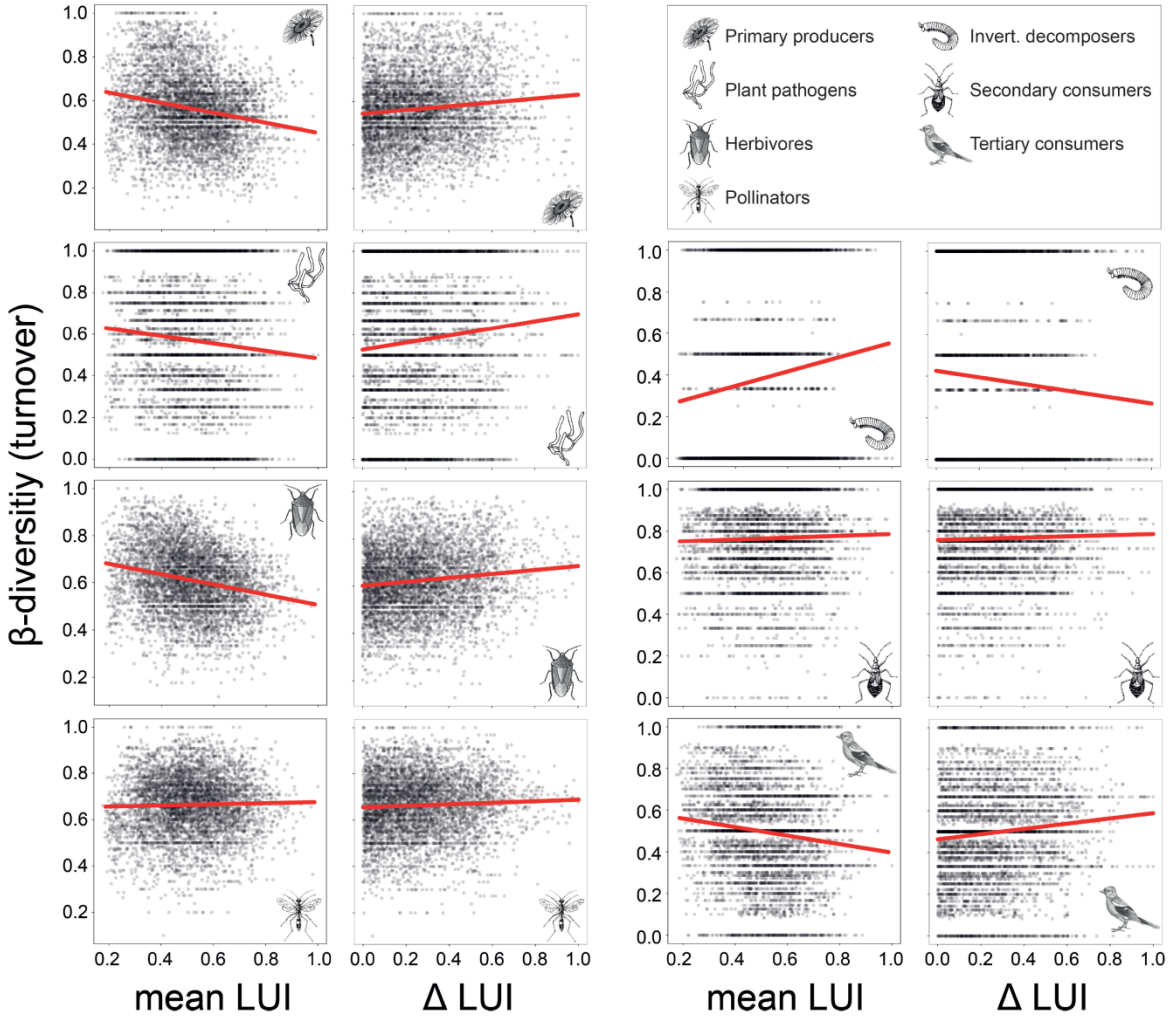
Data availability. The data will become publicly available according to the Rules of Procedure of the German Science Foundation (DFG)-funded Biodiversity Exploratories, that is, five years after completion of the datasets.

29. Fischer, M. *et al.* Implementing large-scale and long-term functional biodiversity research: the Biodiversity Exploratories. *Basic Appl. Ecol.* **11,** 473–485 (2010).
30. Birkhofer, K. *et al.* General relationships between abiotic soil properties and soil biota across spatial scales and different land-use types. *PLoS One* **7,** e43292 (2012).
31. Dray, S. & Dufour, A. B. The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* **22,** 1–20 (2007).
32. *vegan: Community Ecology Package.* R package version 2.3-3. https://CRAN.R-project.org/package=vegan (2016).
33. Socher, S. A. *et al.* Interacting effects of fertilization, mowing and grazing on plant species diversity of 1500 grasslands in Germany differ between regions. *Basic Appl. Ecol.* **14,** 126–136 (2013).
34. Socher, S. A. *et al.* Direct and productivity-mediated indirect effects of fertilization, mowing and grazing on grassland species richness. *J. Ecol.* **100,** 1391–1399 (2012).
35. Boch, S., Prati, D., Schöning, I. & Fischer, M. Lichen species richness is highest in non-intensively used grasslands promoting suitable microhabitats and low vascular plant competition. *Biodivers. Conserv.* **25,** 225–238 (2016).
36. Müller, J. *et al.* Impact of land-use intensity and productivity on bryophyte diversity in agricultural grasslands. *PLoS One* **7,** e51520 (2012).
37. Simons, N. K. *et al.* Resource-mediated indirect effects of grassland management on arthropod diversity. *PLoS One* **9,** e107033 (2014).
38. Simons, N. K. *et al.* Effects of land-use intensity on arthropod species abundance distributions in grasslands. *J. Anim. Ecol.* **84,** 143–154 (2015).
39. Weiner, C. N., Werner, M., Linsenmair, K. E. & Blüthgen, N. Land-use impacts on plant–pollinator networks: interaction strength and specialization predict pollinator declines. *Ecology* **95,** 466–474 (2014).
40. Börschig, C. *Effects of land-use intensity in grasslands on diversity, life-history traits and multitrophic interactions* Dr. rer. nat. thesis, Georg-August-Universität (2012).
41. Börschig, C., Klein, A. M., von Wehrden, H. & Krauss, J. Traits of butterfly communities change from specialist to generalist characteristics with increasing land-use intensity. *Basic Appl. Ecol.* **14,** 547–554 (2013). 10.1016/j.baae.2013.09.002
42. Kempson, D., Lloyd, M. & Ghelardi, R. A new extractor for woodland litter. *Pedobiologia* **3,** 1–21 (1963).
43. Renner, S. C. *et al.* Temporal changes in randomness of bird communities across central Europe. *PLoS One* **9,** e112347 (2014).
44. Rydell, J., Entwistle, A. & Racey, P. A. Timing of foraging flights of three species of bats in relation to insect activity and predation risk. *Oikos* **76,** 243–252 (1996).
45. Denzinger, A., Siemers, B. M., Schaub, A. & Schnitzler, H.-U. Echolocation by the barbastelle bat, *Barbastella barbastellus. J. Comp. Physiol. A* **187,** 521–528 (2001).
46. Russo, D. & Jones, G. Identification of twenty-two bat species (Mammalia: Chiroptera) from Italy by analysis of time-expanded recordings of echolocation calls. *J. Zool.* **258,** 91–103 (2002).
47. Obrist, M. K., Boesch, R. & Flückiger, P. F. Variability in echolocation call design of 26 Swiss bat species: consequences, limits and options for automated field identification with a synergetic pattern recognition approach. *Mammalia* **68,** 307–322 (2004).
48. Jung, K., Kaiser, S., Böhm, S., Nieschulze, J. & Kalko, E. K. V. Moving in three dimensions: effects of structural complexity on occurrence and activity of insectivorous bats in managed forest stands. *J. Appl. Ecol.* **49,** 523–531 (2012). 10.1111/j.1365-2664.2012.02116.x
49. Fenton, M. B. in *Bat Echolocation Research: Tools, Techniques and Analysis* (eds Brigham, M. *et al.*) 133–140 (Bat Conservation International, 2004).
50. Estrada-Villegas, S., Meyer, C. F. J. & Kalko, E. K. V. Effects of tropical forest fragmentation on aerial insectivorous bats in a land-bridge island system. *Biol. Conserv.* **143,** 597–608 (2010).
51. Lueders, T., Manefield, M. & Friedrich, M. W. Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. *Environ. Microbiol.* **6,** 73–78 (2004).
52. Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl. Environ. Microbiol.* **77,** 3846–3852 (2011).
53. Morris, E. K. *et al.* Land use and host neighbor identity effects on arbuscular mycorrhizal fungal community composition in focal plant rhizosphere. *Biodivers. Conserv.* **22,** 2193–2205 (2013).
54. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).
55. Öpik, M. *et al.* The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytol.* **188,** 223–241 (2010).
56. *R: A language and environment for statistical computing v. 3.2.2.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2015).
57. Chao, A. *et al.* Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84,** 45–67 (2014).
58. Chao, A. & Jost, L. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* **93,** 2533–2547 (2012).
59. *iNEXT: iNterpolation and EXTrapolation for species diversity.* R package version 2.0, http://chao.stat.nthu.edu.tw/blog/software-download (2014).
60. Jost, L. Entropy and diversity. *Oikos* **113,** 363–375 (2006).
61. Jost, L. Partitioning diversity into independent alpha and beta components. *Ecology* **88,** 2427–2439 (2007).
62. Hill, M. O. Diversity and eveness: unifying notations and its consequences. *Ecology* **54,** 427–432 (1973).
63. Maurer, B. A. & McGill, B. J. in *Biological Diversity: Frontiers in Measurement and Assessment* Vol. 12 (eds Magurran, A. E. & McGill, B. J.) 55–65 (Oxford Univ. Press, 2011).
64. Jost, L., Chao, A. & Chazdon, R. in *Biological Diversity: Frontiers in Measurement and Assessment* Vol. 12 (eds Magurran, A. E. & McGill, B. J.) 66–84 (Oxford Univ. Press, 2011).
65. *vegan: Community Ecology Package.* R package version 2.2-1. (2015).
66. *nlme: Linear and Nonlinear Mixed Effects Models.* R package version 3.1-125, http://CRAN.R-project.org/package=nlme (2016).
67. *lmPerm: Permutation Tests for Linear Models.* R package version 2.1.0. https://CRAN.R-project.org/package=lmPerm (2016).
68. *gdm: Functions for Generalized Dissimilarity Modeling v.* R-package version 1.1-7 (2016).
69. Goslee, S. C. & Urban, D. L. The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* **22,** 1–19 (2007).

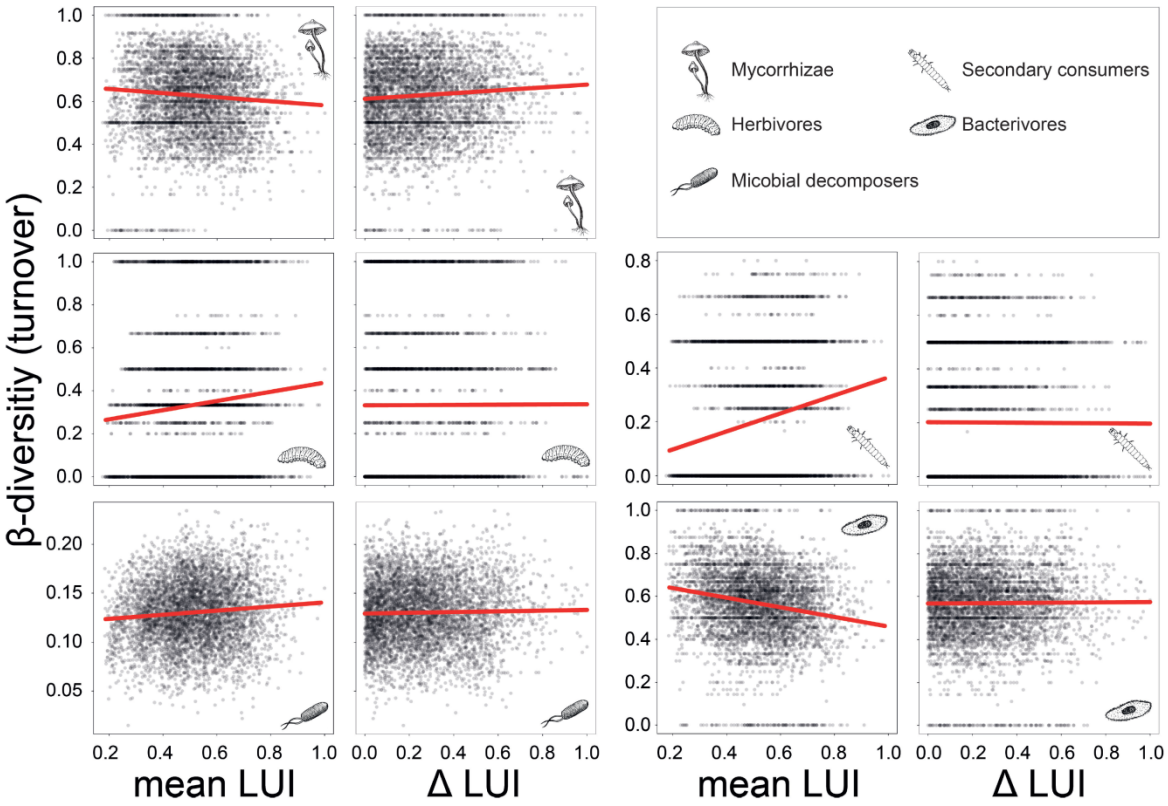


Extended Data Figure 1 | The effect of LUI on higher q-level α-diversity above- and belowground. The partial effect of local LUI comes from a power law model fitted to the exponential Shannon diversity ($q = 1$) and reciprocal Simpson index ($q = 2$) of the seven aboveground (solid lines) and the five belowground trophic groups (dashed lines) ($n = 105$ plots; for more details see Methods). In the model, all parameters of the power law function depended on temporal variation in LUI (sdLUI) and isolation. LUI effects are plotted at the mean values of these two variables. α-diversity and land-use variables were corrected for differences due to region, pH and soil nutrients, by taking residuals, and were then scaled between 0 and 1. The models for protists ($q = 1$ and $q = 2$) and mycorrhizae ($q = 2$) failed to converge and are therefore not shown. Note that plant pathogens are missing because, for this group, no data on abundance was available.

RESEARCH LETTER

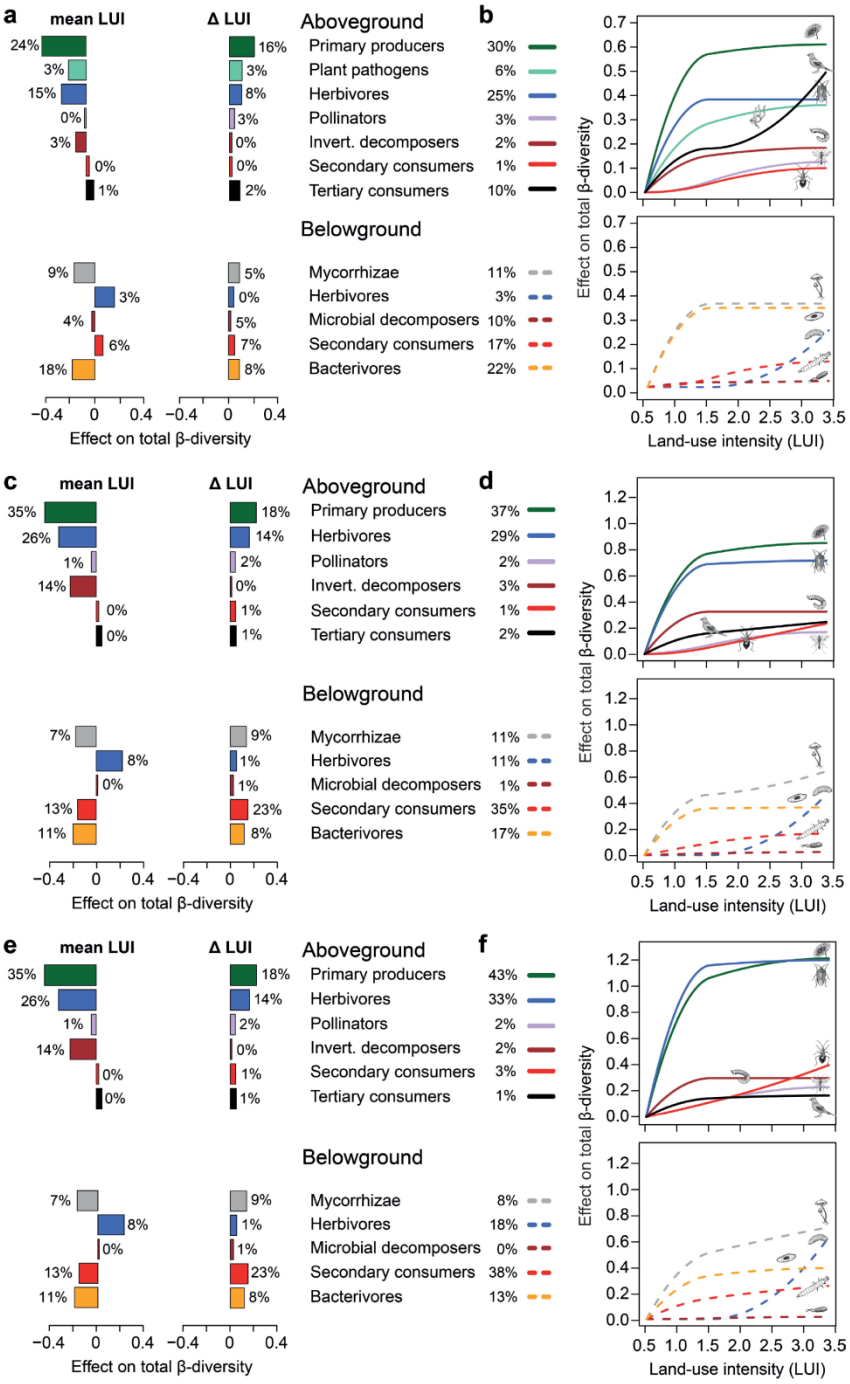


Extended Data Figure 2 | Effects of LUI on turnover of aboveground species. Scatter plots showing the effects of mean LUI and ΔLUI, between plot pairs ($n = 105$ plots), on the species turnover component of β-diversity for seven aboveground groups. Regression lines show predictions from linear models.

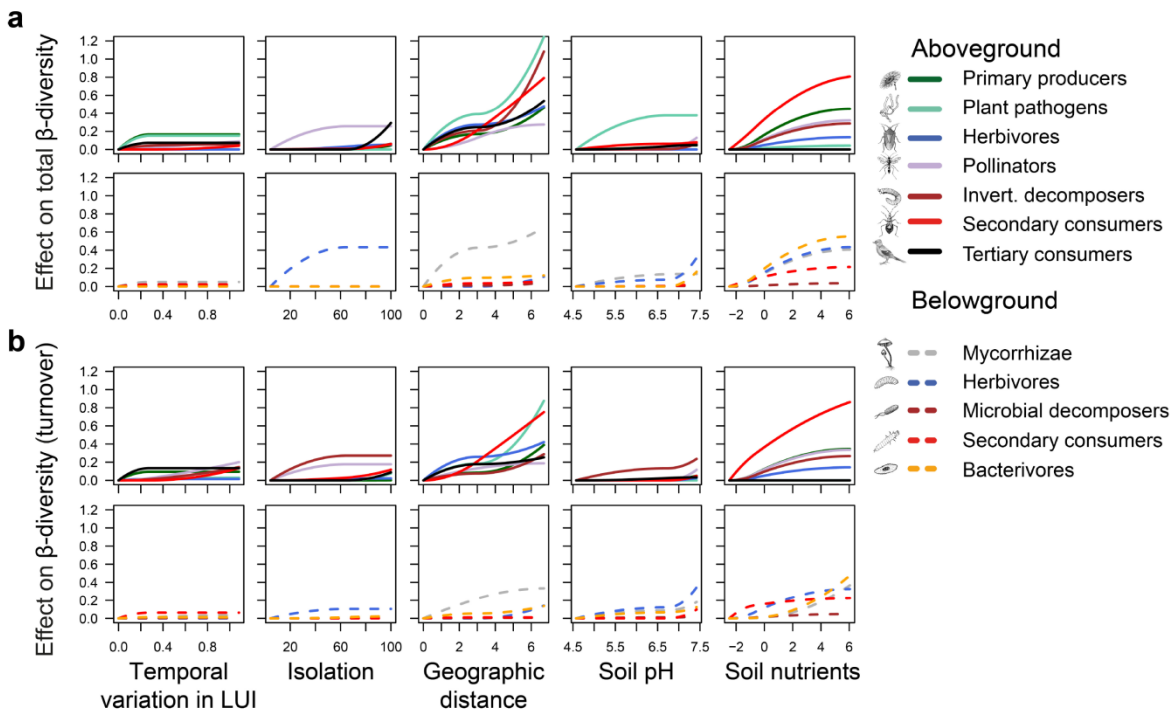


Extended Data Figure 3 | Effects of LUI on turnover of belowground species. Scatter plots showing the effects of mean LUI and ΔLUI, between plot pairs ($n = 105$ plots), on the species turnover component of β-diversity for five belowground groups. Regression lines show predictions from linear models.

RESEARCH LETTER

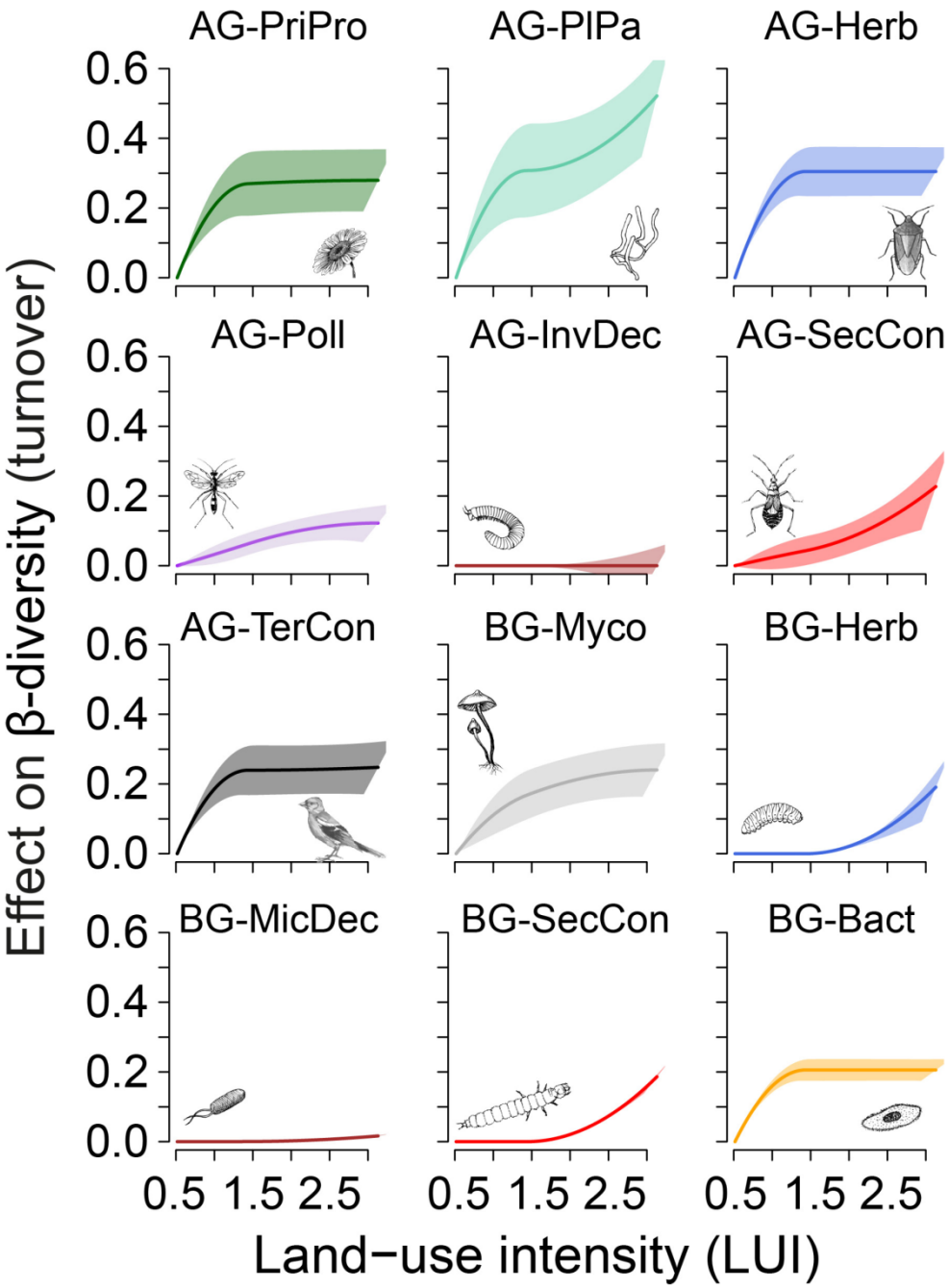


Extended Data Figure 4 | Effects of LUI on total β-diversity above- and belowground. **a**, **c**, **e**, Partial effects of mean LUI and ΔLUI, between plot pairs, on total β-diversity (**a**, Sørensen $q=0$; **c**, Morisita $q=1$; **e**, Morisita–Horn $q=2$) for seven aboveground and five belowground groups from linear models. Negative effects of mean LUI indicate that land-use intensification reduces β-diversity. The bars show coefficients from the models. Numbers adjoining bars are the proportion of explained variance uniquely explained by mean LUI or ΔLUI. **b**, **d**, **f**, Results from the GDMs are shown for total β-diversity (**b**, Sørensen $q=0$; **d**, Morisita $q=1$; **f**, Morisita–Horn $q=2$) for the same trophic groups. The figures show the effect of differences in LUI on β-diversity (calculated between all plot pairs). Effects of differences in LUI can vary nonlinearly along the gradient of LUI. Higher maximum curves indicate larger effects of differences in LUI on β-diversity. The values in the legend are the percentage of deviance that is explained uniquely by LUI. Effects of both linear models and GDMs are corrected for other drivers of β-diversity, and response and explanatory variables are scaled to allow comparisons across trophic groups ($n=105$ plots; for details see Methods).

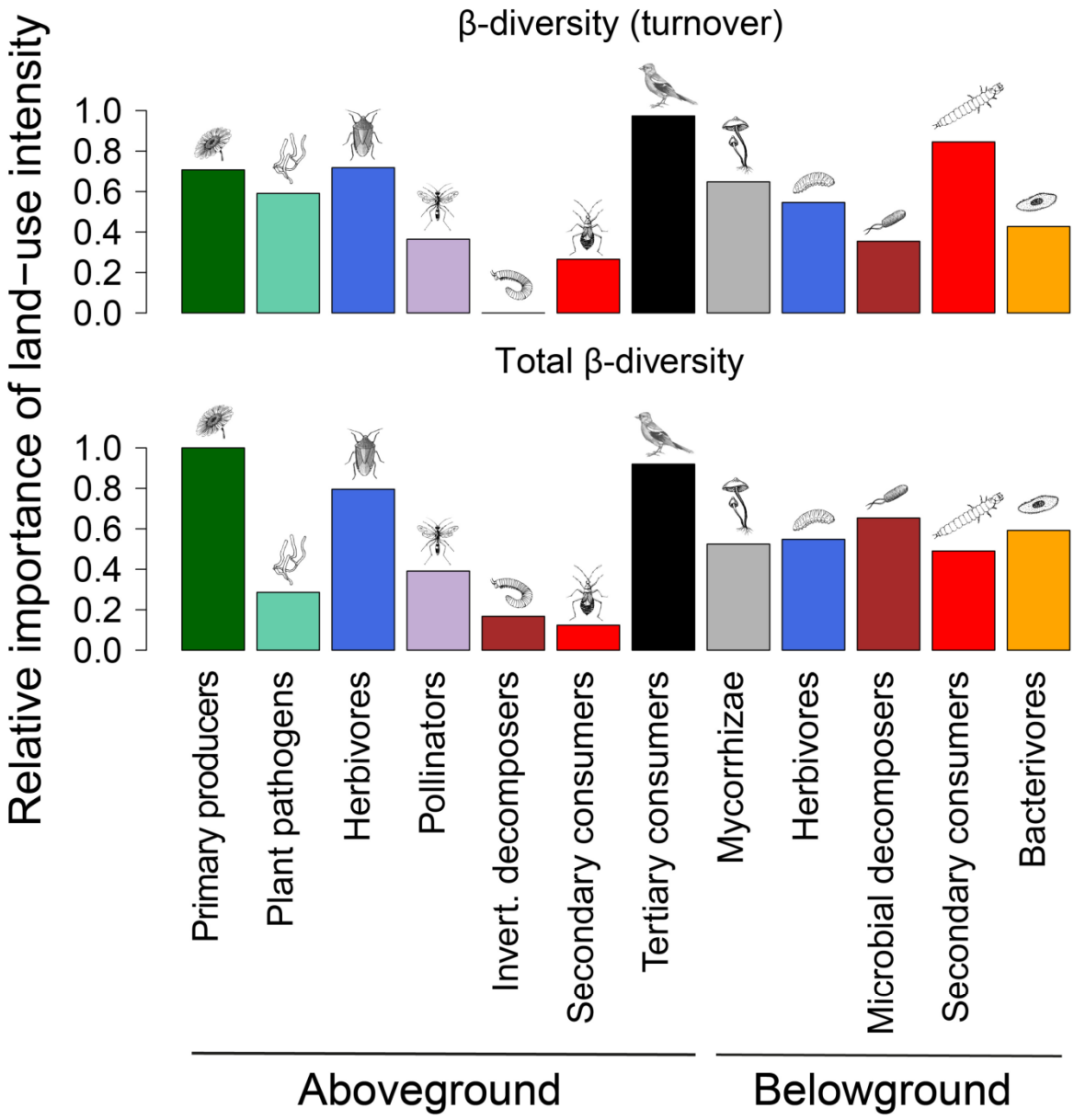


Extended Data Figure 5 | Partial effects of geographic and environmental distances and temporal variation in LUI on β-diversity above- and belowground. a, **b**, Results from GDMs are shown for seven aboveground and five belowground groups, with total β-diversity measured as the Sørensen index β_{sor} (**a**) or as the species turnover component β_{sim} (**b**). The figures show the effect of differences in each of the five variables on β-diversity (calculated between all plot pairs; $n = 105$ plots). Effects of differences in each explanatory variable can vary nonlinearly along the gradient of that variable and each is corrected for all other variables in the model. Higher maximum curves indicate larger effects of differences in a given variable on β-diversity. Soil nutrients refer to the scores of the first PCA axis. Temporal variation in LUI is shown as s.d. Geographic distance has to be multiplied by 100 km.

RESEARCH LETTER

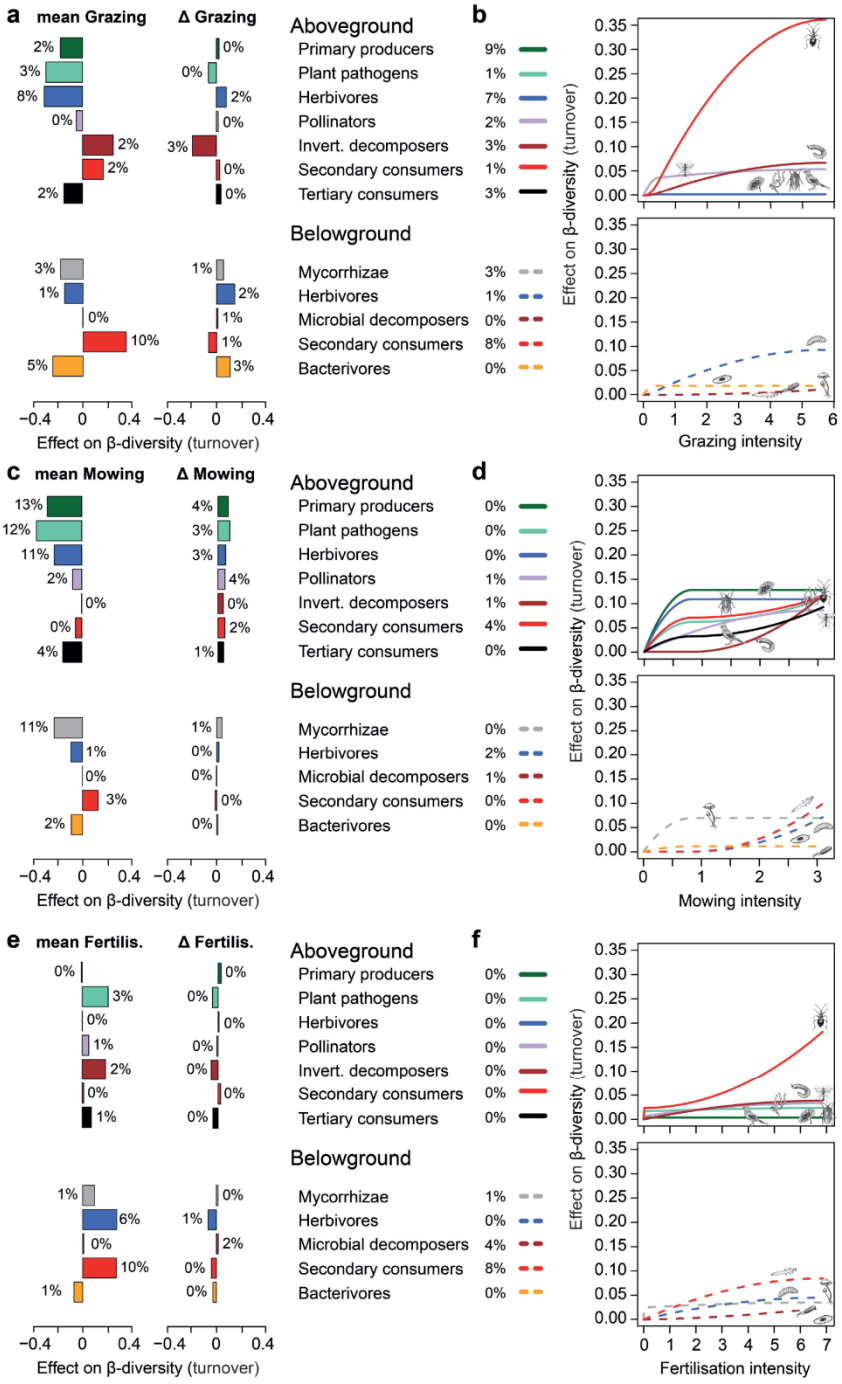


Extended Data Figure 6 | Uncertainty of effects of LUI on β-diversity above (AG) and belowground (BG). The uncertainty is calculated on the basis of 100 bootstraps for each model, each time removing 30% of the plot pairs, then fitting a GDM and extracting the predictions. Predictions are shown as fitted lines and s.d. Uncertainty is shown for all seven above- and five belowground trophic groups based on species turnover β_{sim} ($n = 105$ plots). PriPro, primary producers; PlPa, plant pathogens; Herb, herbivores; Poll, pollinators; InvDec, invertebrate decomposers; SecCon, secondary consumers; TerCon, tertiary consumers; Myco, Mycorrhizae; MicDec, microbial decomposers; Bact, bacterivores.

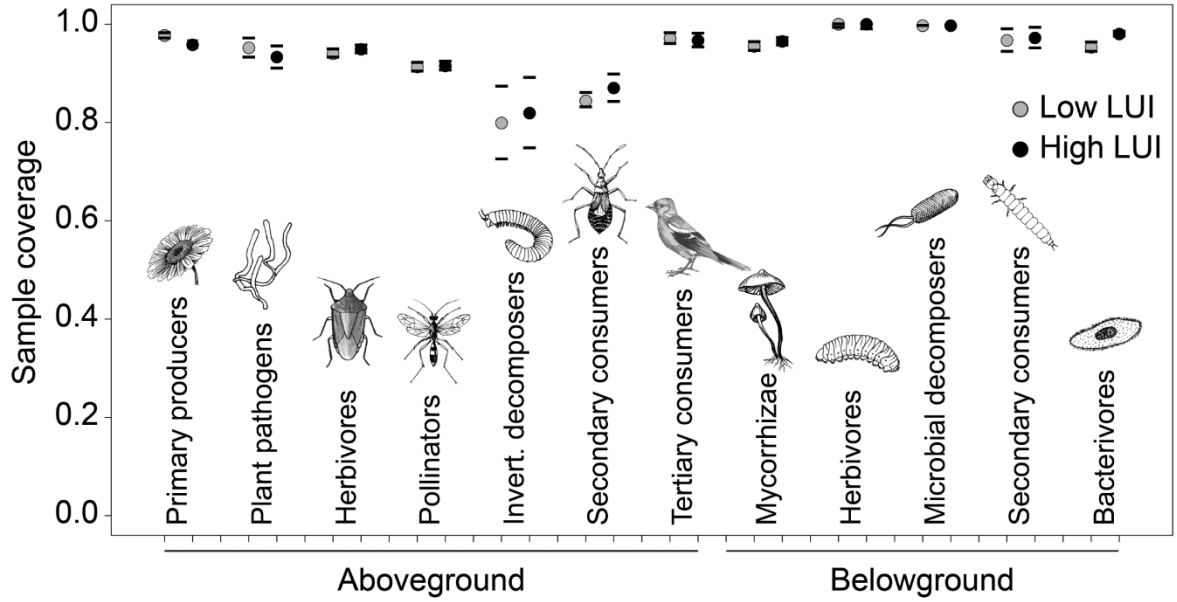


Extended Data Figure 7 | The relative importance of LUI as a driver of β-diversity. The bar plot shows the importance of LUI (in terms of total effect size) relative to the most important variable in the GDM. Results are shown for each trophic group, for the species turnover component (β_{sim}) and total β-diversity (Sørensen index) ($n = 105$ plots).

RESEARCH LETTER

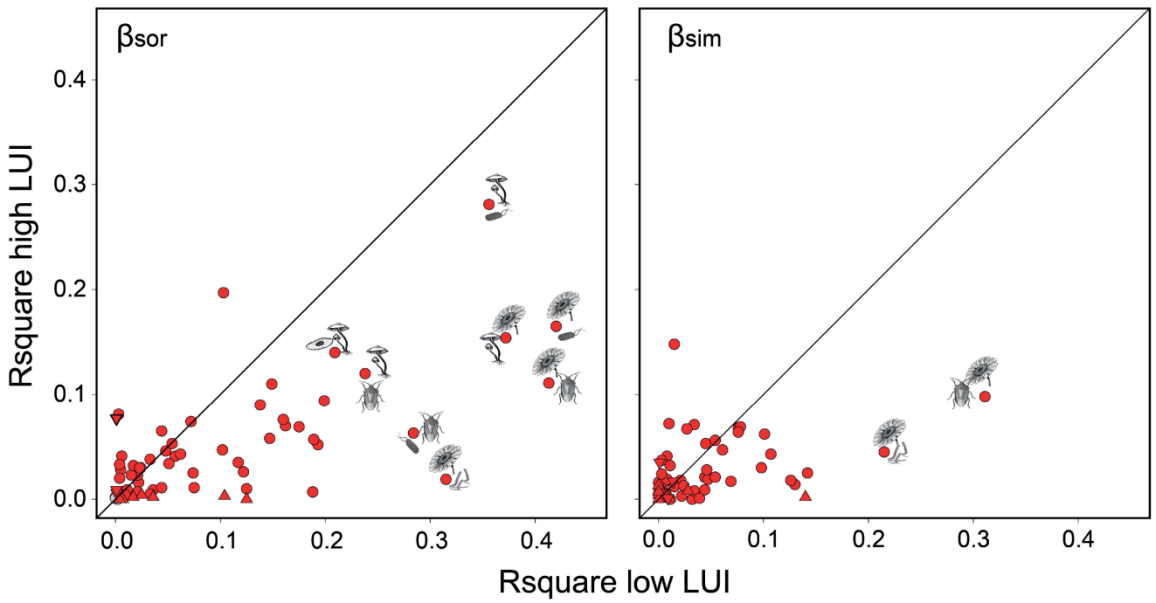


Extended Data Figure 8 | Effects of single land-use components on β-diversity above- and belowground. a, **c**, **e**, Partial effects of minimum LUI (min LUI) and ΔLUI between plot pairs ($n = 105$ plots), on the species turnover component of β-diversity (β_{sim}) for seven aboveground and five belowground groups based on linear models. Negative effects of minimum LUI indicate that land-use intensification reduces β-diversity. The bars show coefficients from the models. Numbers adjoining bars are the proportion of the total explained variance that is uniquely explained by minimum LUI or ΔLUI among plot pairs, on the basis of hierarchical partitioning. **b**, **d**, **f**, Results from GDMs are shown for the turnover component β_{sim} for the same trophic groups. The figures show the effect of ΔLUI on β-diversity (calculated between all plot pairs). Effects of ΔLUI can vary nonlinearly along the gradient of LUI. Higher maximum curves indicate larger effects of ΔLUI on β-diversity. The values in the legend are the percentage of deviance that is explained uniquely by LUI. Effects of both linear models and GDMs are corrected for other drivers of β-diversity, and response and explanatory variables are scaled to allow comparisons across trophic levels (see Methods).



Extended Data Figure 9 | Sample coverage of above- and belowground trophic groups based on species incidences. Sample coverage was calculated for low (52 plots) and high (53) LUI plots based on refs 57, 58. Coverage is defined as the proportion of the total number of individuals in an assemblage that belong to species represented in the sample, and is therefore a measure of sampling completeness. Means and 95% confidence intervals based on 200 bootstraps are shown.

RESEARCH LETTER



Extended Data Figure 10 | The effect of LUI on the correlation between the β-diversities of different trophic groups. Each dot represents the correlation (R^2) between two trophic groups. Correlations are R^2 values from matrix regressions between β-diversity values of different groups (corrected for effects of differences in LUI on β-diversity). Significant correlations ($P < 0.05$) are marked in red. Upward and downward triangles indicate significance under low or high LUI only. Interactions with R^2 values higher than 0.2 in one of the LUI-categories are illustrated by icons. β-diversity was calculated as the Sørensen index (β_{sor}, total β-diversity) and as the species turnover component (β_{sim}) ($n = 105$ plots). For statistical details see Supplementary Information Section 5.

Chapter 3 – Locally rare species influence grassland ecosystem multifuntionality

PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Research

Cite this article: Soliveres S *et al*. 2016 Locally rare species influence grassland ecosystem multifunctionality. *Phil. Trans. R. Soc. B* **371**: 20150269.
http://dx.doi.org/10.1098/rstb.2015.0269

Accepted: 27 December 2015

One contribution of 17 to a theme issue 'Biodiversity and ecosystem functioning in dynamic landscapes'.

Subject Areas:
ecology

Keywords:
biodiversity, common species, ecosystem function, identity hypothesis, land use, multitrophic

Author for correspondence:
Santiago Soliveres
e-mail: santiago.soliveres@ips.unibe.ch

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2015.0269 or via http://rstb.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Locally rare species influence grassland ecosystem multifunctionality

Santiago Soliveres[1], Peter Manning[1,2], Daniel Prati[1], Martin M. Gossner[3,4], Fabian Alt[5], Hartmut Arndt[6], Vanessa Baumgartner[7], Julia Binkenstein[8], Klaus Birkhofer[9], Stefan Blaser[1], Nico Blüthgen[10], Steffen Boch[1], Stefan Böhm[11], Carmen Börschig[12], Francois Buscot[13,14], Tim Diekötter[15], Johannes Heinze[16], Norbert Hölzel[17], Kirsten Jung[11], Valentin H. Klaus[17], Alexandra-Maria Klein[18], Till Kleinebecker[17], Sandra Klemmer[13], Jochen Krauss[19], Markus Lange[3,20], E. Kathryn Morris[21,22], Jörg Müller[16], Yvonne Oelmann[5], Jörg Overmann[7], Esther Pašalić[3], Swen C. Renner[23,24], Matthias C. Rillig[22,25], H. Martin Schaefer[26], Michael Schloter[27], Barbara Schmitt[1], Ingo Schöning[20], Marion Schrumpf[20], Johannes Sikorski[7], Stephanie A. Socher[28], Emily F. Solly[20], Ilja Sonnemann[22], Elisabeth Sorkau[5], Juliane Steckel[19], Ingolf Steffan-Dewenter[19], Barbara Stempfhuber[27], Marco Tschapka[11], Manfred Türke[14,29], Paul Venter[6], Christiane N. Weiner[19], Wolfgang W. Weisser[3,4], Michael Werner[19], Catrin Westphal[12], Wolfgang Wilcke[30], Volkmar Wolters[31], Tesfaye Wubet[14], Susanne Wurst[32], Markus Fischer[1,2] and Eric Allan[1]

[1]Institute of Plant Sciences, University of Bern, Altenbergrain 21, Bern 3013, Switzerland
[2]Senckenberg Gesellschaft für Naturforschung, Biodiversity and Climate Research Centre BIK-F, Senckenberganlage 25, Frankfurt 60325, Germany
[3]Institute of Ecology, Friedrich-Schiller-University Jena, Dornburger Straße 159, Jena 07743, Germany
[4]Terrestrial Ecology Research Group, Department of Ecology and Ecosystem Management, School of Life Sciences Weihenstephan, Technische Universität München, Hans-Carl-von-Carlowitz-Platz 2, Freising 85354, Germany
[5]Geocology, University of Tuebingen, Ruemelinstr. 19-23, Tuebingen 72070, Germany
[6]Department of General Ecology, Institute for Zoology, University of Cologne, Cologne 50674, Germany
[7]Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffenstr. 7B, Braunschweig 38124, Germany
[8]Institute for Biology 1, Albert Ludwigs-University Freiburg, Hauptstr. 1, Freiburg 79104, Germany
[9]Department of Biology, Lund University, Lund, Sweden
[10]Ecological Networks, Biology, Technische Universität Darmstadt, Schnittspahnstr. 3, Darmstadt 64287, Germany
[11]Institute of Experimental Ecology, University of Ulm, Albert-Einstein-Allee 11, Ulm 89069, Germany
[12]Agroecology, Department of Crop Sciences, Georg-August University of Göttingen, Grisebachstr. 6, Göttingen 37077, Germany
[13]Department of Soil Ecology, UFZ-Helmholtz Centre for Environmental Research, Theodor-Lieser-Straße 4, Halle (Saale) 06120, Germany
[14]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, Leipzig 04103, Germany
[15]Department of Landscape Ecology, Kiel University, Kiel, Germany
[16]Biodiversity Research/Systematic Botany, University of Potsdam, Maulbeerallee 1, Potsdam 14469, Germany
[17]Institute of Landscape Ecology, University of Münster, Heisenbergstr. 2, Münster 48149, Germany
[18]Nature Conservation and Landscape Ecology, University of Freiburg, Freiburg, Germany
[19]Department of Animal Ecology and Tropical Biology, Biocentre, University of Würzburg, Am Hubland, Würzburg 97074, Germany
[20]Max-Planck Institute for Biogeochemistry, Hans-Knoell-Str. 10, Jena 07745, Germany
[21]Department of Biology, Xavier University, 3800 Victory Parkway, Cincinnati, OH 45207, USA
[22]Institut für Biologie Funktionelle Biodiversität, Freie Universität Berlin, Königin-Luise-Str. 1-3, Berlin 14195, Germany
[23]Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, VA 22630, USA

[24]Institute of Zoology, University of Natural Resources and Life Science, Gregor-Mendel-Straße 33, 1180 Vienna, Austria
[25]Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Berlin 14195, Germany
[26]Department of Ecology and Evolutionary Biology, Faculty of Biology, University of Freiburg, Hauptstraße 1, Freiburg i. Br 79104, Germany
[27]Research Unit for Environmental Genomics, Helmholtz Zentrum München, Ingolstädter Landstr. 1, Oberschleissheim 85758, Germany
[28]Department of Ecology and Evolution, Universität Salzburg, Kapitelgasse, Salzburg 4-65020, Austria
[29]Institute for Biology, Leipzig University, Johannisallee 21, Leipzig 04103, Germany
[30]Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), Reinhard-Baumeister-Platz 1, Karlsruhe 76131, Germany
[31]Department of Animal Ecology, Justus-Liebig-University Giessen, Gießen, Germany
[32]Institute of Biology, Functional Biodiversity, Freie Universität Berlin, Königin-Luise-Str. 1-3, Berlin 14195, Germany

SS, 0000-0001-9661-7192

Species diversity promotes the delivery of multiple ecosystem functions (multifunctionality). However, the relative functional importance of rare and common species in driving the biodiversity–multifunctionality relationship remains unknown. We studied the relationship between the diversity of rare and common species (according to their local abundances and across nine different trophic groups), and multifunctionality indices derived from 14 ecosystem functions on 150 grasslands across a land-use intensity (LUI) gradient. The diversity of above- and below-ground rare species had opposite effects, with rare above-ground species being associated with high levels of multifunctionality, probably because their effects on different functions did not trade off against each other. Conversely, common species were only related to average, not high, levels of multifunctionality, and their functional effects declined with LUI. Apart from the community-level effects of diversity, we found significant positive associations between the abundance of individual species and multifunctionality in 6% of the species tested. Species-specific functional effects were best predicted by their response to LUI: species that declined in abundance with land use intensification were those associated with higher levels of multifunctionality. Our results highlight the importance of rare species for ecosystem multifunctionality and help guiding future conservation priorities.

1. Introduction

Many studies have demonstrated that high species diversity enhances ecosystem functioning both in experimental and natural assemblages ([1–4]; reviewed in [5] this issue). However, it has been argued elsewhere that it is not the total number of species *per se*, but the functional properties of the most locally abundant ones (hereafter common species) that drive ecosystem functioning (mass–ratio hypothesis [6]). Other work has shown that each common species can only provide a limited number of functions [1,7–8]. Extending the mass–ratio hypothesis to the simultaneous provision of multiple ecosystem functions at high levels (multifunctionality), we might, therefore, predict that several common species would be needed to maintain multifunctionality and that the diversity of common species, rather than overall diversity, would be its main biotic driver. In contrast with this argument, less locally abundant (hereafter rare) species have been shown to play a crucial role in affecting several ecosystem functions [9–11]. Rare species comprise the vast majority of the species in any natural community and are more sensitive to anthropogenic disturbances [12,13]. Thus, quantifying the functional consequences of their loss is of particular importance to predict the provision of ecosystem services in the future. The functional importance of common versus rare species could depend on the ecosystem functions under scrutiny. Studies focused on productivity and pollination have found common species to be the main driver (e.g. [9,14–17]) while those focusing on functions associated with some regulating (e.g. invasion resistance) or recreational (e.g. birdwatching) services highlight the importance of rare species [18–20]. Owing to their contrasting effects depending on the function considered, studies measuring multifunctionality are required to comprehensively assess the relative functional importance of rare and common species [8,11].

Studies across large temporal or spatial scales have shown that the relationship between diversity and ecosystem functioning may change with abiotic conditions or land use intensification [21–25], the level of multifunctionality desired [8,26], or the type of organism being considered [23]. Land use intensification promotes shifts in the functional composition of multiple taxa (e.g. [27]), potentially dampening the generally positive relationship between diversity and ecosystem multifunctionality [25]. The effects of biodiversity might also depend on the level of multifunctionality considered [26,28]. High levels of many functions can be difficult to achieve if there are strong trade-offs between functions or between diversity effects on these functions. Finally, different components of biodiversity may differ in their functional effects. Above- and below-ground organisms differ in their sensitivity to climate or anthropogenic disturbances, with rare above-ground species being the most sensitive (e.g. [13,29]), and can also have different effects on ecosystem multifunctionality, with stronger effects found for above-ground organisms [23,30]. The context dependencies of the biodiversity–functioning relationship are poorly understood, particularly in terms of how they might modify effects of rare and common species. Existing comparisons of the functional role of rare versus common species have seldom been extended beyond single taxa, individual ecosystem functions, or a particular study site (but see [11]). In order to understand the response of natural and semi-natural ecosystems to ongoing global change, we therefore need to examine the relationships between different components of diversity (above- versus below-ground, common versus rare) and ecosystem multifunctionality across environmental gradients [23].

It has also been hypothesized that the presence of certain species can be of particular importance for ecosystem functioning, regardless of their abundance or whether they are above- or below-ground organisms (identity hypothesis [31,32]). This hypothesis has received empirical support from studies focusing on individual functions such as litter decomposition, parasitism or predation [33–35]. However, studies have not yet tested whether there are species that can drive overall ecosystem multifunctionality, which would require lack of trade-offs in their effects on different functions. If there are influential species, it is important to understand the characteristics that they possess and how they respond to land use intensification. If such species decline in abundance as land

rstb.royalsocietypublishing.org Phil. Trans. R. Soc. B 371: 20150269

use intensifies then, in addition to effects of biodiversity loss, compositional change driven by land use may have large effects on ecosystem multifunctionality.

Here, we assess the functional role of the diversity of rare and common species (based on their local abundance), both above- and below-ground, on several multifunctionality indices derived from 14 ecosystem functions, related to the delivery of supporting, provisioning, regulating and cultural services (*sensu* [36]). Our hypotheses are: (i) the diversity of common species is a more important driver of ecosystem multifunctionality than the diversity of rare species; (ii) the positive effect of diversity on multifunctionality will decline with land-use intensity (LUI) due to the associated changes in functional composition; (iii) the diversity of above-ground organisms is the strongest biotic predictor for ecosystem multifunctionality [23]; (iv) there are particular species, across multiple trophic levels, that can promote high multifunctionality, and land use intensification changes the abundance of these species.

2. Material and methods

(a) Study sites

We sampled 150 grassland plots (50 × 50 m) equally distributed over three regions across Germany: the UNESCO Biosphere Reserves Schwäbische Alb (located in the southwest) and Schorfheide-Chorin (in the northeast), and the area in and around the National Park Hainich-Dün (in central Germany; [37]). The 50 grassland plots per region were selected to span a gradient of the full range of land use practices and intensities found in Central European grasslands. Information about LUI was obtained directly from the land owners via questionnaires [37]. We used this information to calculate a compound measure of LUI which summarizes the three major components of land use in these grasslands—intensity of fertilization, mowing and grazing—with the following formula: $\mathrm{LUI} = \mathrm{sqrt}((F_i/F_R) + (M_i/M_R) + (G_i/G_R))$, where F_i, M_i and G_i are the amount of fertilizer applied, frequency of mowing cuts and standardized units of livestock density within each sampling site per year, respectively. All three components were standardized by the average across the 50 grassland plots within each region (F_R, M_R and G_R; see [38] for full methodological details). We averaged LUI across 2006–2010, the period when most diversity and functioning data were collected.

(b) Analyses at the community level

(i) Diversity measures

At each site, we measured the abundance and richness of nine trophic groups using standard methodology (see electronic supplementary material, table S1 for details). Overall, our sampling included approximately 4300 taxa (the taxonomic unit varied between groups, electronic supplementary material, table S1, but we refer to all as 'species', for simplicity). The groups were autotrophs (plants and bryophytes), below-ground herbivores (insect larvae), below-ground predators (insect larvae), detritivores (insects and millipedes), soil microbial decomposers (bacteria), above-ground herbivores (insects), above-ground predators (insects, spiders and centipedes), arbuscular mycorrhizal fungi and below-ground bacterivores (bacterivorous protists). Omnivorous arthropods were not considered in our analyses as there were too few of them.

Using data for each of these nine trophic groups we calculated multidiversity, i.e. a measure of overall diversity at the community level obtained by averaging standardized diversity measures across trophic groups [13]. To calculate multidiversity, we first classified the species into two groups according to their abundance (which was measured differently for the various groups; electronic supplementary material, table S1): common (the top 10% of species in terms of total abundance) and rare species (the bottom 90% of species). Abundance is widely accepted as a measure of rarity (e.g. [39]); therefore, we chose abundance across all study sites to be the most representative measure of the overall rarity of our target species. The top 10% species (common species hereafter) accounted for 80% of the total abundance sampled, whereas the bottom 90% of species (rare species hereafter) made on average 20% of the total abundance (ranging from 6% in bacteria to 30% in below-ground herbivores; electronic supplementary material, figure S1). A second step in the calculation of our multidiversity metric was to standardize all variables to a common scale (between 0 and 1) by subtracting the minimum value and dividing by the maximum value found across the 150 sites to avoid the influence of different ranges in diversity characterizing each group. Third, we classified the trophic groups into above- and below-ground organisms (plants were considered above-ground organisms). Finally, we averaged their standardized values to obtain four measures of multidiversity: above- and below-ground common species multidiversity, and above- and below-ground rare species multidiversity.

(ii) Ecosystem function measures

At each site, we measured 14 different ecosystem functions. These were: above- and below-ground plant biomass, root decomposition rates, potential nitrification, soil phosphorus retention, arbuscular mycorrhizal fungal root colonization, stability of soil aggregates, soil organic carbon, forage quality, resistance to above-ground plant pathogens, above-ground pest control, pollinator abundance, bird diversity and flower cover (see [25] and electronic supplementary material, table S2 for detailed methodology). These ecosystem functions are related to nutrient cycling, food provision, sustainable soil use, pest resistance, or cultural and recreational services. We calculated three ecosystem multifunctionality metrics using these 14 functions and following the multiple threshold approach of Byrnes *et al*. [26], which sums up the number of measured functions that exceed a given threshold. These thresholds are defined as a given percentage of the maximum level found for each function, and we used three thresholds (50, 75 and 90%) to represent a wide spectrum. In order to reduce the influence of outliers, the maximum was defined as an average of the top five values for each function across our study sites.

(iii) Statistical analyses

We used multimodel inference based on information theory [40] to analyse the response of ecosystem multifunctionality to the multidiversity of above- and below-ground common and rare species. We performed a different analysis for each of the three multifunctionality metrics. Large-scale studies such as ours allow the quantifying of the relative importance of diversity regarding other drivers of ecosystem functioning, and the evaluation of changes in diversity–functioning relationships across contrasting environmental conditions. However, it is difficult from observational studies to infer causality as diversity–functioning relationships could be confounded by environmental factors affecting both diversity and ecosystem functioning. To avoid the latter, we controlled for factors that could affect both multidiversity and ecosystem functioning in our analyses; these were: study region, environmental variables (pH, soil depth and topography—an index based upon the position and steepness of each site, which is related to the accumulation of soil material and water availability [41,42]), and LUI. We removed elevation from the set of environmental predictors because it was highly correlated with soil depth (Spearman's rank correlation, $\rho = -0.91$). We also accounted for potential context dependencies in the diversity–

multifunctionality relationship by including the interactions between LUI, region and the four multidiversity predictors.

To analyse the relative importance of environmental conditions, the multidiversity of above- and below-ground common and rare species, and the interactions between them, as drivers of ecosystem multifunctionality, we built a set of competing models including either: environmental variables only, environmental + diversity variables; or environmental + diversity variables and the interactions between diversity and region and/or LUI; see electronic supplementary material, table S3, for the full list of models). From these competing models, we selected those that best fitted our data according to the Akaike information criterion (AICc, corrected for small sample sizes). Thus, those models differing less than 2 AICc units from the most parsimonious model (ΔAICc < 2) were included in the set of best-fitting models. We also calculated the importance of our different predictors as the sum of the AICc weights (a comparison with each model's AICc with the minimum AICc) of the models in which each predictor appears. To allow comparisons between main effects and interaction terms, we divided the importance of each predictor by the number of models in which it was included (16 for the diversity predictors (maximum importance 1/16), and 8 for their interactions with region and LUI (maximum importance 1/8); see [43] for a related approach).

As a sensitivity analysis, we repeated our multimodel selection but used the multidiversity of the bottom 50% of species, instead of the bottom 90%, as an alternative measure of rarity. These bottom 50% of species made up on average 3% of the total abundance (ranging from 0.04% in bacteria to 6% in below-ground herbivores; electronic supplementary material, figure S1 and table S4 for detailed results). We also repeated our analyses using the abundance, instead of the species richness, of above- and below-ground common and rare species (electronic supplementary material, table S5). Results of sensitivity analysis were broadly similar to the main ones and therefore are not further discussed.

(c) Analyses at the species level

(i) Selection of species

We selected a subset of individual species that occurred in all three study areas and in at least 10 of the 150 sites to obtain reliable parameter estimates (see *Estimation of the functional role of each species* below). Some of the trophic groups measured (detritivores, and below-ground herbivores and predators) were not included in these species-level analyses as they contained too few species fulfilling our selection criteria. Of those that did, soil microbial decomposers and bacterivorous protists were overrepresented. Thus, in order to obtain a balanced sampling size for each trophic group, we only selected the most and least abundant 25 species within each trophic group that met the criteria. These species roughly corresponded to those classified as common and rare in the community level analyses (electronic supplementary material, table S6). Thereby, we obtained a balanced sample size of 50 species per trophic group (approx. 50% of them common, approx. 50% rare), with the exception of above-ground predators, for which only 20 species met our criteria (270 species considered overall).

(ii) Estimation of the multifunctional role of each species

To estimate the multifunctional role of each species, we used the null-model approach of Gotelli *et al*. [44] as implemented by the software Impact [45]. This analysis allowed us to identify the presence of influential species, and whether or not the degree of functional influence was related to the average species' abundance, to their functional traits, or to their response to LUI. The latter allowed us to test for the effects of compositional changes, across trophic levels, driven by land use intensification on multifunctionality. This analysis further allows us to identify whether there are functional trade-offs within each trophic level; i.e. whether some species within a group are significantly associated with multifunctionality. The null-model approach used performs linear regressions between the abundance of each species and a given function, and then compares the observed slope with 1000 random permutations of the values of the functional variable. From the randomizations, a standardized effect size (SES) for each species is calculated as: $\mathrm{SES} = (S_{\mathrm{obs}} - S_{\mathrm{sim}})/\mathrm{s.d.}$; where S_{obs} and S_{sim} are the observed and the average of the 1000 simulated regression slopes, respectively, and s.d. is the standardized deviation of the slopes obtained from these 1000 randomizations. SES values higher than 2 or lower than -2 show significant relationships between the abundance of a given species and the function used as a response. We used our three multifunctionality measures as a response and thus obtained three functional effect sizes for each of our target species (270 target species $\times$ 3 multifunctionality scenarios = 810 comparisons). Owing to the increased type II error derived from multiple testing, 40 of these 810 comparisons would be expected to be significant only by chance; we found two times more significant results (electronic supplementary material, table S7), implying that our results were unlikely to be caused by multiple testing only.

By randomizing the functional variable instead of the species abundances, the null-model approach takes partially into account the structure of the biotic community (including species interactions and non-independent effects; see full discussion in [44]). However, this method is purely correlative and prone to confounding factors which could be driving both the function and the abundance of the target species. To control for the latter, we used the residuals of both the abundance of each species and the multifunctionality metrics after filtering for the same environmental variables used in the multimodel selection (region, LUI, soil pH and depth, elevation, and topography). Despite its limitations, this is to our knowledge the best method available to estimate the functional effects of many species (which would be logistically prohibitive to address experimentally).

(iii) Statistical analyses

We performed two complementary analyses at the species level. First, to assess the functional importance of above- and below-ground groups, and of rare and common species, we compared the number of significant positive and negative SES in each of the four categories of species, using Fisher's exact tests (better suited for low sample sizes than χ^2 analyses). This allowed us to assess if there are influential species related to multifunctionality, and if these are either common or rare, or mainly represented by above- or below-ground organisms. The second analysis aimed to understand further which features make a species influential for multifunctionality. To do this we performed multiple regressions with the SES of the functional effect of each species as a response variable, and the 'response to LUI' as a predictor. To correct for other species characteristics that might affect their multifunctional importance, we also included their average abundance (across all sites in which each species occurred) and functional traits (plant height and specific leaf area for plants, and body size for herbivores and predators; obtained from available databases [46,47]). 'Response to land use intensity' was the standardized coefficient of a linear regression between each species' abundance and LUI. The number of sites in which each species occurred (which was correlated also with the range in abundance values; $\rho = 0.45$) was introduced as covariate in our analyses as it could affect slope estimates in the null-model approach used. The traits selected are related to species responses to LUI [25,48,49] and also play an important role for ecosystem functioning [25,50]. Data on functional traits were not available for microbial decomposers, bacterivorous protists and symbionts. Context dependencies in the species-level analysis were accounted for by including the interaction between region, trophic group and abundance or response to LUI as extra predictors. The interactions with region were not

significant in any case and, therefore, they are not considered further. All analyses were performed using R v. 3.0.2 [51].

3. Results

(a) Community level

Between 10 and 18% of the variation in multifunctionality was explained by study region, environmental variables, LUI and our multidiversity metrics (figure 1). All the best models (those with $\Delta AIC < 2$) included at least one of the four multidiversity metrics, with models only including environment and LUI performing less well (ΔAIC between 2.6 and 7.1; electronic supplementary material, table S3). The effects of multidiversity on multifunctionality differed depending on the metrics considered. Above-ground multidiversity of common species was not significantly related to any of the multifunctionality measures, whereas the multidiversity of common species below ground was positively related to multifunctionality at the 50% of threshold, but not to the other multifunctionality measures (electronic supplementary material, table S3). The multidiversity of rare species both above and below ground was significantly, but oppositely (positive for above and negative for below ground), related to multifunctionality at the highest thresholds (75 and 90%; figures 1 and 2).

The relationships found between multidiversity (both above and below ground) and multifunctionality at the highest thresholds did not depend on LUI or study region (figure 1). The best models for both the 75 and 90% thresholds did not include interactions between region and/or LUI and multidiversity (electronic supplementary material, table S3). The best models for the 50% multifunctionality threshold, however, included interactions between region, and/or LUI, and one or more multidiversity metrics (figure 1; electronic supplementary material, table S3), thus demonstrating that multidiversity–multifunctionality relationships were context-dependent for the low threshold measure. Indeed, for multifunctionality at the 50% threshold the interactions were as important as the main effects (figure 1), and not including them increased the AICc by more than 3 units in all cases, suggesting a strong decline in model performance (electronic supplementary material, table S3). Interactions with region or LUI affected the associations between below-ground, but not above-ground, multidiversity and multifunctionality. The association between the multidiversity of below-ground common species and multifunctionality was positive in the southwest, neutral in the central region and negative in the northeast (see interaction coefficients in electronic supplementary material, figure S2). LUI also influenced the effect of below-ground multidiversity, with associations between the multidiversity of both rare and common species and multifunctionality becoming more positive with decreasing LUI (electronic supplementary material, figure S2). Regardless of the interactions with region and LUI, we found a higher importance of above-ground multidiversity for the 75% and 90% thresholds, which shifted towards a higher importance of below-ground multidiversity components at the 50% threshold (figure 1; electronic supplementary material, table S3).

(b) Species level

Apart from the community level effects of multidiversity, we found significant positive associations between the abundance

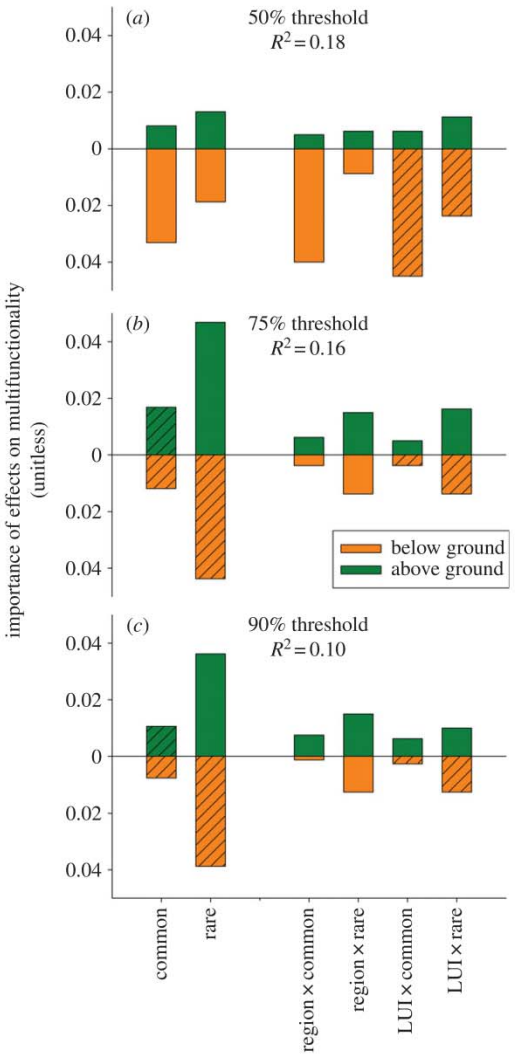


Figure 1. Importance of the different multifunctionality predictors as assessed by the sum of the AIC weights of the models in which each one was retained, divided by the number of models in which each variable was introduced. Green and brown indicate predictors associated with above- and below-ground multidiversity, which included the overall diversity of plants, bryophytes and herbivore, carnivore and decomposer arthropods (above-ground) and soil bacteria, bacterivore protists, arbuscular mycorrhizal fungi, and herbivore and carnivore insects (below-ground). Hatched bars indicate negative effects. Common = multidiversity of the top 10% most abundant species (80% of the individuals sampled), rare = multidiversity of the 90% least abundant species (20% of the individuals sampled). Region × indicates the interaction term between study region and a given multidiversity metric. LUI × indicates the interaction term between land use intensity and a given multidiversity metric. The R^2 of the best model for each multifunctionality metric (first row in electronic supplementary material, table S3) is provided.

of individual species and multifunctionality in 6% of the species tested, whereas we found negative relationships for 4% of the species (electronic supplementary material, table S7). Above-ground rare species had significantly more positive, and fewer negative, relationships with multifunctionality than the above-ground common species did (figure 3), a trend not found in below-ground organisms. The ratio between positive

rstb.royalsocietypublishing.org Phil. Trans. R. Soc. B 371: 20150269

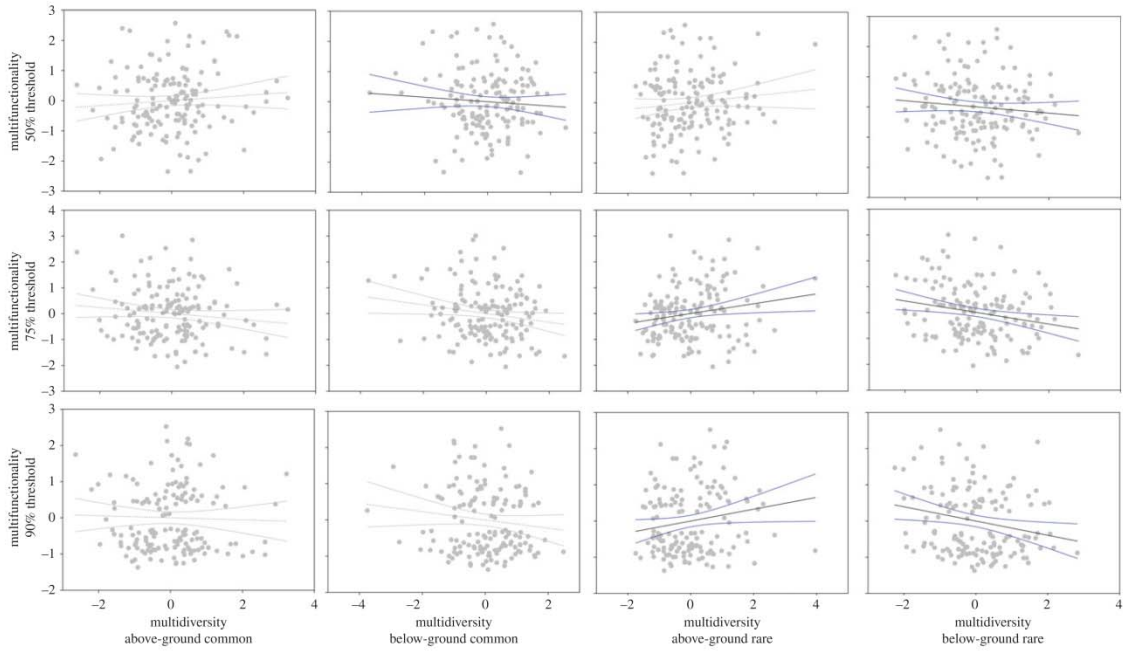


Figure 2. Effect of multidiversity of above- and below-ground, common and rare species on the different levels of multifunctionality. Slopes (with confidence intervals) were calculated after controlling for the other predictors in the model and are shown in blue if selected in the best models (see electronic supplementary material, table S3 and figure S2). Note that dots are residuals of both multidiversity and multifunctionality metrics after filtering by study region, LUI, soil pH and depth, and the topographic wetness index. (Online version in colour.)

and negative relationships differed substantially depending on the trophic group studied. Microbial decomposers had more positive than negative relationships with multifunctionality (11% versus 1%), with the opposite pattern observed in symbionts (4% versus 10%, electronic supplementary material, table S7). The remaining trophic groups showed slightly more positive than negative relationships.

Including the abundance, functional traits and response to LUI of the individual species allowed us to predict 13–16% of the variance in the strength of species–multifunctionality associations. The multiple regressions performed revealed that response to LUI was the strongest predictor of the associations between individual species abundance and multifunctionality. Species that increased in abundance in response to LUI were negatively correlated with multifunctionality at the 75 and 90% thresholds (figure 4; electronic supplementary material, table S8).

4. Discussion

(a) Effects of the multidiversity of rare and common species on multifunctionality

The important role that rare species play in maintaining individual ecosystem functions and, to some extent, multifunctionality has been highlighted previously [8,11,18–20]. Here, we extend those results to multitrophic assemblages in realistic landscapes, and show that (i) the relative importance of rare species increases when multifunctionality is defined using higher thresholds for the functions; and (ii) that this relationship remains relatively consistent across study regions and land use intensities (figure 2). Our results show that the ability of ecosystems to maintain a large number of functions at average levels (50% threshold) is mainly driven by the diversity of common species and, intuitively, by the prevailing environmental conditions (as shown by the significant interactions found in our models). However, the delivery of a smaller number of functions, but at very high levels (75 and 90% thresholds), was mainly related to the multidiversity of rare species. The level of multifunctionality required will depend on stakeholder preferences, but the performance of many functions at their highest potential (high multifunctionality values at high thresholds) can be generally interpreted as a more desirable state of natural ecosystems. Overall, our study shows that the diversity of rare species is consistently and positively related to multifunctionality at the highest levels, thus implying the existence of 'win-win' scenarios between biodiversity conservation and ecosystem service provision.

A high diversity of rare species might be more beneficial for multifunctionality than a high diversity of common species if rare species are less likely to negatively affect ecosystem functions. We found that functional trade-offs between species, where some species have positive effects on multifunctionality and others negative effects, were less common among rare than among common species (figure 3). This could explain the stronger positive effect of rare species diversity on multifunctionality. Our correlative study does not allow us to investigate the mechanisms behind the lower incidence of such functional trade-offs in rare species. However, we speculate that if functional effects are driven by the presence of a given species, rather than by its abundance, they are much less likely to be negative. In the case of the presence-based functional effects, the species is either there and promotes a given function, or is absent and has no effect. For example, the presence of certain species can promote recreational services such as birdwatching [20], or prevent plant invasions [18,19]). Such presence-based effects are likely to be the dominant ones

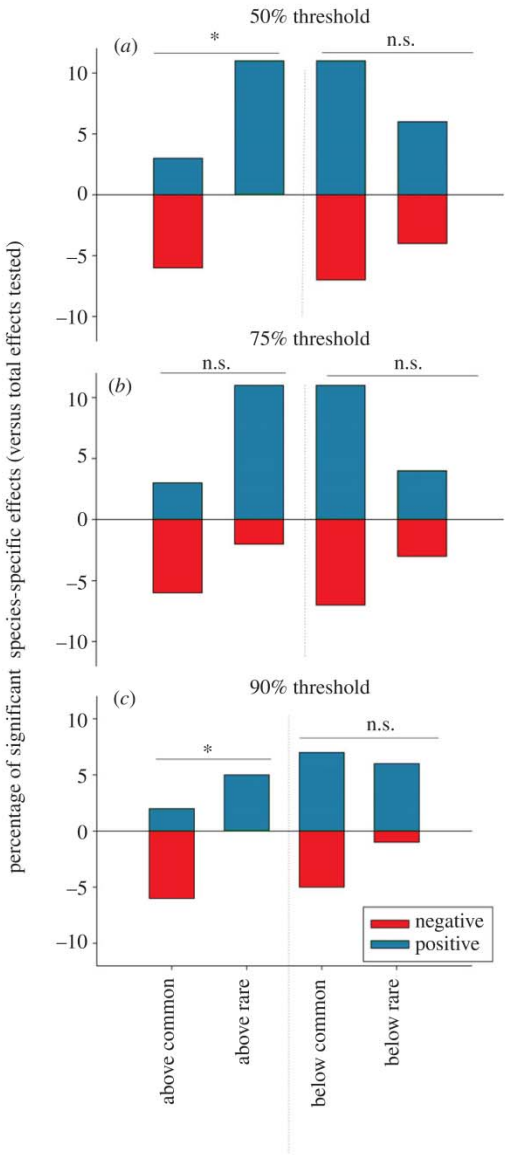


Figure 3. Summary of the relationships between individual species and multifunctionality. The percentage (according to the number of species tested) of significant positive (blue) and negative (red) effects are shown. The averaged results across each category (common versus rare species, above and below ground) are shown. Significant differences ($p < 0.05$) comparing the proportion of positive and negative effects in common and rare species according to Fisher's exact test are highlighted as '*'. n.s., not significant.

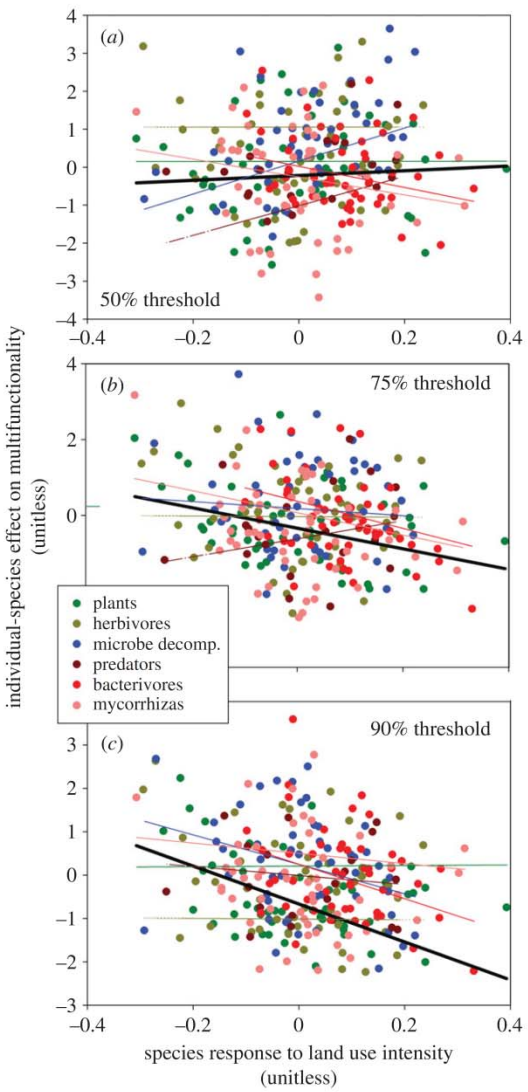


Figure 4. Relationship between the effect of each species (dots) on multifunctionality (standardized effect size) and its response to land use intensity (LUI). Different colours in dots and lines indicate the relationship found for each trophic group (slopes calculated after filtering by number of sites and average abundance). The black lines indicate the overall relationship (after filtering for the same factors and trophic group). Response to LUI was measured as the standardized slope of a regression between LUI and the abundance of each species.

among rare species. By contrast, functional effects that are proportional to a species' abundance [6] may also be negative. Abundant species can reduce, rather than promote, a given function, e.g. species with low specific leaf area can reduce rates of nutrient cycling. Common species are more likely than rare species to have such abundance-related effects. Therefore, even if the common species are functionally relevant, they may have opposing functional effects (e.g. the positive effects of a common productive plant on forage production might be balanced by negative effects of a common herbivorous insect that feeds upon it). Such strong functional trade-offs between common species could therefore result in a small effect of common species diversity on multifunctionality and a greater importance of rare species diversity in promoting multifunctionality. A complementary explanation for the higher functional importance of rare species is that they tend to be less redundant than common species in the functional traits they possess and, therefore, support communities with more distinct combinations of functional traits [52]. This enhanced functional diversity could also explain the positive functional effect of the multidiversity of rare species, as functional diversity is related to the provision and stability of multiple ecosystem functions [53,54].

While our study provides unique insights regarding the role of community level diversity on the provision of multiple

functions simultaneously, the use of these aggregate metrics obscures detailed information regarding the relationships between specific taxa and functions. A full description of such functions is outside the scope of this study but could partially explain the relatively low R^2 of our models (less than 0.20%; see [55] for a full discussion on the topic). The trade-offs we observed in the functional effects of common species would, obviously, not apply when studying ecosystem functions in isolation. Another reason for the relatively low proportion of explained variance could be the influence of factors operating at large spatial (i.e. surrounding landscape) and temporal scales (i.e. legacy effects of past land uses), which were not considered in this study [56].

(b) On the functional role of above- versus below-ground multidiversity

Below- and above-ground biotic components are known to respond differently to anthropogenic disturbances and are likely to differ in their effects on ecosystem functioning (e.g. [13,30]); however, very few studies have explored their separate functional roles [23]. We found that above-ground multidiversity, particularly of rare species, was often positively related to multifunctionality at the highest levels, whereas below-ground multidiversity was negatively associated with it. Above-ground rare species are highly sensitive to anthropogenic disturbances [13,29] and these findings suggest that they are also amongst the most functionally important species. Our results support the crucial role of the diversity of above-ground organisms, e.g. plants [1–4,8], but also herbivores [28] or predators [57], in determining ecosystem multifunctionality.

The stronger positive relationship between above- compared to below-ground diversity with multifunctionality concurs with the only previous study including these two groups separately [23]. It may be argued that the higher importance of above- compared to below-ground components is dictated by the selection of ecosystem functions studied; however, this is unlikely as both our study and Jing *et al.* [23] included a high proportion of soil-related variables. It seems that, when considered alone, below-ground diversity explains variation in multifunctionality that could be mainly due to its correlation with above-ground diversity [23,57], but further studies are required to test whether the pattern we observed holds across a wide variety of ecosystems and environments. Importantly, our snapshot sampling design may have reduced our capacity to compare the effects of both above- and below-ground multidiversity, as below-ground organisms are less sensitive to anthropogenic disturbances [13] and thus they could increase stability in ecosystem functioning by increasing response diversity [58].

The negative relationship between below-ground diversity and ecosystem multifunctionality, however, is surprising and contrasts with previous research (e.g. [23,59,60]). Soil biota effects are often driven more by functional composition than by species richness *per se* (see [59] for a review). Hence, the negative relationship between below-ground multidiversity and ecosystem functioning could reflect compositional changes rather than diversity effects [23,24]. Another potential explanation for these results is that the functional effects of below-ground diversity are context-dependent and change with climate or soil (regional differences in our study sites [61]), or with land use intensification (electronic supplementary material, figure S2; see also [23]). The latter could obscure the overall effect of below-ground multidiversity on ecosystem functioning when it is investigated across wide environmental gradients. In this regard, we found strong context dependency for low (50%) levels of multifunctionality, as the relationship between below-ground multidiversity and multifunctionality changed both with study region and LUI (electronic supplementary material, figure S2). Regardless of the underlying mechanisms, the contrasting relationships between above- and below-ground biotic components and multifunctionality highlight the necessity to consider both in order to better understand the functional consequences of biodiversity loss in realistic landscapes.

(c) Individual species' effects on multifunctionality

Substantial research effort has been devoted to explain the functional role of individual species in natural ecosystems. Previous research suggests that the most abundant species [6], or a few key species with particular functional traits [31], will have the strongest effect on ecosystem functioning. These two hypotheses have received substantial empirical support across a large variety of systems and individual functions [6,14,15,17,32–35], but have rarely been tested for multiple functions simultaneously, or across multiple trophic groups. In addition to the effects of the diversity of the entire community, for 10% of the species tested, we found a significant relationship between their abundances and multifunctionality. This suggests that, despite potentially contrasting functional effects (positive, negative or neutral, depending on the function), some species, even individually, influence the overall ability of ecosystems to simultaneously provide multiple functions. This result supports the identity hypothesis [31,32], extending it to multiple functions and trophic levels. An example of one of these particularly influential species is *Hieracium pilosella*, a plant native to central Europe and locally rare in our study sites. This species was positively associated with multifunctionality according to our method and has previously been shown to increase soil organic carbon, litter decomposition and microbial biomass in comparison to other grassland species [62], to attract a variety of pollinators [63] and to have a relatively high resistance to pathogenic fungal infections [64]. We found a similar number of influential species for both common and rare species, and for both above- and below-ground organisms; indicating that individual species within these biotic components are equally important for multifunctionality. Understanding the attributes of these particularly influential species and their effects on multifunctionality should be a research priority if we are to predict the consequences of biodiversity loss and compositional changes for ecosystem service provision.

The direction of the relationship between the abundance of individual species and multifunctionality was best predicted by the response of individual species to land use intensification, even after accounting for the range in abundance across the plots and important functional traits. Previous studies have shown that land use intensification shifts plant functional composition and leads to an increase in the abundance of productive species, which enhance some provisioning services but compromise regulating and cultural services such as carbon storage or aesthetic value, therefore, reducing overall multifunctionality [25]. Similarly, changes in the ratio between soil fungi and bacteria with land use intensification may speed up nutrient recycling but

reduce ecosystem recovery after disturbances [24]. We show here that, alongside reducing their diversity, land use intensification may substantially influence the effect of multiple trophic levels, via compositional changes, on ecosystem functioning. Our results, therefore, suggest that the negative effect of LUI on multifunctionality at high thresholds could be, at least partly, caused indirectly by the negative response of particularly influential species to land use intensification. Such compositional changes might be particularly relevant for below-ground communities, whose relationships with multifunctionality (50% threshold) became negative under increasing LUI (electronic supplementary material, figure S2).

5. Conclusion

Substantial research effort has raised awareness of the functional consequences of losing biodiversity. However, we are still far from fully understanding which species or biodiversity attributes conservation efforts should focus on if ecosystem services are to be conserved. Our results suggest that locally rare above-ground species are the most important diversity component to preserve high levels of ecosystem multifunctionality in managed grasslands, perhaps due to their lower proportion of negative functional effects. Our multitrophic approach also supports the identity hypothesis, and extends it to multiple trophic groups and functions by showing, for the first time, that approximately 10% of the species tested can be particularly associated with overall ecosystem functioning. We also found that the effect of an individual species on multifunctionality is related to its response to LUI, which will help to anticipate the functional consequences of compositional changes across multiple trophic groups caused by land use intensification.

Authors' contributions. S.S. and E.A. conceived the study, all the authors but S.S. and E.A. gathered the data, M.G. compiled the trait data, S.S. and E.A. did the analyses, S.S. wrote the first draft and all co-authors significantly contributed to improve it.

Competing interests. We declare we have no competing interests.

Funding. This work was funded by the DFG (Deutsche Forschungsgemeinschaft; German Research Foundation) Priority Program 1374 'Infrastructure-Biodiversity Exploratories' (WE 3018/21-1, Li150/22-1).

Acknowledgements. We thank the managers of the three exploratories, Sonja Gockel, Kerstin Wiesner and Martin Gorke for their work in maintaining the plot and project infrastructure; Simone Pfeiffer and Christiane Fischer for giving support through the central office; Birgitta König-Ries and Michael Owonibi for managing the central database, and Eduard Linsenmair, Dominik Hessenmöller, Jens Nieschulze, Ernst-Detlef Schulze and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project. We are further grateful to Boris Büche, Roland Achtziger, Thomas Wagner, Torben Kölckebeck, Frank Köhler, Theo Blick, Franz Schmolke, Michael-Andreas Fritze and Günter Köhler for arthropod species identification. Fieldwork permits were given by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg (according to §72 BbgNatSchG).

References

1. Hector A, Bagchi R. 2007 Biodiversity and ecosystem multifunctionality. *Nature* **448**, 188–190. (doi:10.1038/nature05947)
2. Zavaleta ES, Pasari JR, Hulvey KB, Tilman D. 2010 Sustaining multiple ecosystem functions in grassland communities requires higher biodiversity. *Proc. Natl Acad. USA* **107**, 1443–1446. (doi:10.1073/pnas.0906829107)
3. Maestre FT *et al.* 2012 Plant species richness and ecosystem multifunctionality in global drylands. *Science* **335**, 214–218. (doi:10.1126/science.1215442)
4. Craven D *et al.* 2016 Plant diversity effects on grassland productivity are robust to both nutrient enrichment and drought. *Phil. Trans. R. Soc. B* **371**, 20150277. (doi:10.1098/rstb.2015.0277)
5. Brose U, Hillebrand H. 2016 Biodiversity and ecosystem functioning in dynamic landscapes. *Phil. Trans. R. Soc. B* **371**, 20150267. (doi:10.1098/rstb.2015.0267)
6. Grime JP. 1998 Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.* **86**, 902–910. (doi:10.1046/j.1365-2745.1998.00306.x)
7. Gamfeldt L, Hillebrand H, Jonsson PR. 2008 Multiple functions increase the importance of biodiversity for overall ecosystem functioning. *Ecology* **89**, 1223–1231. (doi:10.1890/06-2091.1)
8. Isbell F *et al.* 2011 High plant diversity is needed to maintain ecosystem services. *Nature* **477**, 199–202. (doi:10.1038/nature10282)
9. Lyons KG, Brigham CA, Traut BH, Schwartz MW. 2005 Rare species and ecosystem functioning. *Conserv. Biol.* **19**, 1019–1024. (doi:10.1111/j.1523-1739.2005.00106.x)
10. McIntyre PB, Jones LE, Flecker AS, Vanni MJ. 2007 Fish extinctions alter nutrient recycling in tropical freshwaters. *Proc. Natl Acad. USA* **104**, 4461–4466. (doi:10.1073/pnas.0608148104)
11. Pendleton R, Hoeinghaus D, Gomes L, Agostinho A. 2014 Loss of rare fish species from tropical floodplain food webs affects community structure and ecosystem multifunctionality in a mesocosm experiment. *PLoS ONE* **9**, e84568. (doi:10.1371/journal.pone.0084568)
12. Suding KN, Collins SL, Gough L, Clark C, Cleland EE, Gross KL, Milchunas DG, Pennings S. 2005 Functional- and abundance-based mechanisms explain diversity loss due to N fertilization. *Proc. Natl Acad. USA* **102**, 4387–4392. (doi:10.1073/pnas.0408648102)
13. Allan E *et al.* 2014 Interannual variation in land-use intensity enhances grassland multidiversity. *Proc. Natl Acad. USA* **111**, 308–313. (doi:10.1073/pnas.1312213111)
14. Smith MD, Knapp AK. 2003 Dominant species maintain ecosystem function with non-random species loss. *Ecol. Lett.* **6**, 509–517. (doi:10.1046/j.1461-0248.2003.00454.x)
15. Vile D, Shipley B, Garnier E. 2006 Ecosystem productivity can be predicted from potential relative growth rate and species abundance. *Ecol. Lett.* **9**, 1061–1067. (doi:10.1111/j.1461-0248.2006.00958.x)
16. Longo G, Seidler TG, Garibaldi LA, Tognetti PM, Chaneton EJ. 2013 Functional group dominance and identity effects influence the magnitude of grassland invasion. *J. Ecol.* **101**, 1114–1124. (doi:10.1111/1365-2745.12128)
17. Kleijn D *et al.* 2015 Delivery of crop pollination services is an insufficient argument for wild pollinator conservation. *Nat. Comm.* **6**, 7414. (doi:10.1038/ncomms8414)
18. Lyons KG, Schwartz MW. 2001 Rare species loss alters ecosystem function-invasion resistance. *Ecol. Lett.* **4**, 358–365. (doi:10.1046/j.1461-0248.2001.00235.x)
19. Zavaleta ES, Hulvey KB. 2004 Realistic species losses disproportionately reduce grassland resistance to biological invaders. *Science* **306**, 1175–1177. (doi:10.1126/science.1102643)
20. Booth JE, Gaston KJ, Evans KL, Armsworth PR. 2011 The value of species rarity in biodiversity recreation: a birdwatching example. *Biol. Conserv.* **144**, 2728–2732. (doi:10.1016/j.biocon.2011.02.018)
21. Steudel B, Hector A, Friedi T, Löftke C, Lorenz M, Wesche M, Kessler M. 2012 Biodiversity effects on ecosystem functioning change along environmental stress gradients. *Ecol. Lett.* **15**, 1397–1405. (doi:10.1111/j.1461-0248.2012.01863.x)

22. Perkins DM, Bailey RA, Dossena M, Gamfeldt L, Reiss J, Trimmer M, Woodward G. 2015 Higher biodiversity is required to sustain multiple ecosystem processes across temperature regimes. *Glob. Change Biol*. **21**, 396–406. (doi:10.1111/gcb.12688)
23. Jing X *et al.* 2015 The links between ecosystem multifunctionality and above- and belowground biodiversity are mediated by climate. *Nat. Comm*. **6**, 8159. (doi:10.1038/ncomms9159)
24. de Vries FT *et al.* 2013 Soil food web properties explain ecosystem services across European land use systems. *Proc. Natl Acad. USA* **110**, 14 296–14 301. (doi:10.1073/pnas.1305198110)
25. Allan E *et al.* 2015 Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. *Ecol. Lett*. **18**, 834–843. (doi:10.1111/ele.12469)
26. Byrnes JEK *et al.* 2014 Investigating the relationship between biodiversity and ecosystem multifunctionality: challenges and solutions. *Methods Ecol. Evol*. **5**, 111–124. (doi:10.1111/2041-210X.12143)
27. Flynn DFB *et al.* 2009 Loss of functional diversity under land use intensification across multiple taxa. *Ecol. Lett*. **12**, 22–33. (doi:10.1111/j.1461-0248.2008.01255.x)
28. Lefcheck JS *et al.* 2015 Biodiversity enhances ecosystem multifunctionality across trophic levels and habitats. *Nat. Commun.* **6**, 6936. (doi:10.1038/ncomms7936)
29. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000 Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858. (doi:10.1038/35002501)
30. Balvanera P *et al.* 2013 Linking biodiversity and ecosystem services: current uncertainties and the necessary next steps. *Bioscience* **64**, 49–57. (doi:10.1093/biosci/bit003)
31. Sala OE, Lauenroth WK, McNaughton SJ, Rusch G, Zhang X. 1996 Biodiversity and ecosystem function in grasslands. In *Functional roles of biodiversity: a global perspective* (eds HA Mooney, JH Cushman, E Medina, OE Sala, ED Schulze), pp. 129–149. New York, NY: John Wiley & Sons.
32. McLaren JR, Turkington R. 2010 Ecosystem properties determined by plant functional group identity. *J. Ecol*. **98**, 459–469. (doi:10.1111/j.1365-2745.2009.01630.x)
33. Mitchell CE, Tilman D, Groth JV. 2002 Effects of grassland plant species diversity, abundance, and composition on foliar fungal disease. *Ecology* **83**, 1713–1726. (doi:10.1890/0012-9658(2002)083[1713:EOGPSD]2.0.CO;2)
34. Vivanco L, Austin AT. 2008 Tree species identity alters forest litter decomposition through long-term plant and soil interactions in Patagonia, Argentina. *J. Ecol*. **96**, 727–736. (doi:10.1111/j.1365-2745.2008.01393.x)
35. Maas B, Tscharntke T, Saleh S, Dwi Putra D, Clough Y. 2015 Avian species identity drives predation success in tropical cacao agroforestry. *J. Appl. Ecol*. **52**, 735–743. (doi:10.1111/1365-2664.12409)
36. Millennium Ecosystem Assessment. 2005 *Ecosystems and human well-being: biodiversity synthesis*. Washington, DC: World Resources Institute.
37. Fischer M *et al.* 2010 Implementing large-scale and long-termfunctional biodiversity research: the biodiversity exploratories. *Basic Appl. Ecol*. **11**, 473–485. (doi:10.1016/j.baae.2010.07.009)
38. Blüthgen N *et al.* 2012 A quantitative index of land-use intensity in grasslands: integrating mowing, grazing and fertilization. *Basic Appl. Ecol*. **13**, 207–220. (doi:10.1016/j.baae.2012.04.001)
39. Gaston KJ. 1994 *Rarity*. Exeter, UK: Springer.
40. Burnham KP, Anderson DR. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. New York, NY: Springer.
41. Gessler PE, Moore ID, McKenzie NJ, Ryan PJ. 1995 Soil-landscape modelling and spatial prediction of soil attributes. *Int. J. Geogr. Info. Syst*. **4**, 421–432. (doi:10.1080/02693799508902047)
42. Sørensen R, Sinko U, Siebert J. 2006 On the calculation of topographic wetness index: evaluation of different methods based on field observations. *Hydrol. Earth Syst. Sci*. **10**, 101–112. (doi:10.5194/hess-10-101-2006)
43. Kittle AM, Fryxell JM, Desy GE, Hamr J. 2008 The scale-dependent impact of wolf predation risk on resource selection by three sympatric ungulates. *Oecologia* **157**, 163–175. (doi:10.1007/s00442-008-1051-9)
44. Gotelli NJ, Ulrich W, Maestre FT. 2011 Randomization tests for quantifying species importance to ecosystem function. *Methods Ecol. Evol*. **2**, 634–642. (doi:10.1111/j.2041-210X.2011.00121.x)
45. Ulrich W. 2010 *Impact—a FORTRAN program for gradient analysis*. Version 1.0. See http::www.umk.pl/~ulrichw.
46. Kleyer M *et al.* 2008 The LEDA Traitbase: a database of life-history traits of the Northwest European flora. *J. Ecol*. **96**, 1266–1274. (doi:10.1111/j.1365-2745.2008.01430.x)
47. Gossner MM *et al.* 2015 A summary of eight traits of Coleoptera, Hemiptera, Orthoptera and Araneae, occurring in grasslands in Germany. *Sci. Data* **2**, 150013. (doi:10.1038/sdata.2015.13)
48. Simmons NK, Weisser WW, Gossner MM. In press. Multi-taxa approach shows consistent shifts in arthropod functional traits along grassland land-use intensity gradient. *Ecology*. (doi:10.1890/15-0616)
49. Birkhofer K, Smith HG, Weisser WW, Wolters V, Gossner M. 2015 Land-use effects on the functional distinctness of arthropod communities. *Ecography* **38**, 1–12. (doi:10.1111/ecog.01141)
50. Garnier E *et al.* 2007 Assessing the effects of land-use change on plant traits, communities and ecosystem functioning in grasslands: a standardized methodology and lessons from an application to 11 European sites. *Ann. Bot*. **99**, 967–985. (doi:10.1093/aob/mcl215)
51. R Core Team. 2013 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
52. Mouillot D *et al.* 2013 Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol*. **11**, e1001569. (doi:10.1371/journal.pbio.1001569)
53. Mouillot D, Villéger S, Scherer-Lorenzen M, Mason NWH. 2011 Functional structure of biological communities predicts ecosystem multifunctionality. *PLoS ONE* **6**, e17476. (doi:10.1371/journal.pone.0017476)
54. Valencia E, Maestre FT, Le Bagousse-Pinguet Y, Quero JL, Tamme R, Börger L, García-Gómez M, Gross N. 2015 Functional diversity enhances the resistance of ecosystem multifunctionality to aridity in Mediterranean drylands. *New Phytol*. **206**, 660–671. (doi:10.1111/nph.13268)
55. Bradford MA *et al.* 2014 Discontinuity in the responses of ecosystem processes and multifunctionality to altered soil community composition. *Proc. Natl Acad. USA* **111**, 14 478–14 483. (doi:10.1073/pnas.1413707111)
56. Gámez-Virués S *et al.* 2015 Landscape simplification filters species traits and drives biotic homogenization. *Nat. Commun*. **6**, 8568. (doi:10.1038/ncomms9568)
57. Birkhofer K, Diekötter T, Boch S, Fischer M, Müller J, Socher S, Wolters V. 2011 Soil fauna feeding activity in temperate grassland soils increases with legume and grass species richness. *Soil Biol. Biochem*. **43**, 2200–2207. (doi:10.1016/j.soilbio.2011.07.008)
58. Karp DS, Ziv G, Zook J, Ehrlich PR, Daily GC. 2011 Resilience and stability in bird guilds across tropical countryside. *Proc. Natl Acad. USA* **108**, 21 134–21 139. (doi:10.1073/pnas.1118276108)
59. Bardgett RD, van der Putten WH. 2014 Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511. (doi:10.1038/nature13855)
60. Wagg C, Bender SF, Widmer F, van der Heijden MGA. 2014 Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proc. Natl Acad. USA* **111**, 5266–5270. (doi:10.1073/pnas.1320054111)
61. Birkhofer K *et al.* 2012 General relationships between abiotic soil properties and soil biota across spatial scales and different land-use types. *PLoS ONE* **7**, e43292. (doi:10.1371/journal.pone.0043292)
62. Saggar S, McIntosh PD, Hedley CB, Knicker H. 1999 Changes in soil microbial biomass, metabolic quotient, and organic matter turnover under *Hieracium* (*H. pilosella* L.). *Biol. Fertil. Soils* **30**, 232–238. (doi:10.1007/s003740050613)
63. Butz Huryn VMH, Moller H. 1995 An assessment of the contribution of honey bees (*Apis mellifera*) to weed reproduction in New Zealand protected natural areas. *N. Z. J. Ecol*. **19**, 111–122. (doi:10.1080/00779962.2010.9722196)
64. Morin L, Syrett P. 1996 Prospects for biological control of *Hieracium pilosella* with the rust *Puccinia hieracii* var. piloselloidarum in New Zealand. In *Proc. of the IX Int. Symp. on biological control of weeds* (eds VC Moran, JH Hoffmann), pp. 199–204. Stellenboch, South Africa: Stellenboch University Press.

Chapter 4 – Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality

LETTER

doi:10.1038/nature19092

Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality

Santiago Soliveres[1], Fons van der Plas[1,2], Peter Manning[1,2], Daniel Prati[1], Martin M. Gossner[3,4], Swen C. Renner[5,6], Fabian Alt[7], Hartmut Arndt[8], Vanessa Baumgartner[9], Julia Binkenstein[10], Klaus Birkhofer[11], Stefan Blaser[1], Nico Blüthgen[12], Steffen Boch[1,13], Stefan Böhm[5], Carmen Börschig[14], Francois Buscot[15,16], Tim Diekötter[17], Johannes Heinze[18,19], Norbert Hölzel[20], Kirsten Jung[21], Valentin H. Klaus[20], Till Kleinebecker[20], Sandra Klemmer[15], Jochen Krauss[22], Markus Lange[3,4,23], E. Kathryn Morris[24,25], Jörg Müller[18], Yvonne Oelmann[7], Jörg Overmann[9], Esther Pašalić[3,4], Matthias C. Rillig[19,25], H. Martin Schaefer[26], Michael Schloter[27], Barbara Schmitt[1], Ingo Schöning[3,23], Marion Schrumpf[23], Johannes Sikorski[9], Stephanie A. Socher[28], Emily F. Solly[23,29], Ilja Sonnemann[30], Elisabeth Sorkau[7], Juliane Steckel[22], Ingolf Steffan-Dewenter[22], Barbara Stempfhuber[27], Marco Tschapka[21,31], Manfred Türke[3,4,16,32], Paul C. Venter[8], Christiane N. Weiner[12], Wolfgang W. Weisser[3,4], Michael Werner[22], Catrin Westphal[14], Wolfgang Wilcke[33], Volkmar Wolters[34], Tesfaye Wubet[15,16], Susanne Wurst[30], Markus Fischer[1,2,13] & Eric Allan[1,35]

Many experiments have shown that loss of biodiversity reduces the capacity of ecosystems to provide the multiple services on which humans depend[1,2]. However, experiments necessarily simplify the complexity of natural ecosystems and will normally control for other important drivers of ecosystem functioning, such as the environment or land use. In addition, existing studies typically focus on the diversity of single trophic groups, neglecting the fact that biodiversity loss occurs across many taxa[3,4] and that the functional effects of any trophic group may depend on the abundance and diversity of others[5,6]. Here we report analysis of the relationships between the species richness and abundance of nine trophic groups, including 4,600 above- and below-ground taxa, and 14 ecosystem services and functions and with their simultaneous provision (or multifunctionality) in 150 grasslands. We show that high species richness in multiple trophic groups (multitrophic richness) had stronger positive effects on ecosystem services than richness in any individual trophic group; this includes plant species richness, the most widely used measure of biodiversity. On average, three trophic groups influenced each ecosystem service, with each trophic group influencing at least one service. Multitrophic richness was particularly beneficial for 'regulating' and 'cultural' services, and for multifunctionality, whereas a change in the total abundance of species or biomass in multiple trophic groups (the multitrophic abundance) positively affected supporting services. Multitrophic richness and abundance drove ecosystem functioning as strongly as abiotic conditions and land-use intensity, extending previous experimental results[7,8] to real-world ecosystems. Primary producers, herbivorous insects and microbial decomposers seem to be particularly important drivers of ecosystem functioning, as shown by the strong and frequent positive associations of their richness or abundance with multiple ecosystem services. Our results show that multitrophic richness and abundance support ecosystem functioning, and demonstrate that a focus on single groups has led to researchers to greatly underestimate the functional importance of biodiversity.

Global change is causing species loss across many trophic groups[3,4], with potential effects on the services that ecosystems provide to humans[1,2]. The functional consequences of a decline in biodiversity across multiple trophic groups are hard to predict from studies focusing on single taxa, as the functional effects of different groups may complement or oppose each other[5,6,9,10]. The effects of the diversity of plants and microbes are complementary, maximizing rates of nutrient cycling[11]; plant and herbivore diversity, on the other hand, have opposing effects on biomass stocks[10,12,13]. Consequently, we know very little about the relative effect of changes in the diversity of different trophic groups on the provision of individual[2,5,6,9,13,14] or multiple (multifunctionality)[11,15] ecosystem services.

In addition to decreasing species richness, global change is altering the total abundance (total number of individuals or amount of biomass within communities) of multiple trophic groups[4]. Changes in

[1]Institute of Plant Sciences, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland. [2]Senckenberg Gesellschaft für Naturforschung, Biodiversity and Climate Research Centre BIK-F, Senckenberganlage 25, 60325 Frankfurt, Germany. [3]Institute of Ecology, Friedrich-Schiller-University Jena, Dornburger Straße 159, D-07743 Jena, Germany. [4]Technische Universität München, Terrestrial Ecology Research Group, Department of Ecology and Ecosystem Management, School of Life Sciences Weihenstephan, Hans-Carl-von-Carlowitz-Platz 2, 85354 Freising, Germany. [5]Institute of Zoology, University of Natural Resources and Life Science, Gregor-Mendel-Straße 33, 1180 Vienna, Austria. [6]Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, Virginia 22630, USA. [7]Geocology, University of Tuebingen, Ruemelinstr. 19-23, 72070 Tuebingen, Germany. [8]University of Cologne, Institute for Zoology, Zülpicher Str. 47b, 50674 Cologne, Germany. [9]Leibniz Institute DSMZ German Collection of Microorganisms and Cell Cultures, Inhoffenstr. 7B, 38124 Braunschweig, Germany. [10]Chair of Nature Conservation and Landscape Ecology, Faculty of Environment and Natural Resources, University of Freiburg, Tennenbacher Straße 4, 79106 Freiburg, Germany. [11]Department of Biology, Lund University, Sölvegatan 35, 22362 Lund, Germany. [12]Ecological Networks, Biology, Technische Universität Darmstadt, Schnittspahnstr. 3, 64287 Darmstadt, Germany. [13]Botanical Garden, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland. [14]Agroecology, Department of Crop Sciences, Georg-August University of Göttingen, Grisebachstr. 6, D-37077, Göttingen, Germany. [15]UFZ-Helmholtz Centre for Environmental Research, Department of Soil Ecology, Theodor-Lieser-Straße 4, 06120 Halle (Saale), Germany. [16]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, D-04103 Leipzig, Germany. [17]Department of Landscape Ecology, Kiel University, Olshausenstr. 75, D-24118 Kiel, Germany. [18]Biodiversity Research/Systematic Botany, University of Potsdam, Maulbeerallee 1, D-14469 Potsdam, Germany. [19]Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), D-14195 Berlin, Germany. [20]Institute of Landscape Ecology, University of Münster, Heisenbergstr. 2, 48149 Münster, Germany. [21]Institute of Evolutionary Ecology and Conservation Genomics, University of Ulm, Albert-Einstein-Allee 11, 89069 Ulm, Germany. [22]Department of Animal Ecology and Tropical Biology, Biocentre, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany. [23]Max-Planck Institute for Biogeochemistry, Hans-Knoell-Str. 10, 07745 Jena, Germany. [24]Xavier University, Department of Biology, 3800 Victory Parkway, Cincinnati, Ohio 45207, USA. [25]Plant Ecology, Institut für Biologie, Freie Universität Berlin, Altensteinstr. 6, D-14195 Berlin, Germany. [26]Department of Ecology and Evolutionary Biology, Faculty of Biology, University of Freiburg, Hauptstraße 1, 79104 Freiburg i. Br., Germany. [27]Research Unit for Environmental Genomics; Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85758 Oberschleissheim, Germany. [28]Department of Ecology and Evolution, Universität Salzburg, Hellbrunnerstrasse 34, 5020 Salzburg, Austria. [29]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland. [30]Functional Biodiversity, Institute of Biology, Freie Universität Berlin. Königin-Luise-Str. 1-3. D-14195 Berlin, Germany. [31]Smithsonian Tropical Research Institute, Balboa, Panama. [32]Institute for Biology, Leipzig University, Johannisallee 21, D-04103 Leipzig, Germany. [33]Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), Reinhard-Baumeister-Platz 1, 76131 Karlsruhe, Germany. [34]Department of Animal Ecology, Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32, D-35392 Giessen, Germany. [35]Centre for Development and Environment, University of Bern, Hallerstrasse, 10, 3012 Bern, Switzerland.

abundance could mitigate or exacerbate the functional consequences of species loss[16,17] by influencing the ability of each trophic group to capture resources. However, studies normally focus on the effects of community evenness or of dominant species[18–20], whereas the simultaneous effects of changes in richness and total abundance on the functioning of ecosystems have been largely unexplored[6,16]. The relative importance of richness and abundance may depend on the function or service of interest. Total abundance could be a main driver of biogeochemical process rates (for example, biomass production[18], nutrient capture and cycling). By contrast, ecosystem services related to biotic interactions, such as pollination or pest control, could be predominantly driven by species richness[16]. Ecosystem services also depend on abiotic factors and, although experiments show that the effects of biodiversity loss on ecosystem functioning are as large as those of abiotic drivers[7,8], it is unclear whether species richness and abundance are similarly important in real-world ecosystems[6,14,21,22].

We adopted a multitrophic approach to evaluate relationships between biodiversity and multifunctionality in 150 real-world grasslands. We measured the richness and abundance of species in nine trophic groups: primary producers, above- and below-ground herbivores and predators, detritivores, soil microbial decomposers, plant symbionts, and bacterivores. These trophic groups comprised 4,600 plant, animal and microbial taxa, and were measured alongside 14 ecosystem variables (proxies for both functions and services, hereafter referred to as services). These are related to the four main types of ecosystem services[23]: provisioning (fodder production and quality), supporting (potential nitrification, phosphorus retention, root biomass and decomposition rate, mycorrhizal colonization and soil aggregate stability), regulating (soil carbon levels, pollinator abundance, pest control, resistance to pathogens) and cultural services (recreation benefits of flower cover and bird diversity). We fitted linear models to our data to test for both positive and negative relationships between the richness and abundance of species within the nine trophic groups and each ecosystem service, the four types of services (provisioning, supporting, regulating and cultural), and ecosystem multifunctionality[22] (see Methods). We accounted for potential confounding factors by performing our analyses on residuals, after controlling for variability in land-use intensity, soils and climate. We compared our results with models that included only plant-species richness, the most commonly used measure of biodiversity[21,24,25], and with models that included the richness and abundance of each individual trophic group. Additional analyses compared the amount of variance explained by, and the effect size (standardized slope) of, multitrophic richness and abundance with those of land-use intensity and environmental variables.

Effects on individual ecosystem services, service types, and multifunctionality were better predicted by changes in multitrophic richness and abundance than by those in the richness or abundance of any individual trophic group (Fig. 1 and Extended Data Fig. 1). The most parsimonious models included the richness and/or abundance of 3.14 ± 0.36 trophic groups (average $\pm$ s.e.m. across all 14 services) to predict the variation in each ecosystem service. These results remained when using raw data instead of environment-corrected residuals (Extended Data Fig. 2), different combinations of ecosystem services (Extended Data Fig. 3), and even when we accounted for well-established links between a predictor and service in our models (for example, plant cover versus biomass). Multitrophic richness had stronger and more positive relationships with the provisioning, regulating and cultural services than plant richness alone (Fig. 1 and Extended Data Fig. 1). For example, both plant and predator richness were related to high levels of pest control, suggesting that combined top-down and bottom-up effects of diversity[26] maximize the provision of this regulating service. Multitrophic richness also had a more positive effect than even the strongest positive-richness effect found across all individual trophic groups on the regulating and cultural services. The findings of our observational study were supported by a quantitative review of the few studies that manipulated the richness of more than one group. Our

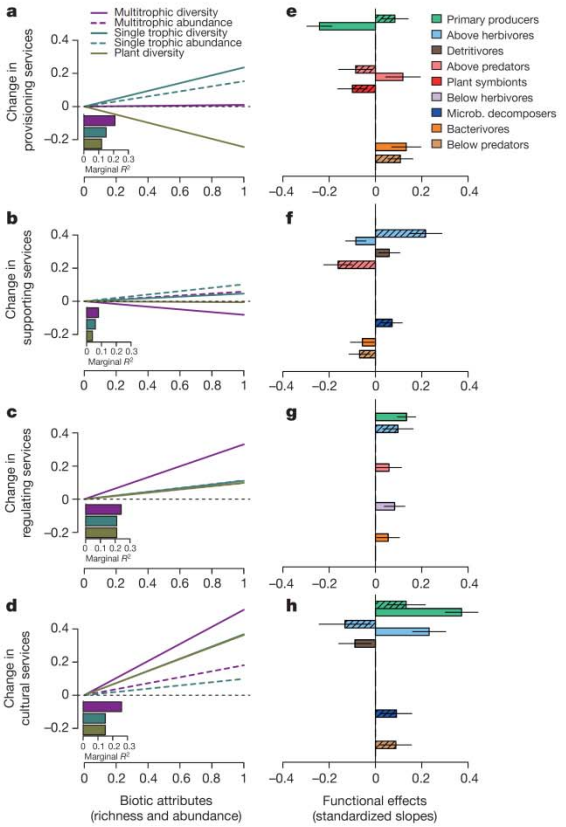


Figure 1 | Effects of multitrophic richness and abundance on grassland functioning. a–d, Variance explained after accounting for the influence of 'site' as random factor (marginal R^2, the equivalent of R^2 for mixed models) and standardized effects for each ecosystem service when models included abundance and richness of multiple or individual trophic groups. **e–h**, Standardized effects (mean $\pm$ s.e.m.) of richness and abundance (full and hatched bars, respectively) of individual trophic groups on each ecosystem service type. Ecosystem services types are plant biomass and forage quality (provisioning); potential nitrification, phosphorus retention, mycorrhizal colonization, soil aggregate stability, root biomass and decomposition (supporting); soil carbon, pollinator abundance, pest control and resistance to pathogens (regulating); flower cover and bird diversity (cultural).

review showed that including the richness of a second trophic group increased the variance in ecosystem functioning by 14–96% for litter decomposition[14], biomass production[2,12,26], or the number of carbon sources used[5] (Extended Data Table 1). Collectively, our results show that high species richness in multiple trophic groups is necessary to maintain high levels of ecosystem functioning, particularly for regulating and cultural services.

Alongside multitrophic richness, the combined effect of a high multitrophic abundance strongly affected ecosystem functioning (according to the amount of variance explained and its effect size). Multitrophic abundance had positive effects on the provisioning and supporting services, but these were generally weaker than those found for the individual trophic group that had the strongest positive effect. This suggests that an abundance of some trophic groups can dampen the effect on ecosystem functioning induced by others. Figure 1, for instance, shows that a higher abundance of predators partially counteracted the positive effects of abundant herbivores on supporting services. Conversely, a high level of richness in a given trophic

RESEARCH LETTER

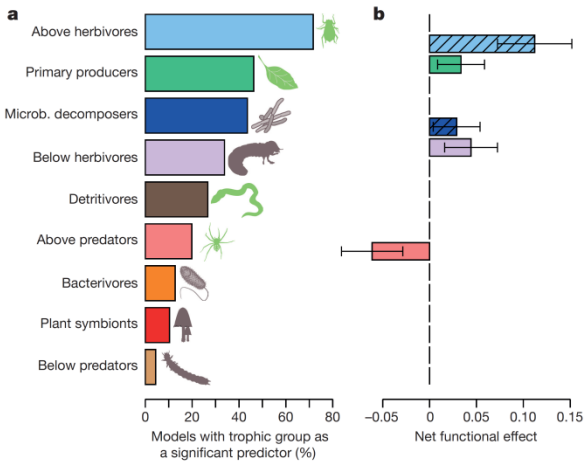


Figure 2 | Functional importance of multiple trophic groups. **a**, Proportion of the multifunctionality metrics (calculated using every possible combination of 1–9 services; $N = 501$; see Methods) in which the biotic attributes (richness and/or abundance) of each trophic group was included in the most parsimonious model. **b**, Functional effects (standardized slopes (mean ± s.e.m.) in the model fitted to all 14 services) of the richness (open bars) and abundance (hatched bars) of each group. Bars are shown only for the predictors included in the most parsimonious models. Green and brown cartoons indicate above- and below-ground trophic groups, respectively.

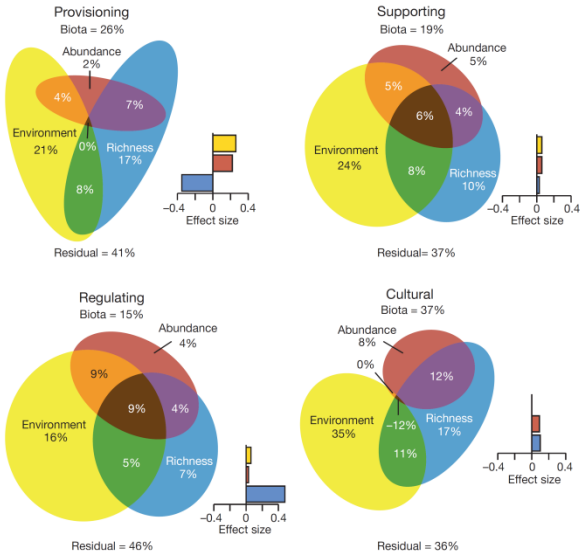


Figure 3 | Biotic versus abiotic drivers of ecosystem functioning. Variation partitioning for three predictor categories in our statistical models: environment, species richness and total abundance (details in Methods). Diagrams show the average across services within each ecosystem service type (detailed results are in Extended Data Figs 2 and 5). Shown are the unique variance explained by each predictor category, the shared variance between these categories (intersections of circles), and the variance not explained by the models (the residual). Biota refers to the total variance explained by species abundance and richness combined. Standardized effect sizes are shown as bar plots.

group generally complements the positive effects of other trophic groups on ecosystem services (see a comparison of multitrophic and unitrophic richness in Fig. 1). These contrasting effects caused multitrophic abundance to increase ecosystem multifunctionality only at low-to-moderate levels (Extended Data Fig. 1). Overall, our results underline the important role of species richness in driving the functioning of ecosystems[1,2,14–17,24,25], while also highlighting the often-overlooked effect of total biomass abundance on the supporting and provisioning services.

To test how generally applicable the trends in relationships between multitrophic richness and abundance were, we calculated multifunctionality metrics using all possible combinations of services. High multitrophic richness or abundance had increasingly positive effects as more services were considered and this effect was consistent across a wide range of levels of multifunctionality (Extended Data Fig. 3). To further explore this result, we calculated the similarities in the identities of the trophic groups driving a given pair of ecosystem services (the functional overlap, $\bar{o}$ (ref. 25)). On average, we found functional overlaps lower than 30% ($\bar{o} = 0.27 \pm 0.03$, mean ± s.e.m.), similar to results found for plant species in grassland experiments ($\bar{o} = 0.19$–0.49)[25]. This demonstrates low multitrophic redundancy and means that different services are supported by different trophic groups (Fig. 1 and Extended Data Fig. 1). We also found that different groups positively affected multifunctionality when it was calculated according to scenarios representing different land-use objectives (Extended Data Fig. 4). Finally, five of the nine trophic groups had the strongest net-positive effects on at least one ecosystem service (for example, primary producers on pest control, soil microbial decomposers on aggregate stability; Extended Data Fig. 1), with each group affecting at least one service. Collectively, these results show the low functional redundancy found between the multiple trophic groups studied, explaining why high multitrophic richness is needed to support high levels of ecological multifunctionality or to promote a larger number of ecosystem services.

The relationships between multitrophic richness, multitrophic abundance and ecosystem services were not always positive (Figs 1, 2 and Extended Data Fig. 1), consistent with previous studies[27,28]. Negative relationships might be explained by interference between species within a given trophic group[10] or by compositional shifts leading to declines in ecosystem functioning[17,20,29]. Despite these negative associations or harm to services, our results suggest that the most important trophic groups for maintenance of the services considered are aboveground herbivorous insects, primary producers and soil microbial decomposers. The richness or abundance of these trophic groups were most often correlated to ecosystem multifunctionality (43–72% of the 501 possible combinations between the services we measured), and had net-positive effects across all services (Fig. 2). These three groups also showed strong and frequent positive associations with the four main ecosystem service types (Fig. 1). These results agree with other studies that have identified plants and soil microorganisms as key drivers of ecosystem functioning[11,14,15,29], extending these findings to the richness and abundance of different trophic groups, including primary producers and consumers both above and below ground. The species richness of some of these functionally important trophic groups relate to whole-ecosystem diversity[3,30] and, thus, management strategies focused on them may foster synergies between biodiversity conservation and high multifunctionality levels.

The relative importance of both multitrophic richness and abundance compared to the environmental drivers of ecosystem functioning has been rarely studied outside of experiments or individual functions[7,8,11,15,21,24]. We therefore calculated the proportion of variance in ecosystem functioning that was explained by multitrophic richness, abundance and environmental (soil, topography and land-use) factors. Our models accounted for a large proportion (54–64%) of the variance in the provisioning, supporting, regulating and cultural ecosystem service types (Fig. 3 and Extended Data Fig. 5). Multitrophic richness and abundance explained at least as much of the variance in ecosystem functioning as abiotic conditions or land-use intensity did, and generally had stronger effects (Fig. 3; Extended Data Fig. 5). These results provide evidence that biodiversity is of comparative importance to environmental factors in driving ecosystem functioning. This is true

not only for individual functions in small-scale experiments[7,8], but also for multiple ecosystem services in realistic landscapes (see refs 15, 24).

Our study shows that the functional importance of biodiversity in real-world ecosystems has been greatly underestimated, as a result of focussing on individual trophic groups. We demonstrate here that the functional effects of multitrophic richness and abundance are as strong as, or even stronger than, those of the environment or land-use intensity. We identified primary producers, above-ground herbivores and soil decomposers as particularly important trophic groups for maintaining a functioning ecosystem. Our results suggest that it is important to preserve high levels of species richness, abundance or both within a wide range of taxa. This must include taxa often ignored by conservation efforts such as soil microbial decomposers[15], or those considered pests in agricultural systems such as herbivorous insects, if we are to promote high levels of the multiple ecosystem services upon which human well-being depends.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 March; accepted 7 July 2016.
Published online 17 August; corrected online 24 August 2016 (see full-text HTML version for details).

1. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486,** 59–67 (2012).
2. Naeem, S., Duffy, J. E. & Zavaleta, E. The functions of biological diversity in an age of extinction. *Science* **336,** 1401–1406 (2012).
3. Allan, E. *et al.* Interannual variation in land-use intensity enhances grassland multidiversity. *Proc. Natl Acad. Sci. USA* **111,** 308–313 (2014).
4. Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity. *Nature* **520,** 45–50 (2015).
5. Naeem, S., Hahn, D. R. & Schuurman, G. Producer-decomposer co-dependency influences biodiversity effects. *Nature* **403,** 762–764 (2000).
6. Balvanera, P. *et al.* Linking biodiversity and ecosystem services: current uncertainties and the necessary next steps. *Bioscience* **64,** 49–57 (2014).
7. Hooper, D. U. *et al.* A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* **486,** 105–108 (2012).
8. Tilman, D., Reich, P. B. & Isbell, F. Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory. *Proc. Natl Acad. Sci. USA* **109,** 10394–10397 (2012).
9. Petchey, O. L., McPhearson, P. T., Casey, T. M. & Morin, P. J. Environmental warming alters food-web structure and ecosystem function. *Nature* **402,** 69–72 (1999).
10. Duffy, J. E. *et al.* The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecol. Lett.* **10,** 522–538 (2007).
11. Jing, X. *et al.* The links between ecosystem multifunctionality and above- and belowground biodiversity are mediated by climate. *Nat. Commun.* **6,** 8159 (2015).
12. Douglass, J. G., Duffy, J. E. & Bruno, J. F. Herbivore and predator diversity interactively affect ecosystem properties in an experimental marine community. *Ecol. Lett.* **11,** 598–608 (2008).
13. Deraison, H., Badenhausser, I., Loeuille, N., Scherber, C. & Gross, N. Functional trait diversity across trophic levels determines herbivore impact on plant community biomass. *Ecol. Lett.* **18,** 1346–1355 (2015).
14. Handa, I. T. *et al.* Consequences of biodiversity loss for litter decomposition across biomes. *Nature* **509,** 218–221 (2014).
15. Delgado-Baquerizo, M. *et al.* Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat. Commun.* **7,** 10541 (2016).
16. Garibaldi, L. A. *et al.* Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science* **339,** 1608–1611 (2013).
17. McGrady-Steed, J., Harry, P. M. & Morin, P. J. Biodiversity regulates ecosystem predictability. *Nature* **390,** 162–165 (1997).
18. Grime, J. P. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.* **86,** 902–910 (1998).
19. Hillebrand, H., Bennett, D. M. & Cadotte, M. W. Consequences of dominance: a review of evenness effects on local and regional ecosystem processes. *Ecology* **89,** 1510–1520 (2008).
20. Soliveres, S. *et al.* Locally rare species influence grassland ecosystem multifunctionality. *Phil. Trans. R. Soc. B* **371,** 20150269 (2016).
21. Grace, J. B. *et al.* Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* **529,** 390–393 (2016).
22. Byrnes, J. E. K. *et al.* Investigating the relationship between biodiversity and ecosystem multifunctionality: challenges and solutions. *Methods Ecol. Evol.* **5,** 111–124 (2014).
23. Millennium Ecosystem Assessment. *Ecosystems and Human Well-Being* 1–86 World Resources Institute (2005).
24. Maestre, F. T. *et al.* Plant species richness and ecosystem multifunctionality in global drylands. *Science* **335,** 214–218 (2012).
25. Hector, A. & Bagchi, R. Biodiversity and ecosystem multifunctionality. *Nature* **448,** 188–190 (2007).
26. Bruno, J. F., Boyer, K. E., Duffy, J. E. & Lee, S. C. Relative and interactive effects of plant and grazer richness in a benthic marine community. *Ecology* **89,** 2518–2528 (2008).
27. Balvanera, P. *et al.* Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9,** 1146–1156 (2006).
28. Lefcheck, J. S. *et al.* Biodiversity enhances ecosystem multifunctionality across trophic levels and habitats. *Nat. Commun.* **6,** 6936 (2015).
29. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368,** 734–737 (1994).
30. Manning, P. *et al.* Grassland management intensification weakens the associations among the diversities of multiple plant and animal taxa. *Ecology* **96,** 1492–1501 (2015).

Acknowledgements We thank B. Schmid, F. T. Maestre and S. Kéfi for comments that helped improve this manuscript. W. Ulrich and N. J. Gotelli provided statistical advice. We thank the people who maintain the Biodiversity Exploratories program: A. Hemp, K. Wells, S. Gockel, K. Wiesner and M. Gorke (local management team); S. Pfeiffer and C. Fischer (central office), B. König-Ries and M. Owonibi (central database management); and E. Linsenmair, D. Hessenmöller, J. Nieschulze, E.-D. Schulze and the late E. Kalko for their role in setting up the project. This work was funded by the Deutsche Forschungsgemeinschaft Priority Program 1374 'Infrastructure-Biodiversity Exploratories'. Fieldwork permits were given by the responsible state environmental offices of Baden-Württemberg, Thüringen and Brandenburg (according to §72 BbgNatSchG). Figure icons were created by R. D. Manzanedo.

Author Contributions S.S. and E.A. conceived the idea of this study. M.F. initiated the Biodiversity Exploratories project aimed at measuring multiple diversities and functions in the field sites. All authors but S.S., E.A. and F.V.D.P. contributed data. S.S. and F.V.D.P. performed the analyses. S.S. and S.C.R. performed the literature search. S.S. wrote the first draft of the manuscript and all the authors (especially E.A., P.M., F.D.V.P., M.M.G. and D.P.) contributed substantially to the revisions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S. (santiago.soliveres@ips.unibe.ch).

Reviewer Information *Nature* thanks Y. Hautier, F. Isbell and the other anonymous reviewer(s) for their contribution to the peer review of this work.

RESEARCH LETTER

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Study sites. We selected a total of 150 grassland sites (50 m × 50 m) in three regions of Germany (50 sites per region) to cover a gradient of land-use intensities, characterized by contrasting grazing, fertilization and mowing levels (www.biodiversity-exploratories.de, ref. 31). The regions in the south-west (Schwäbische Alb) and the north-east (Schorfheide-Chorin) are UNESCO Biosphere Reserves, whereas the central region is in and around the Hainich National Park. The three regions differ substantially in geology, climate and topography[31], covering a range of ~3 °C in mean annual temperature and 500 mm in annual precipitation. Plots in each region cover the range of land-use intensities typical for Central European grasslands. We obtained information on land use via questionnaires sent to land owners, asking about the number and type of livestock (converted to livestock units) and the duration of grazing in each plot, the fertilization (from which we calculated the amount of nitrogen added), and the mowing (number of cuts per year[31,32]). We used this information to calculate three standardized indices summarizing grazing, fertilization and mowing intensity (see ref. 32 for full methodological details).

Diversity measures. At each site, we measured the species richness and abundance of nine functional groups using standard methodologies (Extended Data Table 2). In total we observed about 4,600 taxa on the 150 grasslands studied. The nine trophic groups were: primary producers (vascular plants and bryophytes), below-ground herbivores (herbivorous insect larvae sampled in the soil), below-ground predators (carnivorous insect larvae sampled in the soil), detritivores (insects and Diplopoda feeding on leaf litter and other detritus), soil microbial decomposers (soil bacteria), above-ground herbivores (insects feeding solely on above-ground plant material), above-ground predators (carnivorous insects, spiders and Chilopoda), plant symbionts (arbuscular mycorrhizal fungi), and bacteria-feeding protists (heterotrophic flagellates and ciliates). Lichens and omnivores were not considered in our analyses as they were too rare. We directly measured species richness for most groups, but richness was quantified as family richness for below-ground insects and soil bacteria and as the number of operational taxonomic units (OTUs) for the mycorrhizae and protists. The abundance of each trophic group was also measured using different methods: number of individuals for arthropods, amount of cover for vascular plants and bryophytes, and relative proportion of sequence reads assigned to each family or OTU for protists, soil bacteria and mycorrhiza. To avoid multicollinearity, we did not include the abundances of protists or detritivores as they were highly correlated (Spearman's $\rho > 0.6$) with richness (for more details see Extended Data Table 3).

We also measured the abundance and richness of foliar fungal pathogens, pollinators and birds; however, to include a broader range of ecosystem services in our analyses, we treated these groups as proxies of ecosystem services. Total pollinator abundance and the inverse of pathogen abundance were treated as proxies of regulating services (pollination and disease regulation), and we used bird-species richness as a measure of a cultural ecosystem service. Lepidoptera behave as herbivores during juvenile stages and as pollinators when adults. To avoid accounting for them twice, we assigned them to only one trophic group (pollinators), as the data were counts of the adult butterflies, not the caterpillars.

Ecosystem functioning measures. At each site, we measured 14 different ecosystem variables (both functions and service proxies; Extended Data Table 2) and classified them into four types of services following the Millennium Ecosystem Assessment[23]. These 14 ecosystem services were: i) supporting services related to nutrient capture and cycling (root biomass, root decomposition rates, potential nitrification (based on urease activity in soil samples), phosphorus retention (calculated as a ratio between shoot and microbial phosphorus stock and soil extractable phosphorus), arbuscular mycorrhizal fungal root colonization (measured as hyphal length), soil aggregate stability (proportion of water-stable soil aggregates)); ii) provisioning services related to agricultural value (forage production (above-ground plant biomass) and forage quality (based on crude protein and relative forage value); iii) regulating services for neighbouring crop production or climate regulation (that is, regulating services: resistance to plant pathogens, pest control, pollinator abundance and soil organic carbon); or iv) cultural services linked to recreation (bird diversity and flower cover). Because the values for trophic groups and ecosystem functions varied widely, we standardized all variables to a common scale ranging from 0 to 1 according to the following formula: $\mathrm{STD} = (X - X_{\min})/(X_{\max} - X_{\min})$; where STD is the standardized variable and X, $X_{\min}$ and $X_{\max}$ are the target variable, and its minimum and maximum value across all sites, respectively. This made slope estimates for different predictors comparable.

We calculated ecosystem multifunctionality metrics from the 14 services as the percentage of measured services (measured services only to correct for the fact that some services that had not been measured in all sites) that exceeded a given threshold of their maximum observed level across all study sites. To reduce the influence of outliers we calculated the maximum observed level as the average of the top five sites[22,33]. Given that any threshold is likely to be arbitrary, the use of multiple thresholds is recommended to better understand the role that biodiversity plays in affecting ecosystem multifunctionality and to understand trade-offs between functions of interest[22]. Therefore, we used four different thresholds (25%, 50%, 75% and 90%) to represent a wide spectrum in the analyses performed (Extended Data Figs 1–3). As an alternative approach we also calculated multifunctionality scenarios, weighting the services differently according to the different potential views of stakeholders (that is, stakeholders willing only to promote provisioning services versus those trying to maximize cultural and recreation services or the sustainability of soils and crops; Extended Data Fig. 4)[34].

Effects of multitrophic richness and abundance on grassland ecosystem services and multifunctionality. We used linear models to evaluate the relationships between species richness and abundance in the nine trophic groups and each of the 14 individual ecosystem services, along with the different multifunctionality metrics (four thresholds and the metrics were obtained by weighting each ecosystem service according to different potential stakeholders' needs; for example, only provision, sustainable soils and crops, or cultural scenarios, see ref. 34). In all cases, we used a Gaussian error distribution as the errors of our response variables were normally distributed. We report the effects of the different trophic groups on the different functions as slopes from the multiple regression model; these are corrected for the effects of all other variables in the model. Since our main focus was on calculating the independent effects of the richness and abundance of the different trophic groups, we corrected them for co-varying factors. Thus, we calculated residuals for all our variables (both biotic predictors and functioning measures) from linear models including region, land-use intensity (standardized measures of mowing, grazing and fertilization intensity) and other important environmental factors (soil type and depth, pH, a topographic wetness index based on position within the slope and orientation, and elevation). As an alternative to using residuals, we also fitted models with all the environmental and land-use predictors (standardized to give comparable coefficients) alongside the diversity and abundance measures. These approaches gave very similar results (Extended Data Fig. 2). Standardized coefficients of the functional effects of richness were very similar, whether or not abundance was included ($\rho = 0.80$, $P < 0.0001$, $N = 162$; data not shown). We also fitted models with the abundance and richness of only one individual trophic group to compare the results of the best individual trophic group with the multitrophic analyses (Extended Data Fig. 1). Finally, we fitted models with only richness of vascular plant species as a predictor. The latter is the most common measure of biodiversity[7,8,21,24,25,35–38] and we used it to compare our results with those found in previous studies on biodiversity–ecosystem functioning relationships.

We performed model simplification using the stepAIC function in R, and further simplified the minimal models produced using AIC by removing all terms that were not significant according to F-ratio tests (Extended Data Table 4). Results using alternative approaches for model selection are presented in Extended Data Fig. 6. We did not fit interactions between the richness and abundance of different trophic groups, or between those and environmental factors, as this would require a large number of coefficients, would be difficult to interpret and would require an even larger data set than ours (see ref. 20 for a study evaluating the interaction between land-use and diversity). We did not find evidence of nonlinear relationships between our predictors and the ecosystem services measured when checking all bivariate relationships; thus we did not include nonlinear terms in the models to keep them simple.

Not all trophic groups or ecosystem services were measured on all sites; thus different services were analysed using different sized data sets (N ranged between 111 and 54, depending on the service). The different sampling sizes used were not related to the number of trophic groups included in the most parsimonious model (Spearman's rank correlation coefficient $\rho = 0.32$), the increase in variance explained by vascular plant species richness ($\rho = -0.21$) or the net effect of richness or abundance ($\rho = -0.01$ or 0.05, respectively; $N = 14$ and $P > 0.25$ in all cases). Thus, fitting models differing in sample size for different services did not affect our results.

The inclusion of many predictors in statistical models increases the chance of type I error (false positives). To account for this we used a Bernoulli process to detect false discovery rates, where the probability (P) of finding a given number of significant predictors (K) just by chance is a proportion of the total number of predictors tested ($N = 16$ in our case: the abundance and richness of 7 and 9 trophic groups, respectively) and the P value considered significant ($\alpha = 0.05$ in our case)[39,40]. The probability of finding three significant predictors on average, as we did, is therefore, $P = [16!/(16-3)!3!] \times 0.053(1-0.05)^{(16-3)} = 0.0359$, indicating

that the effects we found are very unlikely to be spurious. The probability of false discovery rates when considering all models and predictors fit (14 ecosystem services × 16 richness and abundance metrics) and the ones that were significant amongst them (52: 25 significant abundance predictors and 27 significant richness predictors) was even lower ($P < 0.0001$). All analyses were performed using R version 3.0.2 (ref. 41).

Net functional effects of the different trophic groups across ecosystem service types. We calculated the net effect of each trophic group on each ecosystem service type (provisioning, supporting, regulating and cultural) by fitting all services belonging to these types into a single model. To do so, we added two extra predictors to our models: 'service identity' as a fixed factor, to account for differences between individual services, and 'site' as a random factor, to account for correlations between services, abundance and richness values measured on the same site. Since we were interested in the net effects of each group across all services, we did not fit interactions between our multitrophic predictors and service identity. The net effect across all services was analysed using the same approach, while fitting a single model for the 14 ecosystem services at the same time. This approach corrects for the fact that the individual service models vary in their explanatory power and in the predictor variables included. Fitting all services into a single model allows us to obtain a robust estimate of the net functional effect (the standardized coefficient from the model) of the abundance and richness of each trophic group on the service type of each ecosystem and on ecosystem multifunctionality, together with an estimate of its error. If the standardized coefficient was positive, we interpreted it as a net overall positive effect of either richness or abundance across all services, or on a given service type (Figs 1 and 2). In all cases, we used standardized coefficients of the most parsimonious models after model reduction. However, our results remained when using other approaches that account for differences in model fit, such as multi-model averaging coefficients (coefficients were weighted according to the AIC weight of the models in which each predictor is included) or when weighting the standardized coefficient for each ecosystem service by the adjusted R^2 of each model (which should also be comparable across models with different response variables; Extended Data Fig. 6).

Variance partitioning analyses. Variance partitioning analyses (also known as commonality analyses) were performed with standard techniques[42,43] based on the comparison of variance explained by models including every possible combination of variables. Variables were organized by environment (study region, soil type, pH, topographic wetness index, grazing and fertilization, with the remaining environmental predictors removed to prevent multicollinearity; shown in Extended Data Table 3), species richness (standardized species richness of the nine trophic groups) and abundance (standardized abundance of those trophic groups in which abundance and richness were not strongly correlated ($\rho < 0.6$; shown in Extended Data Table 3)). Thus, we fitted a series of seven models for each service and multifunctionality metric (at the 25%, 50%, 75% and 90% thresholds) to extract the unique and shared variance for each combination of variables (environment only, richness only, abundance only, environment + abundance, environment + richness, richness + abundance, and all predictors together). Variance-partitioning analyses were performed with the full models (without model simplification) to allow us to compare between different services. As a consequence, we used R^2 rather than the adjusted R^2 because, owing to the large number of predictors, some adjusted R^2 values were negative, complicating the extraction of unique variance explained by each predictor. Venn diagrams were drawn using Euler APE for Windows[44].

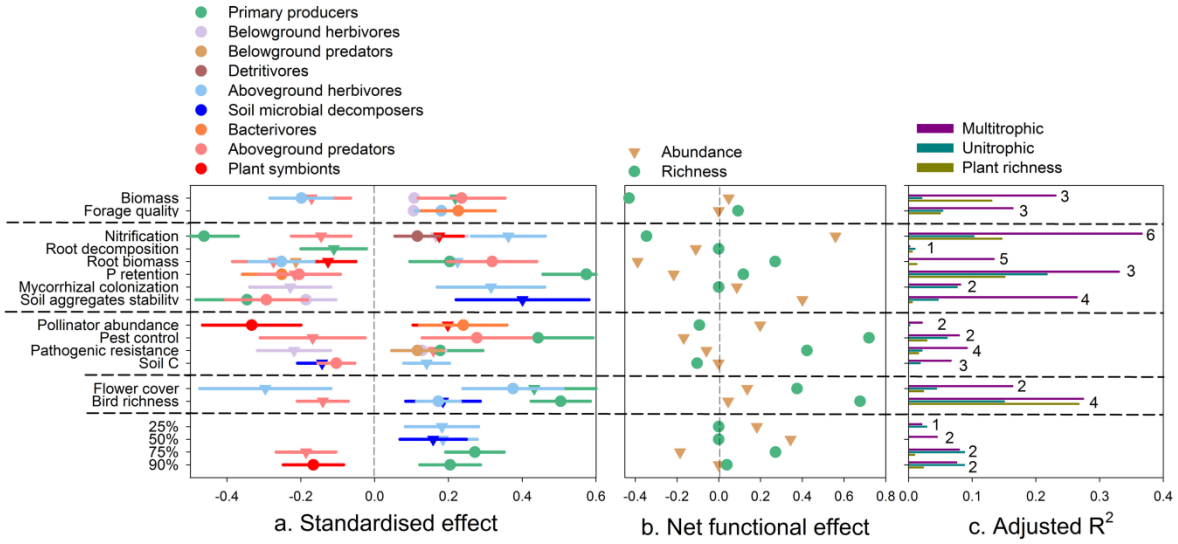
To compare the effect size among richness, abundance and environment on the different ecosystem services and multifunctionality metrics, we summed the standardized coefficients of all predictors from each component (the abundance of five trophic groups (abundance), the richness of nine trophic groups (richness), and pH, fertilization, grazing, and topographic wetness index (environment)). We excluded study region and soil type when summing effects, as these were categorical predictors and their coefficients were not straightforward to interpret. We performed these calculations for each of the 14 ecosystem services and 4 multifunctionality metrics in isolation (Extended Data Fig. 4), and for each ecosystem service type (Fig. 2) by using models containing all the ecosystem services belonging to each type into a single model (again, adding 'service identity' and 'site' as fixed and random predictors, respectively).

Analysing every possible combination of ecosystem services. Studies on multifunctionality are difficult to compare as they include different measures of ecosystem functioning. To allow us to generalize our results and to test whether multitrophic richness and abundance are more important in supporting higher numbers of services simultaneously, we also calculated multifunctionality indices using every possible combination of the services we measured. We did this after removing those services with more than 20 missing sites, leaving a total of 9 services (501 combinations) as response variables. We calculated multifunctionality at the 25%, 50%, 75% and 90% thresholds for all these combinations (Extended Data Fig. 3). We also tested the sensitivity of our analyses to missing data by repeating our analyses for every possible combination of 1–13 of the 14 measured services (16,368 combinations; results for multifunctionality calculated with all 14 ecosystem services are presented in Extended Data Figs 1, 2). To allow the comparison of models with different services, data gaps were filled with the average value of a given service in a given region, which is a conservative approach. In both cases (combinations of 1–13 or 1–9 functions), the most parsimonious models possible were selected on the basis of their AIC. This avoids inflated type I error, caused by fitting a large number of models, as model selection was not based on P values. Results using 9 or 14 functions were qualitatively the same and therefore only the former are shown here.

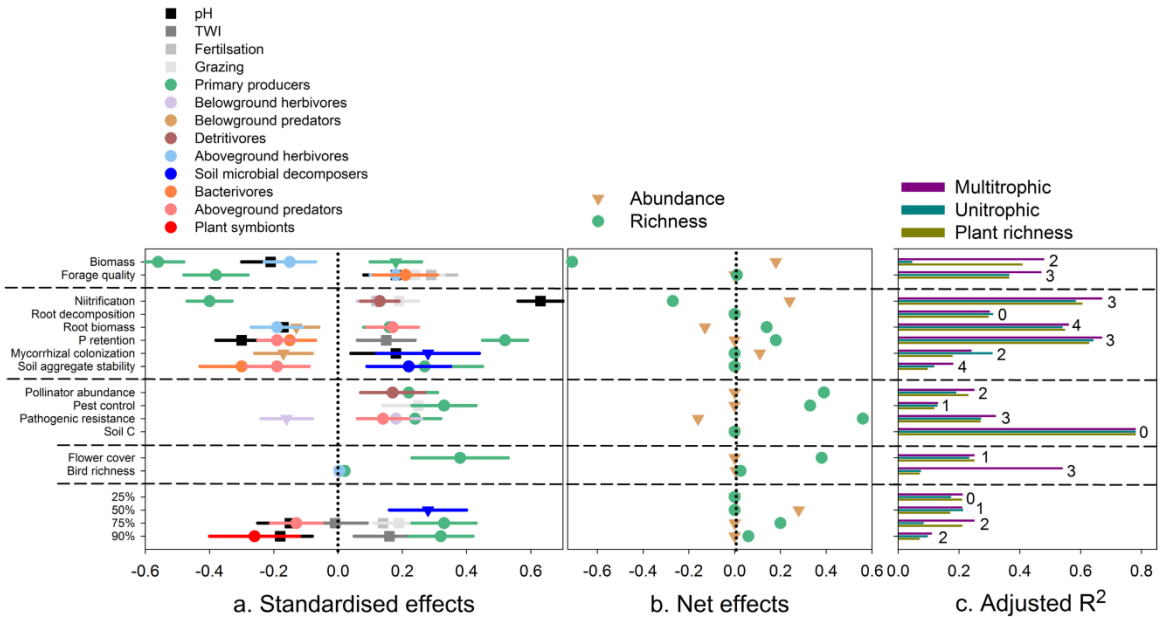
Review of multitrophic manipulative approaches. Manipulative experiments including as many groups and services as we considered in this study do not yet exist. However, we compared our correlational results with available evidence from experiments manipulating the diversity of more than one trophic group. To do this we performed a bibliographic research in the Web of Knowledge and in Google Scholar using all combinations of the terms 'multitrophic' or 'trophic groups' + 'functioning' or 'multifunctionality' or 'biomass' or 'ecosystem services' or 'diversity'. We also screened references within available reviews on multitrophic diversity–ecosystem functioning relationships[10,27,45]. Of the papers found, we selected those which fulfilled the following criteria: i) it was a manipulative study, ii) it included a range in species richness (not only presence or absence) of, at least, two different trophic groups and iii) it provided enough information to calculate the increase in variance explained by the addition of a second trophic group. Only four studies, including seven ecosystem functions, fulfilled these criteria (Extended Data Table 1). Some of these manipulative studies did not include plants, so we calculated the percentage increase in variance seen when comparing a model with the trophic group that had the strongest explanatory power in models containing two trophic groups. When the same function was measured across several studies (that is, biomass), we calculated the average increase in variance explained for this variable when another trophic level was added. These results were used to qualitatively compare the limited evidence from multitrophic manipulations with our results.

31. Fischer, M. *et al.* Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. *Basic Appl. Ecol.* **6,** 473–485 (2010).
32. Blüthgen, N. *et al.* A quantitative index of land-use intensity in grasslands: integrating mowing, grazing and fertilization. *Basic Appl. Ecol.* **13,** 207–220 (2012).
33. Zavaleta, E. S., Pasari, J. R., Hulvey, K. B. & Tilman, G. D. Sustaining multiple ecosystem functions in grassland communities requires higher biodiversity. *Proc. Natl Acad. Sci. USA* **107,** 1443–1446 (2010).
34. Allan, E. *et al.* Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. *Ecol. Lett.* **18,** 834–843 (2015).
35. Hautier, Y. *et al.* Eutrophication weakens stabilizing effects of diversity in natural grasslands. *Nature* **508,** 521–525 (2014).
36. Gamfeldt, L., Hillebrand, H. & Jonsson, P. R. Multiple functions increase the importance of biodiversity for overall ecosystem functioning. *Ecology* **89,** 1223–1231 (2008).
37. Gamfeldt, L. *et al.* Higher levels of multiple ecosystem services are found in forests with more tree species. *Nat. Commun.* **4,** 1340 (2013).
38. Cardinale, B. J. *et al.* Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* **443,** 989–992 (2006).
39. Moran, M. D. Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos* **100,** 403–405 (2003).
40. Tylianakis, J. M. *et al.* Resource heterogeneity moderates the biodiversity-function relationship in real world ecosystems. *PLoS Biol.* **6,** e122 (2008).
41. R Development Core Team. *R: A Language And Environment For Statistical Computing* (R Foundation For Statistical Computing, 2014).
42. Borcard, D., Legendre, P. & Drapeau, P. Partialling out the spatial component of ecological variation. *Ecology* **73,** 1045–1055 (1992).
43. Peres-Neto, P. R., Legendre, P., Dray, S. & Borcard, D. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87,** 2614–2625 (2006).
44. Micallef, L. & Rodgers, P. eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* **9,** e101717 (2014).
45. Worm, B. & Duffy, J. E. Biodiversity, productivity and stability in real food webs. *Trends Ecol. Evol.* **18,** 628–632 (2003).

RESEARCH LETTER

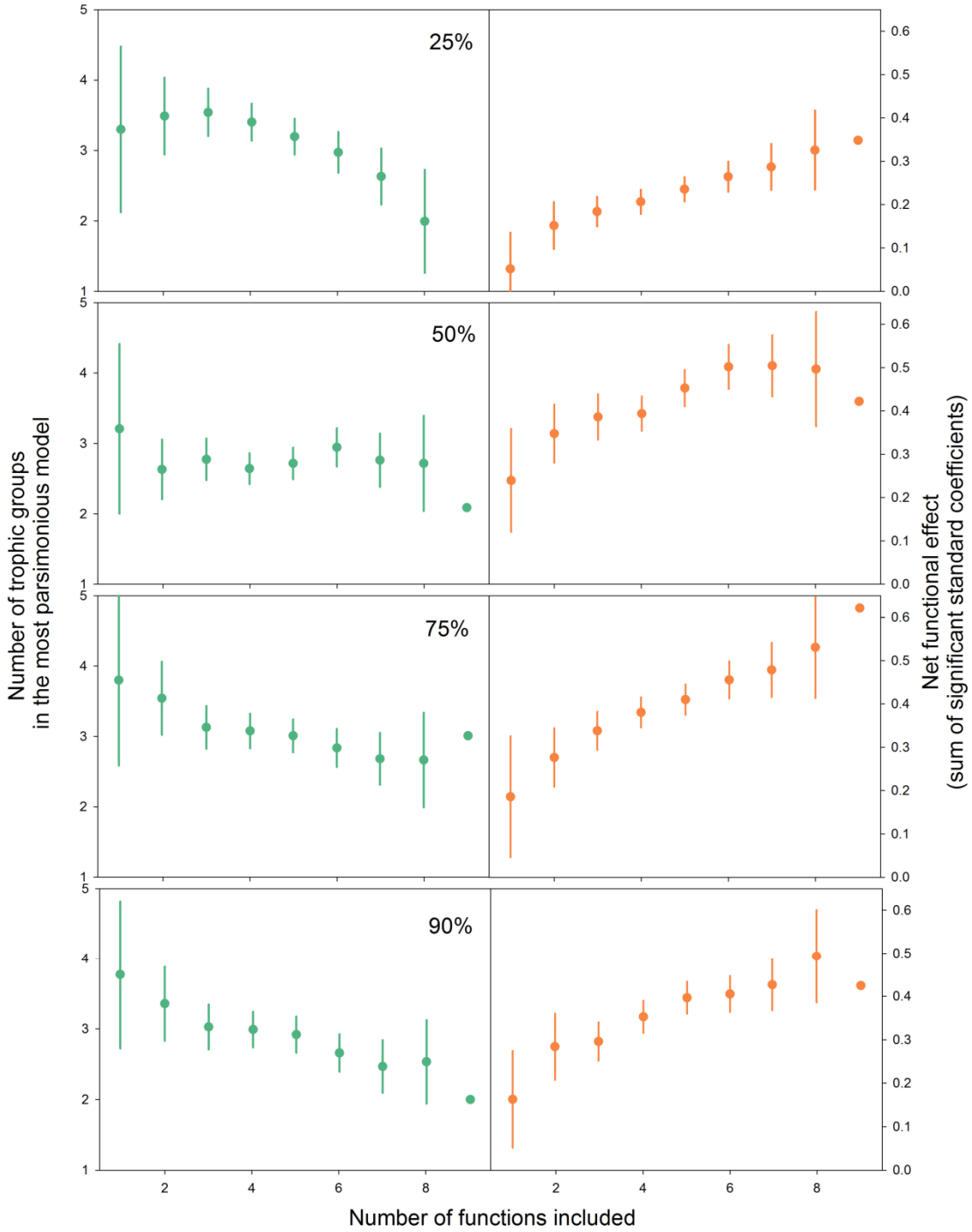


Extended Data Figure 1 | Functional effects of multitrophic richness and abundance on 14 grassland ecosystem services. a, Standardized coefficients (mean ± s.e.m.) of the abundances (triangles) and richness (circles) of those trophic groups that significantly affect a given function are shown. **b**, The net effect (that is, the sum of significant standardized effects). **c**, Difference in adjusted R^2 between the final multitrophic models and those models using the abundance and richness of the best performing individual trophic group (unitrophic) or plant species richness (plant richness). Ecosystem services are organized by the main four types of services they associate with (provisioning, supporting, regulating and cultural). The number of trophic groups included in the most parsimonious model is given next to their adjusted R^2. Multifunctionality results at 25%, 50%, 75% and 90% thresholds are also shown (see Methods).

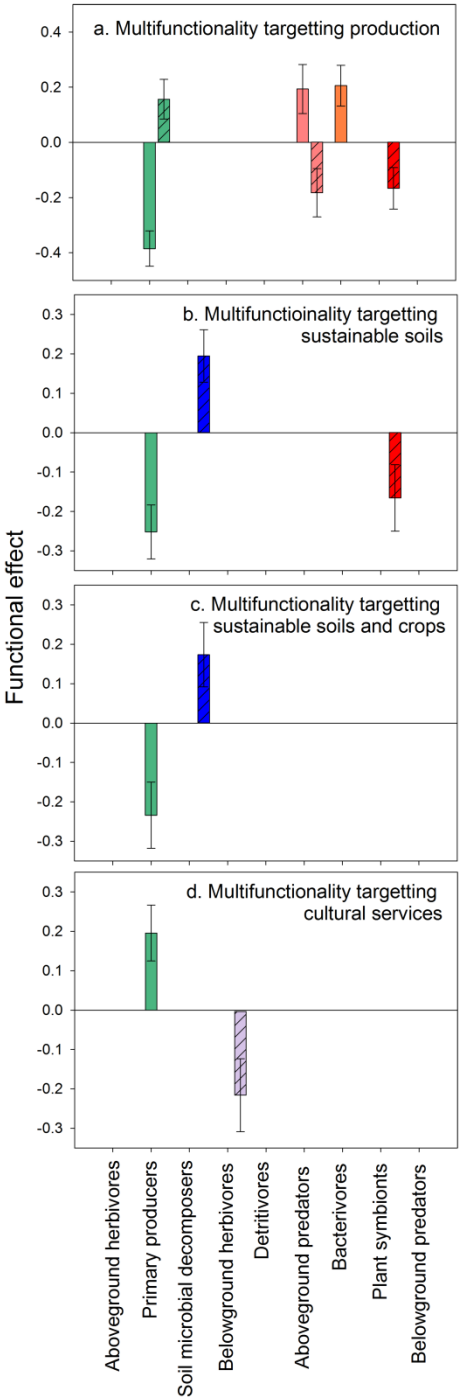


Extended Data Figure 2 | Functional effects of environmental factors and multitrophic richness and abundance on 14 grassland ecosystem functions. a, Standardized slope estimates (mean ± s.e.m.) for each significant predictor are shown, with the exception of study region and soil type, which were retained in all models. **b**, Net effect (sum of significant standardized effects) for multitrophic richness and abundance. **c**, The total amount of variance explained by either environmental + plant species richness, environmental + the abundance and richness of the best individual trophic predictor, or by environmental + multitrophic diversity and abundance are shown for each function (adjusted R^2, to control for the high number of predictors included). The number of trophic groups included in the best models (2.15 ± 1.2 across functions, and 1.94 ± 1.2 across functions and multifunctionality indices) is given next to the adjusted R^2 value. The increase in the adjusted R^2 values in models with plant-species-richness averaged 0.07 ± 0.12 (across functions) and 0.06 ± 0.11 (across functions and multifunctionality indices). Ecosystem services are organized by the main four types of services they associate with (top–bottom: provisioning, supporting, regulating and cultural). TWI, topographic wetness index, based on the aspect and position in the slope, and the inclination of the slope. Multicollinearity between the predictors introduced is unlikely (Extended Data Table 3).

RESEARCH LETTER

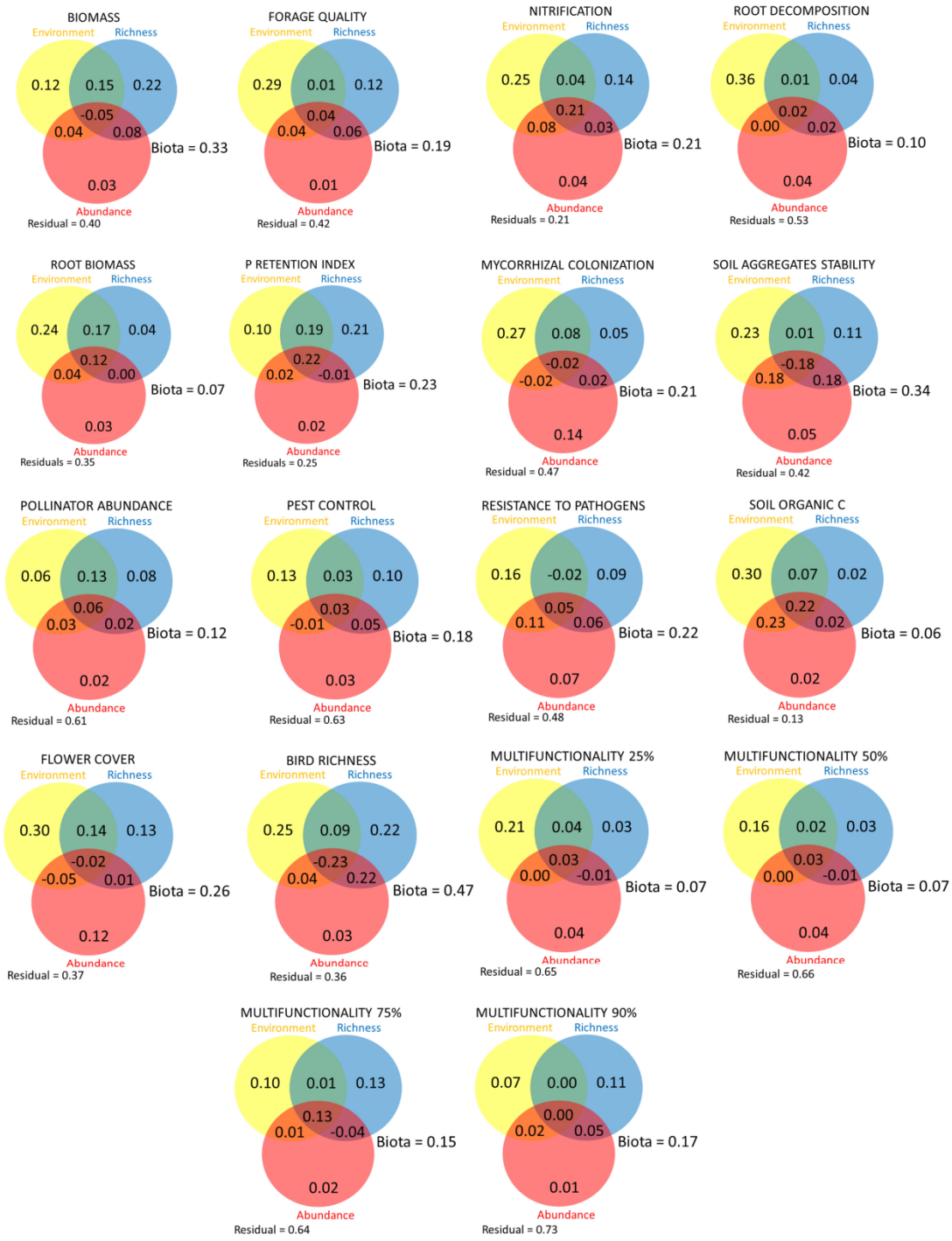


Extended Data Figure 3 | Number of trophic groups necessary to predict multifunctionality measures calculated with all possible combinations of 1–9 services, and their net effects. The number of predictors selected in the best models (left) and their overall effects (sum of standardized coefficients; right) across all possible combinations of 1–9 services ($N = 501$) are shown. Error bars show the 95% confidence intervals, estimated for all possible combinations of n (1 to 9) functions in both cases. Only the 9 services with fewer than 20 data gaps were considered in these analyses (see details in Methods). Multifunctionality for these combinations was calculated at the 25% (upper panel), 50%, 75% and 90% (bottom panel) thresholds. Services removed were flower cover, arbuscular mycorrhizal colonization, soil aggregate stability, phosphorous retention index and pest control.

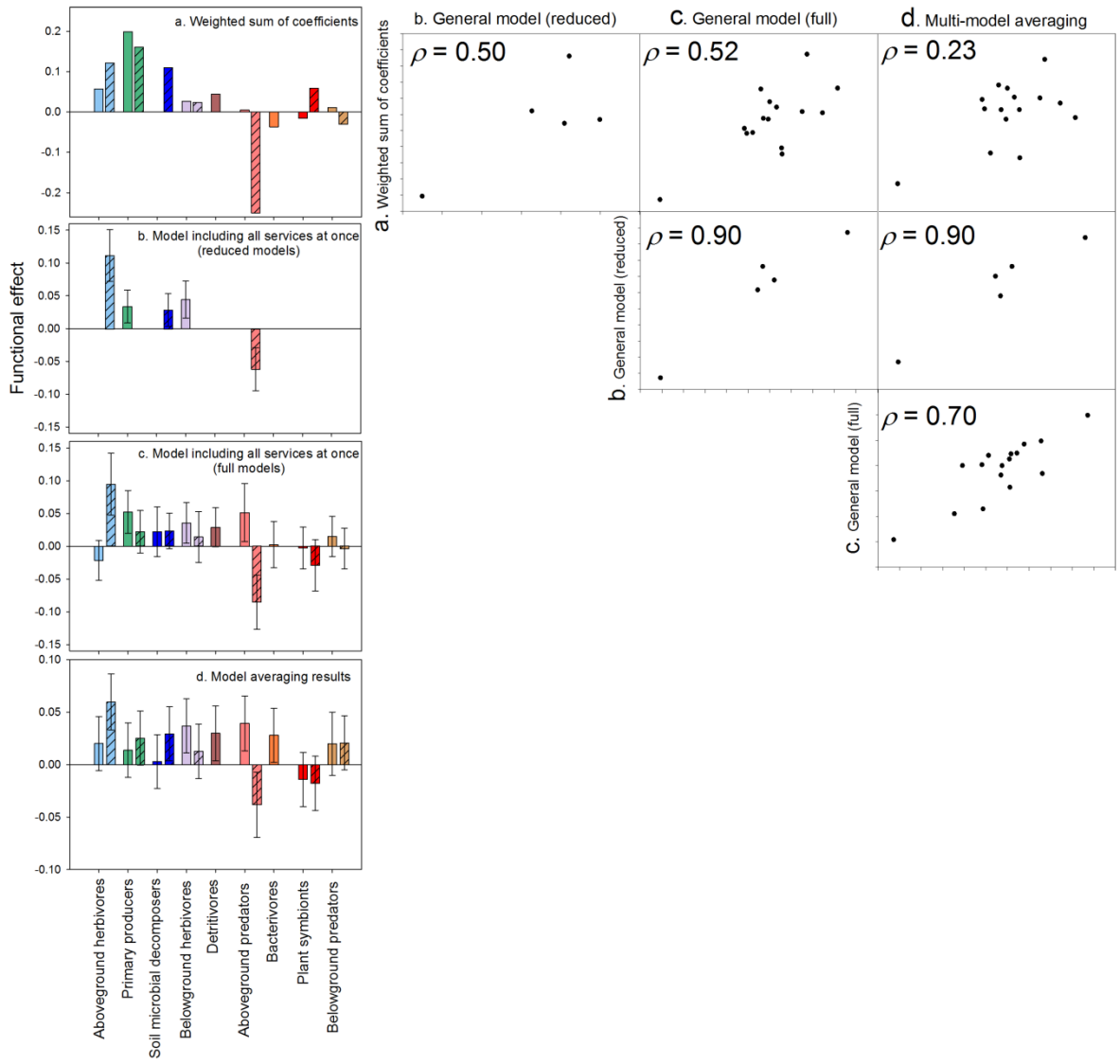


Extended Data Figure 4 | Functional effect of the different trophic groups on contrasting multifunctionality scenarios. Overall functional effects (significant standardized coefficients; mean ± s.e.m.) from the most parsimonious model) of the richness (open bars) and abundance (hatched bars) of each group are shown according to ref. 34.

RESEARCH LETTER



Extended Data Figure 5 | Functional importance of species richness and abundance compared to environmental drivers. Venn diagrams showing the variance partition for the four components of our statistical models (environment: climate, soil and land-use intensity; species richness of the nine trophic groups, abundance of primary producers, above- and below-ground predators, below-ground herbivores and soil microbial decomposers). The variance not explained by the model (the residual) is also shown. The variance explained by richness, abundance and their overlap is summed up as Biota. Each panel represents an individual function or multifunctionality metric.



Extended Data Figure 6 | Functional effect of the different trophic groups. Overall functional effects (mean ± s.e.m. of the standardized slopes obtained from the model; with the exception of **a**, where error could not be estimated) of the richness (open bars) and abundance (hatched bars) of each group. **a**, The values were calculated after weighting each standardized coefficient (those in Extended Data Fig. 1) by the adjusted R^2 of the model to account for differences in model performance. **b**, **c**, The values were calculated as the standardized coefficients in a general model fitted to all services at once, including 'service identity' as an extra predictor and 'plot' as random factor to control for pseudo-replication (reduced models (**b**); the ones presented in the main text), or full models (**c**) and, **d**, calculated as multi-model average parameters from a model fitted to all services at once. Correlations (Spearman's rank correlation coefficients) between the different approaches are given.

RESEARCH LETTER

Extended Data Table 1 | Re-analysis of manipulative multitrophic studies

Study ID	Reference	System	# trophic groups manipulated	Approach	Response variable	Variance explained best group	Variance explained two trophic groups	% increase with multitrophic approach	comment
1	Douglass et al. 2008 Ecolett	aquatic	2	mesocosm	grazer abundance	55.0	61.0	10.9	based on ω^2
1	Douglass et al. 2008 Ecolett	aquatic	2	mesocosm	predator abundance	5.0	7.0	40.0	based on ω^2
2	Bruno et al. 2008. Ecology	aquatic	2	mesocosm	autotroph biomass	19.5	42.7	119.0	based on F
3	Naeem et al. 2000. Nature	terrestrial	2	microcosm	autotroph biomass	13.6	23.5	72.8	based on F
3	Naeem et al. 2000. Nature	terrestrial	2	microcosm	detritivore biomass	5.6	11.3	101.8	based on F
3	Naeem et al. 2000. Nature	terrestrial	2	microcosm	# C sources used	7.2	13.5	87.5	based on F
4	Handa et al. 2014 Nature	both	2	field.expt	Litter C loss	5.8	6.6	13.8	based on %SS

For each study, an ID number and full reference are given. The system in which each study was performed (aquatic or terrestrial), the number of trophic groups manipulated and the approach used (controlled mesocosms or field studies) are provided. The ecosystem functions ('response variable') measured within each study were grouped in biomass production (the first five rows), nutrient cycling (sixth row) and decomposition (seventh row). Variance explained (according to the statistic mentioned in comments; ω^2 = proportion of variance explained according to the authors; F = Fisher's F, SS = sum of squares) for the single trophic group with the most explanatory power, and the difference between the variance explained by this group and the inclusion of a second group are given (grey column).

Extended Data Table 2 | Details of the sampling procedure for each trophic group and function

Trophic groups			
Trophic group	**Subgroup**	**Sampling method**	**Author**
Primary producers	Plants, bryophytes	Measurement of % cover in a 4×4 m subplot, done in 2009	Boch, Heinze, Hölzel, Klaus, Kleinebecker, Müller, Prati, Socher, Fischer
Aboveground herbivores	Herbivorous insects	Sweep netting (Hemiptera: Heteroptera/Auchenorrhyncha, Hymenoptera, Neuroptera and Orthoptera). Transects of 150m with 60 double sweeps, done twice per plot in 2008-2010.	Lange, Paŝalić, Türke, Gossner, Weisser
Aboveground predators	Carnivorous insects	Sweep netting (Hemiptera: Heteroptera/Auchenorrhyncha, Hymenoptera, Neuroptera and Orthoptera). Transects of 150m with 60 double sweeps done twice per plot in 2008-2010.	Lange, Paŝalić, Türke, Gossner, Weisser
	Spiders	Sweep netting. Transects of 150m with 60 double sweeps, done twice per plot in 2008-2010.	Lange, Paŝalić, Türke, Gossner, Weisser
Detritivores	Chilopoda	Kempson extraction from one soil core of 20 ×5 cm per plot, done in 2008	Birkhofer, Diekötter, Wolters
	Annelids	Hand sorting from two soil cores of 20 ×10 cm per plot, done in 2008	Birkhofer, Diekötter, Wolters
	Diplopoda	Kempson extraction from one soil core of 20 ×5 cm per plot, done in 2008	Birkhofer, Diekötter, Wolters
	Detritivorous insects	Sweep netting (Hemiptera: Heteroptera/Auchenorrhyncha, Hymenoptera, Neuroptera and Orthoptera). Transects of 150m with 60 double sweeps, done twice per plot in 2008-2010.	Lange, Paŝalić, Türke, Gossner, Weisser
Microbial decomposers	Soil bacteria	cDNA amplicon sequencing of partial (V3) 16S rRNA gene transcripts, done in 2011	Baumgartner, Sikorski, Overmann
Bacterivores	Bacterivorous protists	18S rDNA gene PCR and amplicon sequencing (454) filtering for rhizarians, alveolates, stramenopiles and opisthokonts, done in 2011	Venter, Arndt
Symbionts	Arbuscular mycorrhizal fungi	Pyrotag sequencing of the NS31 - AM1 fragment of the 18S rDNA genes, done in 2011	Klemmer, Wubet, Buscot
Belowground herbivores	Insect larvae	Extracted from a heat/moisture gradient in one soil core of 20 x 5 cm per site, done in 2011 over a period of eight days.	Sonnemann, Wurst
Belowground predators	Insect larvae	Extracted from a heat/moisture gradient in one soil core of 20 x 5 cm per site, done in 2011 over 8 days.	Sonnemann, Wurst
Functions			
Function		**Sampling method**	**Author**
Aboveground plant biomass		Harvested in four 0.5 m × 0.5 m quadrats per plot, done in May-June in 2008-2012.	Schmitt, Prati, Fischer, Klaus, Kleinebecker, Hölzel
Belowground plant biomass		Measured in 14 soil cores (0-10 cm). Fine roots were sorted and weighted after drying in the oven, done in samples collected in May 2011	Solly, Schöning, Schrumpf
Root decomposition rate		Measured as the mass loss from root litter bags after 6 months, from October 2011 to April 2012.	Solly, Schöning, Schrumpf
Potential nitrification		10 mM ammonium sulphate solution was added as substrate to 2.5g of soil composite samples (i.e. the same samples as for soil carbon; see below).	Stempfhuber, Schloter
Phosphorus uptake and retention		Proportion of P in plants and microbes (shoot P stock + microbial P stock) / (shoot P stock + microbial P stock + soil extractable P [$NaHCO_3$]).	Alt, Sorkau, Oelman, Wilcke, Klaus, Kleinebecker, Hölzel
Arbuscular mycorrhizal fungal root colonization		Cultured in sterile soil in the field from April to October 2011 and then extracted with sodium hexametaphosphate (35 g l^{-1}). Hyphal length was quantified after staining with trypan blue.	Morris, Rillig
Stability of soil aggregates		A subsample of the same soil than above (AMF colonization) was passed through a 250 µm sieve under water to determine the percentage of water stable macroaggregates.	Morris, Rillig
Soil organic Carbon		Measured in 14 soil cores (0-10 cm). Calculated as the difference between total carbon (measured with a CN analyzer "Vario Max" [Elementar Analysensysteme GmbH, Hanau, Germany]) and inorganic carbon (determined after combustion of organic carbon in a muffle furnance; 450°C for 16 h), done in samples collected in May 2011	Schöning, Solly, Schrumpf
Forage quality		Was calculated as a function of mean of scaled crude protein concentration and scaled relative forage value, done in May-June in 2008-2012.	Klaus, Kleinebecker, Hölzel
Resistance to plant pathogens		Calculated as the inverse of the total cover of foliar fungal pathogens. The cover of pathogens was measured in four 25 × 1 m transects per plot, were proportion of plants infected, and leaf area infected of these individuals was measured; done in October 2011,	Blaser, Prati, Fischer
Pest control		Number of trap nesting wasps known to feed on pest insects, done between April and October 2008.	Steckel, Westphal, Steffan-Dewenter
Pollinator abundance		Estimated as the total abundance of flower visitors, measured in one 200 × 3 m transect per plot, done in May 2008	Krauss, Klein, Weiner, Werner, Blüthgen
Bird diversity		Measured as the cumulative species richness estimated by audio-visual point-counts, done in May-June 2008-2010	Renner, Böhm, Tschapka
Flower cover		Measured as the number of inflorescences in four 50 × 3 m transects per plot. Flower area for each species was obtained from the literature.	Binkenstein, Schaefer

Note that for some groups the taxonomic unit was either operational taxonomic units (OTU: fungi and protists) or families (bacteria and below-ground insect larvae). Abundance measures were: per cent cover (plants, bryophytes), number of individuals captured (arthropods) and relative proportion of sequence reads assigned to each family among all reads within each plot (protists, soil bacteria and mycorrhiza).

RESEARCH LETTER

Extended Data Table 3 | Correlations between diversity predictors from the models in the main text

Models using residuals after controlling for environmental predictors

		Belowground herbivores	Belowground predators	Detritivores	Aboveground herbivores	Soil microbial decomposers	Bacterivores	Aboveground predators	Plant symbionts	Primary producers	Belowground herbivores	Belowground predators	Detritivores	Aboveground herbivores	Soil microbial decomposers	Bacterivores	Aboveground predators	Plant symbionts
Abundance	Primary producers	-0.06	0.09	0.03	-0.07	-0.01	0.01	-0.13	-0.05	0.06	0.06	-0.05	0.05	-0.06	0.10	-0.02	-0.04	-0.16
	Belowground herbivores		0.21	0.08	0.19	0.00	0.18	-0.02	-0.02	-0.15	0.37	0.10	0.13	0.08	0.06	0.18	-0.05	0.10
	Belowground predators			0.14	-0.16	0.05	0.09	-0.23	-0.04	-0.16	0.22	0.54	0.12	0.09	0.20	0.15	-0.09	0.12
	Detritivores				0.12	-0.10	-0.02	0.20	-0.02	0.04	0.15	0.09	0.74	0.12	-0.13	0.02	0.03	-0.15
	Aboveground herbivores					0.01	0.16	0.32	-0.05	-0.05	0.05	0.08	0.22	0.54	-0.04	0.14	0.34	-0.06
	Soil microbial decomposers						0.27	-0.02	0.05	0.00	0.01	0.04	-0.05	-0.11	0.17	0.25	-0.05	0.08
	Bacterivores							-0.13	0.06	-0.24	0.11	0.12	0.00	-0.10	0.13	0.76	-0.14	0.25
	Aboveground predators								-0.04	0.14	0.02	-0.11	0.14	0.38	-0.16	-0.18	0.59	-0.18
	Plant symbionts									-0.06	0.03	-0.10	0.04	-0.12	0.03	-0.06	0.03	0.46
Richness	Primary producers										-0.13	-0.13	-0.09	0.16	-0.38	-0.26	0.02	-0.22
	Belowground herbivores											0.24	0.18	0.11	-0.03	0.09	0.11	0.13
	Belowground predators												0.08	0.13	0.19	0.25	-0.02	-0.07
	Detritivores													0.19	-0.12	0.02	0.09	-0.08
	Aboveground herbivores														0.02	-0.03	0.48	-0.15
	Soil microbial decomposers															0.21	-0.06	0.17
	Bacterivores																-0.17	0.21
	Aboveground predators																	-0.17

Models including environmental predictors

		Soil depth	Fertilisation	Mowing	Grazing	Elevation	TWI	Primary producers	Belowground herbivores	Belowground predators	Detritivores	Aboveground herbivores	Soil microbial decomposers	Bacterivores	Aboveground predators	Plant symbionts	Primary producers	Belowground herbivores	Belowground predators	Detritivores	Aboveground herbivores	Soil microbial decomposers	Bacterivores	Aboveground predators	Plant symbionts
Environment	pH	0.05	0.07	0.05	-0.17	-0.24	0.22	-0.19	-0.22	-0.09	0.12	-0.21	-0.40	-0.30	0.25	-0.03	0.08	-0.18	0.00	0.21	-0.07	-0.05	-0.35	0.21	-0.07
	Soil depth		-0.21	0.02	-0.01	-0.82	0.52	-0.28	-0.22	-0.11	0.22	-0.29	-0.15	-0.06	-0.10	0.60	-0.62	-0.16	0.20	0.26	-0.39	0.42	0.10	-0.02	0.37
	Fertilisation			0.61	-0.22	0.15	0.00	0.01	-0.06	0.22	0.06	-0.06	0.00	0.10	0.01	-0.16	-0.23	-0.04	0.09	0.13	-0.06	-0.08	-0.01	-0.10	-0.06
	Mowing				-0.68	0.05	0.18	-0.03	-0.09	0.19	0.09	-0.15	-0.17	-0.01	-0.04	-0.12	-0.40	-0.13	0.10	0.09	-0.24	-0.02	-0.04	-0.17	-0.16
	Grazing					-0.09	-0.13	0.01	0.04	-0.02	-0.04	0.16	0.26	0.27	0.01	0.11	0.13	0.12	0.14	-0.04	0.11	0.07	0.32	0.12	0.25
	Elevation						-0.55	0.31	0.18	0.11	-0.28	0.29	0.22	0.05	-0.04	-0.61	0.60	0.19	-0.20	-0.33	0.39	-0.32	-0.13	-0.10	-0.42
	TWI							-0.19	-0.20	-0.02	0.15	-0.28	-0.36	-0.19	0.09	0.35	-0.50	-0.11	0.19	0.27	-0.28	0.41	-0.08	0.14	0.10
Abundance	Primary producers								-0.06	0.13	-0.06	0.22	0.06	-0.08	-0.27	-0.21	0.12	0.06	-0.15	-0.11	0.10	-0.10	-0.13	-0.14	-0.41
	Belowground herbivores									0.26	-0.02	0.11	0.26	0.28	-0.05	-0.08	0.19	0.48	0.08	-0.03	0.22	-0.09	0.24	-0.12	0.20
	Belowground predators										0.09	-0.01	0.08	0.14	-0.19	-0.09	-0.08	0.24	0.51	0.06	0.09	0.09	0.13	-0.13	0.05
	Detritivores											0.02	-0.15	-0.05	0.16	0.15	-0.16	0.09	0.20	0.88	0.01	-0.06	0.02	0.08	0.03
	Aboveground herbivores												0.07	0.06	0.23	-0.25	0.17	0.05	-0.06	-0.01	0.64	-0.14	0.03	0.31	-0.29
	Soil microbial decomposers													0.47	-0.18	0.05	0.17	0.17	0.01	-0.21	0.09	-0.11	0.46	-0.18	0.32
	Bacterivores														-0.17	0.05	-0.07	0.15	0.12	-0.08	-0.07	0.08	0.82	-0.20	0.40
	Aboveground predators															-0.11	0.23	-0.03	-0.09	0.18	0.32	-0.13	-0.15	0.68	-0.07
	Plant symbionts																-0.38	0.01	0.06	0.19	-0.29	0.28	0.13	-0.05	0.64
Richness	Primary producers																	0.14	-0.22	-0.24	0.42	-0.31	-0.13	0.14	-0.14
	Belowground herbivores																		0.22	0.08	0.18	-0.08	0.11	0.03	0.19
	Belowground predators																			0.17	-0.01	0.27	0.20	-0.02	0.12
	Detritivores																				0.04	-0.01	-0.05	0.15	0.08
	Aboveground herbivores																					-0.17	-0.09	0.41	-0.22
	Soil microbial decomposers																						0.14	-0.04	0.24
	Bacterivores																							-0.18	0.46
	Aboveground predators																								-0.09

Correlations between residuals (after controlling for the effect of study region, soil type, pH, topographic wetness index and the three land-use intensity components: fertilization, mowing and grazing) of abundance and species richness of the nine different trophic groups considered (top) or of the raw data (bottom). Those predictors removed owing to multicollinearity problems are shaded grey, with the correlation responsible highlighted. TWI = topographic wetness index, obtained from P. M., unpublished data.

Extended Data Table 4 | Model selection

		Biomass	Forage quality	Potential nitrification	Root biomass	Root decomposition	Phophorus retention	Mycorrhizal colonization	Soil aggregate stability	Soil C	Pest control	Resistance to pathogen	Pollinator abundance	Bird diversity	Flower cover	Multifunctionality 25%	Multifunctionality 50%	Multifunctionality 25%	Multifunctionality 90%
Abundance	Primary producers	3.26	-1.92	-1.09	-1.08	-2.34	0.59	-1.90	-2.00	-1.48	-0.35	6.09	-1.84	-1.99	3.53	-1.48	-1.76	-1.48	-1.69
	Belowground herbivores	-1.85	-0.50	1.35	-2.00	-1.88	-0.13	-0.02	-1.99	-1.70	-2.00	1.88	0.13	-0.10	-1.24	-0.65	-2.00	-0.98	-1.95
	Belowground predators	-1.90	0.03	-1.95	5.88	-1.97	-1.75	3.10	-0.70	-2.00	-1.74	-1.62	-1.71	-0.19	-1.98	-1.90	-1.34	-0.30	-0.06
	Aboveground herbivores	-0.74	-1.71	11.71	0.94	-0.46	-1.96	4.83	3.71	3.26	-1.96	-0.77	-1.97	-2.00	1.29	1.46	-0.03	-0.63	-1.74
	Soil microbial decomposer	-1.57	-0.86	0.38	-0.20	-0.14	-1.29	5.10	7.13	1.45	-1.99	0.66	-0.75	1.55	-1.99	-2.00	1.68	-1.98	-1.80
	Aboveground predators	0.74	-1.25	1.18	4.38	-1.84	-1.88	-1.94	2.68	0.48	1.79	6.41	-1.97	2.10	-1.81	-0.25	0.59	2.23	-1.28
	Plant symbionts	-0.80	-1.22	5.99	0.55	-0.39	1.82	-0.93	1.39	0.03	0.49	-1.96	2.92	-1.85	-1.96	-1.87	-0.97	-1.83	-2.00
Richness	Primary producers	14.90	5.43	12.00	1.94	0.13	16.91	-0.32	4.28	-0.97	5.01	1.56	-1.74	33.27	-1.66	-1.83	-1.35	6.88	2.45
	Belowground herbivores	-0.09	-1.95	-1.93	-0.23	-1.92	-1.94	-1.99	2.13	-1.61	-0.71	1.13	-2.00	-1.22	-1.99	-1.51	-1.88	-0.61	-1.74
	Belowground predators	-0.08	-1.78	-0.59	-1.65	-1.97	-0.56	-1.83	-1.65	-0.25	-1.59	-0.04	-1.84	-1.98	-1.02	-1.76	-1.53	-1.47	0.04
	Detritivores	0.74	-1.89	2.56	-1.79	-1.63	-1.22	-1.15	-1.35	-2.00	-1.32	-1.68	-1.63	-1.63	-0.44	-1.32	-1.66	-1.70	-1.64
	Aboveground herbivores	5.68	5.14	-1.85	6.91	-1.99	-1.67	-1.66	-1.66	0.10	-1.94	-0.65	-1.57	5.90	6.40	-1.49	-0.76	-1.98	-1.98
	Soil microbial decomposer	-1.97	-0.80	0.47	-3.86	-1.77	-1.20	4.95	-1.41	-1.37	-1.99	-1.54	-1.66	-1.66	-1.63	-1.92	0.01	-0.43	-0.77
	Bacterivores	-1.93	2.98	-0.86	-1.55	-1.80	0.74	-1.84	3.58	-1.75	-1.58	-1.91	1.44	-1.87	-1.90	-1.56	-1.33	-1.05	-1.90
	Aboveground predators	2.31	-1.93	-1.43	5.79	-1.81	10.25	-0.23	-1.97	6.17	3.29	-0.75	-1.38	-1.88	-1.34	-0.53	-2.00	-0.48	-1.60
	Plant symbionts	-1.91	-1.87	-1.91	-0.35	-1.58	-1.36	-1.39	-1.38	0.55	-1.99	2.64	7.34	-2.00	-2.00	-1.32	-2.00	-0.03	1.27

Difference in AIC when subtracting each term regarding the full model according to the backward step AIC procedure used (models using the environmental-corrected residuals, as presented in Fig. 1 and Extended Data Fig. 1). Green shade indicates the terms included in the most parsimonious models. Orange shade indicates terms included in the model with the lowest AIC but further removed using *F*-ratio tests.

Chapter 5 – Methodological studies on estimates of abundance and diversity of heterotrophic flagellates from the deep-sea floor

Journal of *Marine Science and Engineering*

MDPI

Review

Methodological Studies on Estimates of Abundance and Diversity of Heterotrophic Flagellates from the Deep-Sea Floor

Alexandra Schoenle, Alexandra Jeuck, Frank Nitsche, Paul Venter, Dennis Prausse and Hartmut Arndt *

Department of General Ecology, Institute for Zoology, Biocenter Cologne, University of Cologne, Zülpicher Straße 47b, Cologne D-50674, Germany; aschoenl@uni-koeln.de (A.S.); alexandra.jeuck@uni-koeln.de (A.J.); fnitsche@uni-koeln.de (F.N.); pventer@uni-koeln.de (P.V.); praussedennis@yahoo.de (D.P.)
* Correspondence: hartmut.arndt@uni-koeln.de; Tel.: +49-221-470-3100; Fax: +49-221-470-5932

Academic Editor: Angelika Brandt
Received: 30 September 2015; Accepted: 15 February 2016; Published: 3 March 2016

Abstract: Extreme environmental conditions in the deep sea hamper access to protist communities. In combination with the potentially highly diverse species composition, it demands a wide range of methods to be applied at the same time to guarantee a high resolution of quantitative and qualitative studies of deep-sea heterotrophic flagellates (HF). Within this study, we present a possible combination of several culture-independent and culture-dependent methods available for investigating benthic deep-sea HF communities. Besides live-counting and fixation of HF, we refer to cultivation methods and molecular surveys using next generation sequencing. Laboratory ecological experiments under deep-sea conditions (high pressure, low temperature) could allow the approval of the potential deep-sea origin of sampled HF. The combination of different methods offers a unique possibility to receive detailed information on nanofaunal life in the deep sea. Specific fixation techniques to preserve samples directly at the sampling depth must be applied in further studies to reflect the real biodiversity of the largest habitat on earth.

Keywords: live-counting; liquid-aliquot; fixation; next generation sequencing; pressure; deep-sea nanofauna

1. Introduction

Although deep-sea ecosystems represent the largest and most remote biome of the Earth [1], only about 5% have been explored so far, even less have been sampled [2]. We lack a firm understanding of species-level distribution (cosmopolitan *vs.* local) for deep-sea communities [3] as well as the functioning of and the interactions between biodiversity and ecological processes in this vast environment [2]. Marine heterotrophic flagellates contribute not only a major part within the microbial food web and are important nutrient remineralizers in biogeochemical cycles in surface waters [4,5] with densities ranging between 10^2 and 10^4 cells mL^{-1}, but are also potentially important regarding material flux and bacterial consumption in the deep sea. Recent studies of microbes have shown that even the deepest parts of our Earth are populated with a large variety of life [6]. Nevertheless, qualitative studies from eukaryotic deep-sea communities concerning diversity, spatial distribution and ecological function are still scarce. Concerning investigations of the bathypelagic deep sea, analysis revealed the occurrence of heterotrophic protists (11 ± 1 cells mL^{-1}) at depths down to 4000 m [7]. A global survey of bathypelagic microbial eukaryote communities identified a few groups as the dominant part of deep-sea communities, whereas the proportional composition of the dominant groups varied on a global scale [8]. Until the end of the last century, besides morphology based studies of foraminiferans [9], only anecdotal reports for other protists existed [10,11] mainly due to

methodological issues. While some authors reported a lack of flagellate occurrence in samples from the deep-sea floor [11], other studies revealed densities of HF [10,12,13] up to 10^5 cells cm^{-3} [14,15].

There is no standardized protocol for the sampling and analysis of benthic deep-sea protists available at present. Cultivation-based methods miss a majority of taxa since most species require specific cultivation conditions [16]. Molecular barcoding approaches employing PCR introduce significant biases in reported community compositions of marine protists due to the restriction of 'general' primers to detect all protist groups [17]. Since molecular surveys cannot yet provide any information on the morphology and abundance of the organisms, culture-dependent and culture-independent investigations are required to gain quantitative and qualitative results concerning deep-sea protist biodiversity.

The aim of this review is the study of benthic flagellated protists, those living in deep-sea sediments/seafloor. We will illustrate the above mentioned methodological problems presenting own recent results and will recommend a combination of methods to get a more reliable estimate of deep-sea benthic nanofauna.

2. Quantification and Qualification of Deep-Sea Protists

Several methods have been applied during the last decades to characterize deep-sea communities. Main procedures for flagellate detection in the past included live-counting of samples immediately after sampling, occurrences in laboratory cultures and molecular surveys using Sanger or next generation sequencing (NGS) (Figure 1). However, the taxonomic identification of protists, especially nanoprotists, in routine samples is difficult due to the general lack of conspicuous morphological features and the selectivity of sampling and counting methods [18–20]. The taxonomic identity of heterotrophic flagellates is generally based on cultivated strains, on which ultrastructural, physiological and molecular studies have been performed [16,21]. However, most deep-sea organisms are extremely difficult to cultivate due to their slow *in-situ* growth rates and their likely strict adaptation to extreme environmental conditions (oligotrophy, low temperatures, high pressure, anoxia) [2]. The role of these cultured strains as representatives within deep-sea protist communities is unclear. Molecular surveys frequently recover novel eukaryotic lineages that have not been recorded from cultures so far [16,22]. Environmental molecular surveys in microbial ecology have revolutionized our knowledge, indicating how far we are from understanding this "untapped reservoir" [23] of microbial diversity in the depth [19,22,24]. A major problem up to now is the assignment of these obtained sequences to species level with existing databases. A better annotation will improve the knowledge that comes from such analysis and sequence libraries. Therefore, such data will become more valuable as better gene annotations become available [25]. Molecular environmental diversity studies of the deep-sea floor have mainly been focusing on assumed "hot spots" of activity (e.g., hydrothermal vents, methane seeps) mostly from the bathyal zone carried out on a local scale [26–28]. Our previous studies of deep-sea nanofauna [29–33] indicated the existence of a specific abyssal nanofauna which contains a large number of endemic taxa [31,33].Recent comprehensive studies [34] indicated protists as the most diverse eukaryotic organisms. The diversity of phyla (Figure 1) with their specific differences (e.g., ultrastructure) makes it necessary to consider specifically designed fixatives or molecular techniques.

2.1. How to Sample Deep-Sea Protists

The main tool used up to now to collect benthic deep-sea protist communities is the Multi-Corer system. Due to a closing mechanism at the top and bottom of the cores, the risk of contamination with organisms and cysts from upper water layers is reduced. However, the problem is that samples have to be treated immediately after sampling which means within minutes. We microscopically observed living nanoflagellates within the first 30 min after sampling. Protists are stressed by tremendous physical changes, e.g., varying pressures and temperatures, during sampling. Therefore, it is likely that several flagellate species adapted to deep-sea conditions die, while being raised through the

J. Mar. Sci. Eng. **2016**, *4*, 22

water column. Morgan-Smith *et al.* [35] sampled deep-sea protists with 200 mL titanium chambers retaining *in situ* pressure from depths of 2750 and 4000 m to investigate the effect of pressure on protist abundances prior to fixation. Although depressurization pre-versus post fixation did not significantly affect the number of eukaryotes counted, cell physiology might be greatly impacted by changes in pressure. Future methodological studies must be applied to solve these problems. Potential solutions could be the usage of specific fixations of samples already in the depth of sampling. Furthermore, samples might generally be obtained under pressure in special containers to ensure observation of living flagellates under prevailing environmental conditions.

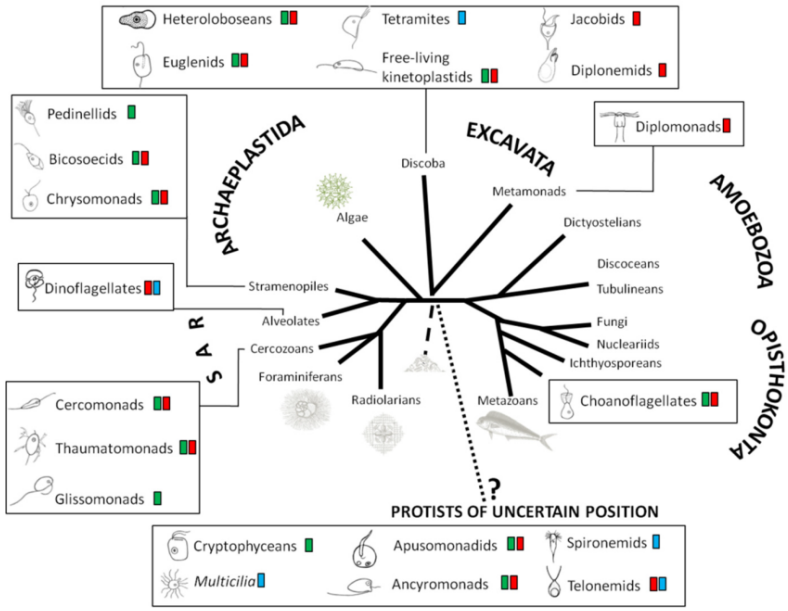


Figure 1. Taxonomic composition of heterotrophic flagellate groups (according to [25]) known from the deep sea. Colors indicate different kinds of quantification and/or qualification methods used for detection: Live-counting (blue), cultures (green), molecular surveys (red) (compilation of literature and own results; scheme derived from [33]).

2.2. Live-Counting

Live-counting techniques offer the opportunity to detect living cells up to the morphospecies level in addition to quantitative estimates. Although this method is difficult regarding a limited available time frame for observation and the need of a high amount of expertise, it is useful for obtaining high taxonomic morphotype resolutions [36].

Generally, untreated sediment samples are stored on ice and used to detect living flagellates immediately after sampling. The direct counts can serve as an estimate of deep-sea protistan abundance and as a cultivation-independent record of species. Inspections and counting of 5–10 µL subsamples of sediment suspensions can be conducted using light microscopes (40–63× phase-contrast objectives) combined with video recording [20]. However, it has to be considered that several flagellates die under the microscope during counting, probably caused either by rising temperatures due to microscopic light exposure or exposure to low atmospheric pressure (1 bar). These observations also underline the limitation of culture-dependent studies discussed in more detail in section 2.4. Due to the fact that only a few individuals can be detected within this short time frame after sampling, the low abundances lead to possibly severe underestimations of actual protist abundances as can be seen by comparing the numbers obtained from live-counts with those obtained from the analysis of fixed and stained samples (Figure 2). On the other hand, counts of fixed samples could overestimate real abundances when not fluorescence in-situ hybridization techniques are applied [35], but unspecifically binding

J. Mar. Sci. Eng. **2016**, *4*, 22

fluorochromes are used which may also stain free-floating nuclei and other DNA containing particles. An advantage of live-counting and observation is that the presence of living specimens of genotypes only known from clone libraries and metagenomic studies can be verified. In addition, new taxa can be detected.

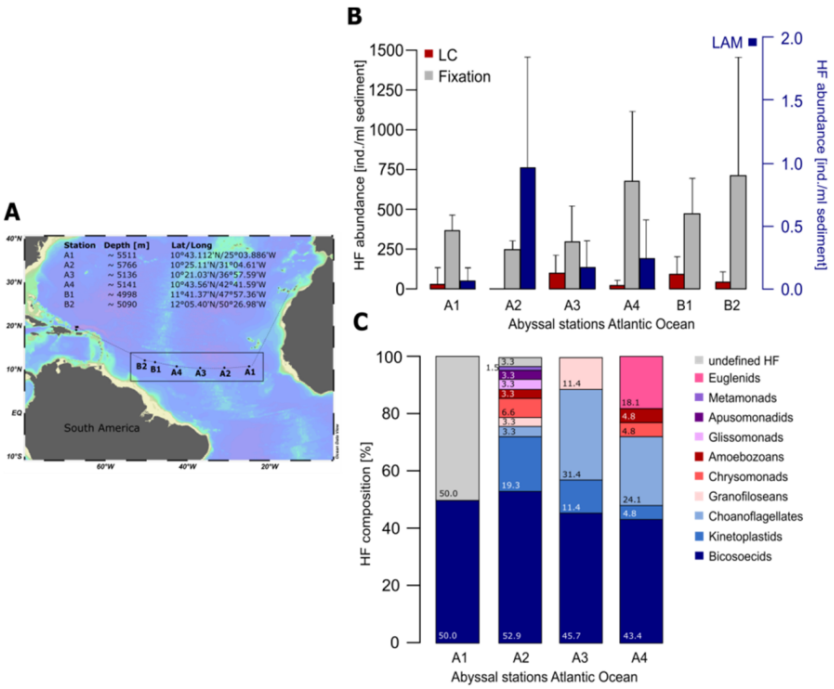


Figure 2. Comparison of methods applied for investigations of deep-sea heterotrophic flagellates in the VEMA fracture zone, southern North Atlantic. (**A**) Station map (created with Ocean Data View [37]) of the research cruise with *R/V Sonne II* (SO 237, 14.12.2014–26.01.2015). Sampling stations are indicated by black dots and station labelling (A1–A4, B1, B2). (**B**) Mean heterotrophic flagellates (HF) abundance (n = 3) of live, fixed and cultivation (liquid aliquot method, LAM) counts (ind./mL sediment). LAM counts were plotted with a separate y-axis. (**C**) Percentage of taxonomic HF group composition for stations A1–A4 revealed with LAM within the first 2–4 weeks. Live-counting: Inspections and counting of 5–10 µL subsamples of sediment suspensions was conducted using light microscopes (40–63× phase-contrast objectives) combined with video recording. Fixation: Sediment subsamples were fixed with formaldehyde (2%), stained with DAPI (4′,6-Diamidin-2-phenylindol, Sigma-Aldrich, Munich, Germany) and filtered on 0.2 µm membrane filters. Following criteria were defined for the detection of flagellates: roundish shape, larger than 1.5 µm and clear blue coloration. Cultivation (LAM): Subsamples of 2 mL of the sediment suspension were cultivated in 50 mL tissue-culture flasks (Sarstedt, Nümbrecht, Germany) filled with 30 mL autoclaved sea water. Additionally, 650 mL culture flasks were filled with 400–500 mL overlaying water.

2.3. Fixation and Staining

Fixation and staining methods are advantageous due to the possible long-term storage and observation of samples. Generally, glutaraldehyde (1%) or formaldehyde (1%–2%) are used as fixatives combined with staining fluorochromes (e.g., DAPI, FITC, Proflavine) which bind to cell components such as DNA or proteins [38–40] to detect potentially eukaryotic cells under epifluorescent microscopes [41]. Morgan-Smith et al [35] suggested the fixation of deep-sea water column samples overnight at room temperature followed by filtration on polycarbonate filters at a vacuum of −200 mbar to ensure the escape of supersaturated gases and, thus, avoid bubble formation on filters. Hondefeld *et al.* [42] proposed a suitable method to detect protists in marine sediment samples resuspending fixed samples and taking subsamples of the supernatant after a few minutes when

the majority of inorganic particles had been settled, alternative methods could be density gradient centrifugation to separate protists from inorganic particles [43,44]. Although fixed counts are up to three orders of magnitude larger than live counts (see Figure 2), no methodological tests of the reliability of this method exist for deep-sea conditions [30]. Quantifying eukaryotic cells in fluorescently stained fixed samples is accompanied by several disadvantages. Critical comparisons of fixed samples of heterotrophic flagellates have found significant discrepancies between parallel counts of each other (*cf.* Figure 2). One has to keep in mind that obtained abundances might be underestimated due to the fact that cells may not survive the pressure changes during sampling. Several groups of HF are very sensitive to the fixation procedure. A significant part of HF might be disrupted by fixation and difficult to detect [45]. Although several authors emphasize the more accurate identification of protists with DAPI-staining due to the discrimination between the nucleus and cytoplasm and sometimes the display of flagella, a clear identification of all protists is still limited and in several cases doubtful, because large bacteria [7] and free-floating nuclei might also be stained. Thus, staining the nucleus with DAPI in combination with FITC [39,40] or Proflavine [38], which stain entire cell body, seem to be a more accurate way to explicitly identify flagellates.

One improvement in detecting protists is the usage of fluorescence-*in-situ*-hybridization (FISH). Although FISH is a huge development in identifying microbial eukaryotes as well as prokaryotes, there are still several disadvantages of FISH such as insufficient sensitivity due to the low number of target molecules in cells, low probe permeability of cells and poor probe hybridization efficiency [46]. The catalyzed reporter deposition fluorescence-*in-situ*-hybridization (CARD FISH) allows characterization of communities in terms of abundance and taxonomy and specifically targets protists, while large bacteria are not confounded [47]. This technique has already been used for analyzing eukaryotic deep-sea microbes together with universal oligonucleotide probes (e.g., EUK516) [35]. The universal probes EUK516 (5′-ACCAGACTTGCCCTCC-3′, [48,49]) and EUK1209 (5′-GGGCATCACAGACCTG-3′, [50,51]) are missing the detection of kinetoplastids within the eukaryotic phylogenetic tree. Thus, the exclusive usage of these two probes would lead to a lack of detection of some free-living protists in at least some marine systems. The overall specificity and reliability of the detection of protists can be increased with a combination of oligonucleotide probes KIN516 (5′-ACCAGACTTGTCCTCC-3′, [52]) and EUK516.

2.4. Cultivation

Cultivation methods offer the possibility of detailed morphological characterizations and the establishment of clonal cultures for molecular studies. Water originating from the sampling depth is autoclaved and bacterial growth is supported by adding organic substances (e.g., yeast extract, glucose) to allow cultivation of bacterivorous species. Generally, not all species appear in cultures due to selective conditions like enrichment of bacteria or the lack of suitable other food sources (e.g., other protists). This results in a support of r-strategists among HF favoring similar genera/species such as *Cafeteria, Caecitellus, Rhynchomonas, Neobodo* during cultivation [53]. However, sometimes even seldom recorded species may appear, showing that a massive cultivation effort is needed to enhance successful cultivation. To partially overcome this problem, molecular investigations such as next generation sequencing are applied to detect uncultivable organisms.

One suitable method of cultivating protists is the liquid aliquot method (LAM, [54]) inoculating defined aliquots small enough to place one cultivable organism into each culture vessel. Aliquots of deep-sea sediment or overlaying water can be cultivated to estimate the abundance and diversity of cultivable deep-sea nanoprotists (Figure 3B,C).

2.5. Next Generation Sequencing (NGS)

Molecular surveys have revolutionized our understanding of deep-sea protist communities. The methodological spectrum of next-generation sequencing (NGS) and DNA-barcoding for HF has increased significantly in the last years [17,34]. Conserved samples for bulk analysis of RNA (active

J. Mar. Sci. Eng. **2016**, *4*, 22

organisms, metatranscriptomics) and DNA (whole metagenome studies) can be used to analyze the presence of protist genotypes in the deep sea [31,33]. However, there are still some unsolved problems like specific instead of general primers, different rRNA copy numbers for protists, PCR biases, the difficulty of differentiating active from inactive forms (e.g., cysts), and incomplete databases containing incorrect labeled species [45,55,56].

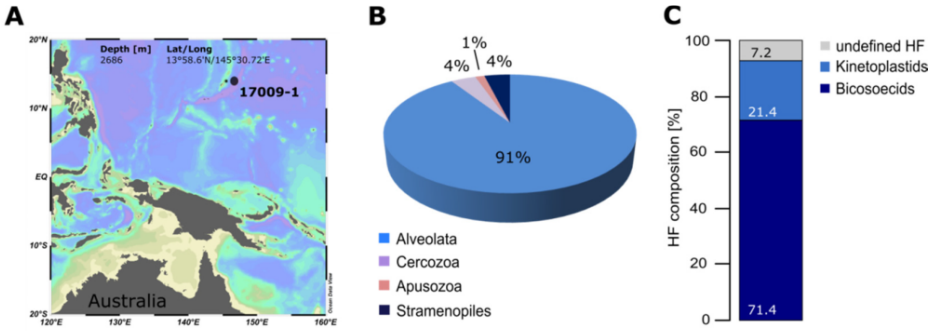


Figure 3. Comparison of methods applied for investigations of deep-sea heterotrophic flagellates in the Mariana Basin, Central Northern Pacific. (**A**) Position of the sampling station (2686 m depth), *R/V Sonne I* (SO223T, 09.09.2012–10.10.2012). Created with Ocean Data View [37]. (**B**) Percentage contribution of sequence reads of HF groups obtained by 454 sequencing. (**C**) Percentage contribution of HF groups revealed with the liquid aliquot method. NGS: Whole genomic DNA extracted from sediment samples using the PowerSoil® DNA isolation kit (Mo Bio Laboratories, Inc., Carlsbad, CA, USA) and quantified using a spectrophotometer. The highly variable V4 region of the 18S rRNA gene was directly amplified from the samples using the eukaryotic specific primers 590F (5′-CGGTAATTCCAGCTCCAATAGC-3′) and 1300R (5′-CACCAACTAAGAACGGCCATGC-3′). Each sample was PCR'ed in triplicate and pooled to a final concentration of 20 ng/µL to reduce possible PCR biases. NGS using the GS-FLX Titanium sequencer (Roche, Mannheim, Germany) was performed by GATC Biotech AG, (Cologne, Germany). Sequencing was done as from adaptor A (forward primer or 5′-end). Obtained sequences (100% query coverage) were clustered in OTUs using a pairwise identity of 80% on the "class" level, since most reads did not yet have hits in public databases. Cultivation: see Figure 2.

While the usage of quantitative PCR of 18S rRNA genes in conjunction with FISH for marine picoeukaryotes, seemed to be a very promising way to quickly obtain data on the ecological distribution of important phytoplankton groups, primer specificity and varying rRNA gene copy numbers among eukaryotes need to be considered [57]. The potentially selective amplification needs to be incorporated in the interpretation of obtained results concerning species composition and abundances. Thus, the use of multiple sets of primers is required to recover the major part of environmental microbial diversity. Comparison of artificial and environmental 18S rRNA gene libraries revealed, that environmental PCR-based techniques might be sufficient to compare samples, but the total diversity will probably always be underestimated [58]. High amounts of ribosomal sequence data can be obtained by next generation sequencing (e.g., 454-pyrosequencing, Illumina), which has the potential to uncover more organisms including rare species. Both methods, 18S clone libraries and 18S amplicon sequencing, showed significant similarities in protist community composition [59].

An alternative which does not require PCR steps is the metagenomic approach. The analysis of bulk DNA from deep sea-sediments [3] allows for both a qualitative analysis and a rough assignment to trophic functions of deep-sea nanofauna. However, it has to be considered that a large proportion (estimations up to 90%) of the DNA in deep-sea sediments is extracellular [60,61]. Thus, it is uncertain, if detected benthic protist communities are actually thriving under these conditions or are rather an artifact by deposited cells from the upper water column, encysted cells or extracellular DNA [62]. Thus, metagenomics might introduce biases in actual protist biodiversities, because they are accompanied by two major issues, rDNA copy number and extracellular DNA [63]. One major bias of

rDNA diversity surveys, the extracellular rDNA, is reduced significantly by rRNA libraries, but such libraries are exclusively recovering the active part of the communities. A solution might be the RNA and DNA extraction from the same sample to assess the composition of the microeukaryotic assemblage by distinguishing between active cells and signals from inactive or even dead organisms [62]. A comparison with transcriptome data from similar sampling sites in the deep sea could help to detect a "passive seed bank" which might contain species which are able to grow in the sampled habitat, but might be inactive due to actually unfavorable conditions. Investigations of sympagic as well as surface protist communities revealed activity patterns of specific groups by comparing rDNA and rRNA libraries [64,65].

The results from our study of deep-sea sediments from the central Pacific indicate the advantages and disadvantages of molecular and cultivation methods. Cultivation recovered only a very minor part in comparison with the diversity obtained by NGS. Cultivation (and Sanger sequencing) allowed the assignment of sequences to species level (Figure 3). Within our studies 91% of all NGS sequences belonged to alveolates (undetermined dinoflagellates) which amplify preferentially with universal primers, whereas they did not occur in cultures. The same was true for cercozoans. Contrary, kinetoplastids were not recovered by NGS due to primer mismatch, but could be detected in cultures (new species). Bicosoecids (new species) occurred in cultures but were obviously too seldom for registration by NGS.

3. Protocol for Detecting Nanofaunal Abundance and Diversity

Estimates of abundance and diversity should be accomplished by culture independent methods such as live-counting of untreated samples as well as counting of fixed and stained samples. Furthermore, cultivation of defined aliquots of the diluted sample (LAM) offer the possibility of morphological characterization and later molecular surveys (PCR, single-cell genomics/transcriptomics) for identifying corresponding genotypes. This addition of known sequences to molecular database is a very important step to increase the knowledge of diversity of protists in the deep ocean. To get an idea regarding the active genotypes in deep-sea samples, NGS applied to RNA is necessary. Clone libraries or next generation sequencing are helpful tools to detect diversity but the results must be verified regarding the origin of the organisms. From an ecological point of view, pressure (>200 bar) and temperature (<4 °C) experiments may confirm the deep-sea origin of sampled HF [66]. At least for some organisms isolated from the deep sea it should be tested in the laboratory whether they are able to survive at deep-sea conditions. Thus, a combination of several methods is recommended when analyzing deep-sea nanofauna (Figure 4).

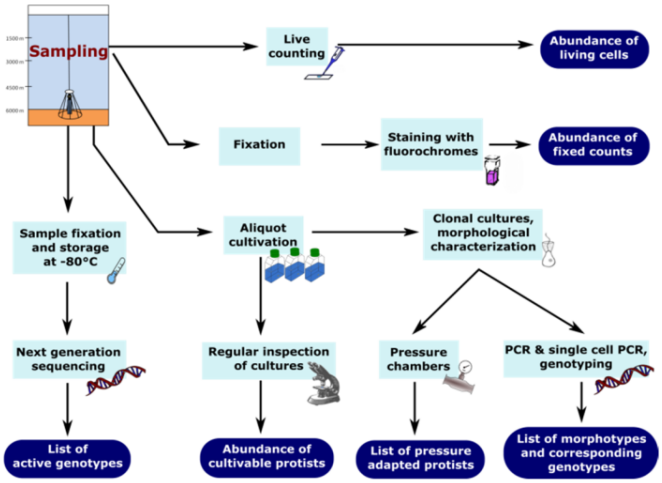


Figure 4. Proposed protocol for diversity and abundance estimates of deep-sea nanoprotists including molecular surveys (such as environmental RNA), fixation, live-counting and aliquot cultivation.

4. Conclusions

Molecular methods are an appropriate way to investigate deep-sea protistan diversity. With metagenome analysis and 18S rDNA amplicon sequencing, the active as well as inactive fraction of protist communities in the deep sea can be recorded. However, one has to keep in mind, that extracellular DNA is also detected. Thus, it is recommended to add analysis of the RNA (rRNA amplicon sequencing and metatranscriptomics) of the recorded genepool to filter for the active organisms. As DNA is well preserved in this environment and protists may form cysts in the deep sea because of unfavorable conditions, one should consider data from metagenomic analysis as a seed bank analysis. This way, a comparison between spatial and temporal separated samplings in the deep sea can be used to detect theoretically viable protists which were not active during sampling due to environmental factors such as lack of resources. To apply NGS for the analysis of species and hence biodiversity, the need of reliable reference databases is a major hindrance, which has to be overcome. A close combination of NGS together with culture dependent methods, morphological observations, single-cell investigations, as well as ecological studies is a prerequisite for a profound understanding of the diversity and the role of protists in deep-sea food webs. We tried to provide a recommendation of methods for investigating abundance and diversity of deep-sea nanoprotists by combining six different techniques available at present (Figure 4). Each method has its own advantages and disadvantages concerning investigations of HF. The combination of different methods offers a unique possibility to receive detailed information on nanofaunal life in this extreme and hardly accessible environment. However, the fact that protists are usually exposed to high variations in pressure and temperature during sampling procedures compared to their constant original environment may potentially lead to a disruption of flagellates. Future studies must solve these methodological problems. Therefore, the usage of specific fixations of samples already at the depth of sampling should be considered. For investigation concerning diversity and the ecological role of HF, samples should be obtained under pressure from the deep sea to ensure observation of living flagellates.

Acknowledgments: We are extremely grateful to Capt. Oliver Meyer and his crew for valuable help during sampling and the excellent support during the SO223-T *(R/V Sonne I)* and SO237 *(R/V Sonne II)* cruise. We thank the whole scientific crew from the research cruise SO237 and SO223-T for a great assistance during all kind of issues. We thank Rosita Bieg, Brigitte Gräfe and Bärbel Jendral for valuable technical support. We thank three anonymous reviewers for valuable suggestions and comments. This work is supported by two grants from BMBF (ProtAbyss-03G0237B & TransGeoBiOc-03G0828A) to H.A.

Author Contributions: All stated authors conceived, designed and carried out the experiments. Schoenle, A. and Arndt, H. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gage, J.; Tyler, P.A. *Deep-Sea Biology: A Natural History of Organisms at the Deep-Sea Floor*; Cambridge University Press: Cambridge, UK, 1991.
2. Danovaro, R.; Corinaldesi, C.; Rastelli, E.; Dell'Anno, A. Towards a better quantitative assessment of the relevance of deep-sea viruses, *Bacteria* and *Archaea* in the functioning of the ocean seafloor. *Aquat. Microb. Ecol.* **2015**, *75*, 81–90. [CrossRef]
3. Bik, H.M.; Sung, W.; De Ley, P.; Baldwin, J.G.; Sharma, J.; Rocha-Olivares, A.; Thomas, W.K. Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. *Mol. Ecol.* **2012**, *21*, 1048–1059. [CrossRef] [PubMed]
4. Caron, D.A.; Peele, E.R.; Lim, E.L.; Dennett, M.R. Picoplankton and nanoplankton and their trophic coupling in surface waters of the Sargasso Sea south of Bermuda. *Limnol. Oceanogr.* **1999**, *44*, 259–272. [CrossRef]
5. Sherr, E.B.; Sherr, B.F. Bacterivory and herbivory: Key roles of phagotrophic protists in pelagic food webs. *Microb. Ecol.* **1994**, *28*, 223–235. [CrossRef] [PubMed]
6. Jørgensen, B.B.; Boetius, A. Feast and famine-microbial life in the deep-sea bed. *Nat. Rev. Micro.* **2007**, *5*, 770–781. [CrossRef] [PubMed]

7. Pernice, M.C.; Forn, I.; Gomes, A.; Lara, E.; Alonso-Sáez, L.; Arrieta, J.M.; del Carmen Garcia, F.; Hernando-Morales, V.; MacKenzie, R.; Mestre, M.; *et al*. Global abundance of planktonic heterotrophic protists in the deep ocean. *ISME J.* **2015**, *9*, 782–792. [CrossRef] [PubMed]
8. Pernice, M.C.; Giner, C.R.; Logares, R.; Perera-Bel, J.; Acinas, S.G.; Duarte, C.M.; Gasol, J.M.; Massana, R. Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J.* **2015**. [CrossRef] [PubMed]
9. Gooday, A.J.; Rathburn, A.E. Temporal variability in living deep-sea benthic foraminifera: A review. *Earth Sci. Rev.* **1999**, *46*, 187–212. [CrossRef]
10. Burnett, B.R. Quantitative sampling of nanobiota (microbiota) of the deep-sea benthos-III. The bathyal San Diego Trough. *Deep Sea Res.* **1981**, *28*, 649–663. [CrossRef]
11. Lighthart, B. Planktonic and benthic bacterivorous protozoa at eleven stations in Puget Bay, Australia. *J. Fish. Res. Can.* **1969**, *26*, 299–304. [CrossRef]
12. Alongi, D.M. Bacterial growth rates, production and estimates of detrital carbon utilization in deep-sea sediments of the Solomon and Coral Seas. *Deep Sea Res.* **1990**, *37*, 731–746. [CrossRef]
13. Burnett, B.R. Quantitative sampling of microbiota of the deep-sea benthos-I. Sampling techniques and some data from the abyssal central North Pacific. *Deep Sea Res.* **1977**, *24*, 781–789. [CrossRef]
14. Bak, R.P.M.; Nieuwland, G. Seasonal variation in bacterial and flagellate communities of deep-sea sediments in a monsoonal upwelling system. *Deep Sea Res. II* **1997**, *44*, 1281–1292. [CrossRef]
15. Danovaro, R.; Marrale, D.; Della Croce, N.; Dell'Anno, A.; Fabiano, M. Heterotrophic nanoflagellates, bacteria, and labile organic compounds in continental shelf and deep-sea sediments of the Eastern Mediterranean. *Microb. Ecol.* **1998**, *35*, 244–255. [CrossRef] [PubMed]
16. Del Campo, J.; Balagué, V.; Forn, I.; Lekunberri, I.; Massana, R. Culturing bias in marine heterotrophic flagellates analyzed through seawater enrichment incubations. *Microb. Ecol.* **2013**, *66*, 489–499. [CrossRef] [PubMed]
17. Pawlowski, J.; Audic, S.; Adl, S.M.; Bass, D.; Belbahri, L.; Berney, C.; Bowser, S.S.; Cepicka, I.; Decelle, J.; Dunthorn, M.; *et al*. CBOL Protist working group: Barcoding eukaryotic richness beyond the animal, plant and fungal kingdoms. *PLoS Biol.* **2012**, *10*, e1001419. [CrossRef] [PubMed]
18. Jürgens, K.; Massana, R. Protistan grazing on marine bacterioplankton. In *Microbial Ecology of the Oceans*, 2nd ed.; Wiley: New York, NY, USA, 2008; pp. 383–441.
19. Massana, R.; Guillou, L.; Diez, B.; Pedrós-Alió, C. Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. *Appl. Environ. Microbiol.* **2002**, *68*, 4554–4558. [CrossRef] [PubMed]
20. Arndt, H.; Dietrich, D.; Auer, B.; Cleven, E.-J.; Gräfenhan, T.; Weitere, M.; Mylnikov, A.P. Functional diversity of heterotrophic flagellates in aquatic ecosystems. In *The Flagellates—Unity, Diversity and Evolution*; Leadbeater, B.S.C., Green, J.C., Eds.; Taylor & Francis Ltd: London, UK, 2000; pp. 240–268.
21. Nitsche, F.; Arndt, H. A new choanoflagellate species from Taiwan: Morphological and molecular biological studies of *Diplotheca elongata* nov. spec. and *D. costata*. *Eur. J. Protistol.* **2008**, *44*, 220–226. [CrossRef] [PubMed]
22. López-García, P.; Rodríguez-Valera, F.; Pedrós-Alió, C.; Moreira, D. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **2001**, *409*, 603–607. [CrossRef] [PubMed]
23. Arístegui, J.; Gasol, J.M.; Duarte, C.M.; Herndl, G.J. Microbial oceanography of the dark ocean's pelagic realm. *Limnol. Oceanogr.* **2009**, *54*, 1501–1529. [CrossRef]
24. López-García, P.; Philippe, H.; Gail, F.; Moreira, D. Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 697–702. [CrossRef] [PubMed]
25. Adl, S.M.; Simpson, A.G.B.; Lane, C.E.; Lukeš, J.; Bass, D.; Bowser, S.S.; Brown, M.W.; Burki, F.; Dunthorn, M.; Hampl, V.; *et al*. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **2012**, *59*, 429–493. [CrossRef] [PubMed]
26. Edgcomb, V.; Orsi, W.; Leslin, C.; Epstein, S.S.; Bunge, J.; Jeon, S.; Yakimov, M.M.; Behnke, A.; Stoeck, T. Protistan community patterns within the brine and halocline of deep hypersaline anoxic basins in the eastern Mediterranean Sea. *Extremophiles* **2009**, *13*, 151–167. [CrossRef] [PubMed]
27. Stoeck, T.; Epstein, S. Novel eukaryotic lineages inferred from small-subunit rRNA analyses of oxygen-depleted marine environments. *Appl. Environ. Microbiol.* **2003**, *69*, 2657–2663. [CrossRef] [PubMed]

28. Edgcomb, V.P.; Kysela, D.T.; Teske, A.; de Vera Gomez, A.; Sogin, M.L. Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7658–7662. [CrossRef] [PubMed]
29. Scheckenbach, F.; Wylezich, C.; Weitere, M.; Hausmann, K.; Arndt, H. Molecular identity of strains of heterotrophic flagellates isolated from surface waters and deep-sea sediments of the South Atlantic based on SSU rDNA. *Aquat. Microb. Ecol.* **2005**, *38*, 239–247. [CrossRef]
30. Arndt, H.; Hausmann, K.; Wolf, M. Deep-sea heterotrophic nanoflagellates of the Eastern Mediterranean Sea: Qualitative and quantitative aspects of their pelagic and benthic occurrence. *Mar. Ecol. Prog. Ser.* **2003**, *256*, 45–56. [CrossRef]
31. Scheckenbach, F.; Hausmann, K.; Wylezich, C.; Weitere, M.; Arndt, H. Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 115–120. [CrossRef] [PubMed]
32. Hausmann, K.; Selchow, P.; Scheckenbach, F.; Weitere, M.; Arndt, H. Cryptic species in a morphospecies complex of heterotrophic flagellates: The case study of *Caecitellus* spp. *Acta Protozool.* **2006**, *45*, 415–431.
33. Salani, F.S.; Arndt, H.; Hausmann, K.; Nitsche, F.; Scheckenbach, F. Analysis of the community structure of abyssal kinetoplastids revealed similar communities at larger spatial scales. *ISME J.* **2012**, *6*, 713–723. [CrossRef] [PubMed]
34. De Vargas, C.; Audic, S.; Henry, N.; Decelle, J.; Mahé, F.; Logares, R.; Lara, E.; Berney, C.; Le Bescot, N.; Probert, I.; *et al*. Eukaryotic plankton diversity in the sunlit ocean. *Science* **2015**, *348*. [CrossRef] [PubMed]
35. Morgan-Smith, D.; Herndl, G.; van Aken, H.; Bochdansky, A. Abundance of eukaryotic microbes in the deep subtropical North Atlantic. *Aquat. Microb. Ecol.* **2011**, *65*, 103–115. [CrossRef]
36. Jeuck, A.; Arndt, H. A short guide to common heterotrophic flagellates of freshwater habitats based on the morphology of living organisms. *Protist* **2013**, *164*, 842–860. [CrossRef] [PubMed]
37. Schlitzer, R. Ocean Data View. 2012. Available online: http://odv.awi.de (accessed on 28 August 2015).
38. Tanaka, T.; Rassoulzadegan, F. Full-depth profile (0–2000 m) of bacteria, heterotrophic nanoflagellates and ciliates in the NW Mediterranean Sea: Vertical partitioning of microbial trophic structures. *Deep Sea Res. II* **2002**, *49*, 2093–2107. [CrossRef]
39. Fukuda, H.; Sohrin, R.; Nagata, T.; Koike, I. Size distribution and biomass of nanoflagellates in meso-and bathypelagic layers of the subarctic Pacific. *Aquat. Microb. Ecol.* **2007**, *46*, 203. [CrossRef]
40. Paffenhöfer, G.-A. Abundance and distribution of nanoplankton in the epipelagic subtropical/tropical open Atlantic Ocean. *J. Plankton Res.* **2003**, *25*, 1535–1549. [CrossRef]
41. Sherr, E.B.; Caron, D.A.; Sherr, B.F. Staining of heterotrophic protists for visualization via epifluorescence microscopy. In *Handbook of Methods in Aquatic Microbial Ecology*; Taylor & Francis: London, UK, 1993; pp. 213–228.
42. Hondeveld, B.J.; Nieuwland, G.; Van Duyl, F.C.; Bak, R.P. Temporal and spatial variations in heterotrophic nanoflagellate abundance in North Sea sediments. *Mar. Ecol. Prog. Ser.* **1994**, *109*, 235–235. [CrossRef]
43. Epstein, S.S. Simultaneous enumeration of protozoa and micrometazoa from marine sandy sediments. *Aquat. Microb. Ecol.* **1995**, *9*, 219–227. [CrossRef]
44. Scherwass, A.; Wickham, S.A.; Arndt, H. Determination of the abundance of ciliates in highly turbid running waters—An improved method tested for the River Rhine. *Arch. Hydrobiol.* **2002**, *156*, 135–143. [CrossRef]
45. Jeuck, A.; Nitsche, F.; Wylezich, C.; Wirth, O.; Hennemann, M.; Nopper, N.; Bergfeld, T.; Monir, S.; Scherwass, A.; Arndt, H. A comparison of some methods to quantify heterotrophic flagellates of different taxonomic groups. In Proceedings of the VII ECOP-ISOP joint meeting, Sevilla, Spain, 5–10 October 2015.
46. Kubota, K. CARD-FISH for environmental microorganisms: Technical advancement and future applications. *Microbes Environ.* **2013**, *28*, 3–12. [CrossRef] [PubMed]
47. Pernthaler, A.; Pernthaler, J.; Amann, R. Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* **2002**, *68*, 3094–3101. [CrossRef] [PubMed]
48. Amann, R.I.; Ludwig, W.; Schleifer, K.-H. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microb. Rev.* **1995**, *59*, 143–169.
49. Beardsley, C.; Knittel, K.; Amann, R.I.; Pernthaler, J. Quantification and distinction of aplastidic and plastidic marine nanoplankton by fluorescence *in situ* hybridization. *Aquat. Microb. Ecol.* **2005**, *41*, 163–169. [CrossRef]
50. Biegala, I.C.; Not, F.; Vaulot, D.; Simon, N. Quantitative assessment of picoeukaryotes in the natural environment by using taxon-specific oligonucleotide probes in association with tyramide signal

amplification-fluorescence *in situ* hybridization and flow cytometry. *Appl. Environ. Microbiol.* **2003**, *69*, 5519–5529. [CrossRef] [PubMed]
51. Giovannoni, S.J.; De Long, E.F.; Olsen, G.J.; Pace, N.R. Phylogenetic group-specific oligodeoxynucleotide probes for identification of single microbial cells. *J. Bact.* **1988**, *170*, 720–726. [PubMed]
52. Bochdansky, A.B.; Huang, L. Re-Evaluation of the EUK516 Probe for the Domain Eukarya Results in a Suitable Probe for the Detection of Kinetoplastids, an Important Group of Parasitic and Free-Living Flagellates. *J. Eukaryot. Microbiol.* **2010**, *57*, 229–235. [CrossRef] [PubMed]
53. Patterson, D.J.; Lee, W.J. Geographic distribution and diversity of free-living heterotrophic flagellates. In *The Flagellates - Unity, Diversity and Evolution*; Taylor & Francis Ltd: London, UK, 2000; pp. 269–287.
54. Butler, H.; Rogerson, A. Temporal and spatial abundance of naked amoebae (Gymnamoebae) in marine benthic sediments of the Clyde Sea Area, Scotland. *J. Eukaryot. Microbiol.* **1995**, *42*, 724–730. [CrossRef]
55. Weber, A.A.-T.; Pawlowski, J. Can abundance of protists be inferred from sequence data: A case study of foraminifera. *PLoS ONE* **2013**, *8*, e56739. [CrossRef] [PubMed]
56. Prosdocimi, E.M.; Novati, S.; Bruno, R.; Bandi, C.; Mulatto, P.; Giannico, R.; Casiraghi, M.; Ferri, E. Errors in ribosomal sequence datasets generated using PCR-coupled "panbacterial" pyrosequencing, and the establishment of an improved approach. *Mol. Cell. Probes* **2013**, *27*, 65–67. [CrossRef] [PubMed]
57. Zhu, F.; Massana, R.; Not, F.; Marie, D.; Vaulot, D. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol.* **2005**, *52*, 79–92. [CrossRef] [PubMed]
58. Potvin, M.; Lovejoy, C. PCR-based diversity estimates of artificial and environmental 18S RNA gene libraries. *J. Eukaryot. Microbiol.* **2009**, *56*, 174–181. [CrossRef] [PubMed]
59. Wolf, C.; Kilias, E.S.; Metfies, K. Evaluating the potential of 18S rDNA clone libraries to complement pyrosequencing data of marine protists with near full-length sequence information. *Mar. Biol. Res.* **2014**, *10*, 771–780. [CrossRef]
60. Dell'Anno, A. Extracellular DNA plays a key role in deep-sea ecosystem functioning. *Science* **2005**, *309*, 2179–2179. [CrossRef] [PubMed]
61. Torti, A.; Lever, M.A.; Jørgensen, B.B. Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Mar. Genom.* **2015**, *24*, 185–196. [CrossRef] [PubMed]
62. Stoeck, T.; Zuendorf, A.; Breiner, H.-W.; Behnke, A. A molecular approach to identify active microbes in environmental eukaryote clone libraries. *Microb. Ecol.* **2007**, *53*, 328–339. [CrossRef] [PubMed]
63. Not, F.; Gausling, R.; Azam, F.; Heidelberg, J.F.; Worden, A.Z. Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ. Microbiol.* **2007**, *9*, 1233–1252. [CrossRef] [PubMed]
64. Stecher, A.; Neuhaus, S.; Lange, B.; Frickenhaus, S.; Beszteri, B.; Kroth, P.G.; Valentin, K. rRNA and rDNA based assessment of sea ice protist biodiversity from the central Arctic Ocean. *Eur. J. Phycol.* **2016**, *51*. [CrossRef]
65. Logares, R.; Audic, S.; Bass, D.; Bittner, L.; Boutte, C.; Christen, R.; Claverie, J.-M.; Decelle, J.; Dolan, J.R.; Dunthorn, M.; *et al*. Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* **2014**, *24*, 813–821. [CrossRef] [PubMed]
66. Schoenle, A.; Werner, J.; Nitsche, F.; Arndt, H. Ciliates in the abyss: Occurrence and survival at high hydrostatic pressures. University of Cologne, Cologne, Germany, Unpublished work, 2015.

Chapter 6 – Discrepancies between molecular and morphological databases of soil ciliates studied for temperate grasslands of central Europe

Discrepancies between molecular and morphological databases of soil ciliates studied for temperate grasslands of central Europe

Paul Christiaan Venter, Frank Nitsche, Anja Scherwass, Hartmut Arndt*.

University of Cologne, Biocenter Cologne, Institute of Zoology, Department of General Ecology, Zuelpicher Str. 47b, D-50674 Koeln (Cologne), Germany

Running title: Molecular techniques reveal the hidden diversity of soil ciliates

Abstract

By measuring the change in soil protist communities, the effect of human land use on grasslands can be monitored to promote sustainable ecosystem functioning. Protists form the active link in the rhizosphere between the plant roots and higher trophic organisms; however, only few morphological species and their ecological values have yet been described in this context. To investigate the communicability between morphological and molecular databases used in the molecular barcoding of protists and in the biomonitoring of grassland soil, the present high-throughput sequencing (HTS) study (N=150) covered the area of central Europe (mesoscale) known to be well studied for ciliated protists. The 2,404 unique individual HTS reads identified taxa in all major ciliophoran classes but exact reference matches were few. The study identified clear discrepancies between databases for well-studied taxa, where molecular databases contained multiple gene variants for single morphospecies of dominant taxa. Gene variants presented own biogeography – the eukaryotic microdiversity along gradients (e.g., land-use intensity, soil water). It is possible that many of the so called novel phylogenetic lineages and hidden diversity pointed out in environmental surveys could be evidence for the severe lack of molecular data for already known and morphologically described species, present in morphological databases.

Keywords: molecular ecology; soil; diversity; land-use intensity; spatial distribution; Ciliate

Introduction

Land use is a major recognized driver of community structure in terrestrial habitats like grasslands (Blüthgen et al. 2012; Soliveres et al. 2016), where community stability is very important for sustainable ecosystem functioning (Gossner et al. 2016). The communities of agricultural meadows and pastures are governed by land use types (fertilization, mowing and grazing by livestock) and intensity, which impacts negatively on the belowground diversity of organisms like unicellular eukaryots which build an important part of the soil microbial food web being necessary for optimal plant growth (Bonkowski 2004; Ekelund et al. 2002; Gossner et al. 2016). Protists, forming the active link between the rhizosphere and higher trophic organisms, are usually studied and taxonomically classified by their morphology, using microscopic and culture based techniques (Esteban et al. 2006). Morphological techniques are used for protist identification however, their description is time consuming and requires a thorough knowledge of taxonomy. Due to this, the taxa found can hide cryptic species and, moreover, may exclude non-cultivable organisms (Foissner et al. 2008). Additionally, it is believed that there are also many undescribed species which are yet undiscovered and their taxonomical description constitute a major part of all protists (Pawlowski et al. 2014; Pawlowski et al. 2016).

Ciliophora, especially in soil, is the best described group of all protists based on morphological techniques (Chao et al. 2006; Foissner 2006; Lara and Acosta-Mercado 2012). For several reasons, ciliates have been suggested as ideal bioindicators of land-use change and environmental stress in soil (Foissner 1999; Lara and Acosta-Mercado 2012), because they are so well investigated in central European soils (Chao et al. 2006; Mahé et al. 2017). Ciliates have high turnover rates and biomass in soil and are very sensitive to environmental conditions (Foissner 1999, 2016). A lot of ecological information is connected to the individual morphospecies (Esteban et al. 2006; Foissner et al. 2011; Foissner et al. 2014; Foissner 2016; Grattepanche et al. 2016). E.g. Colpodea with its high growth rates and

tolerance to unfavorable conditions are expected to occur in disturbed soils, whereas the important predatory groups of haptorians and suctorians are expected to occur in stable soils in significant amounts/biomass (Lara and Acosta-Mercado 2012). Altogether about 8,000 ciliates morphospecies (of which 4,500 are free-living) were described until 2008 (Foissner 2008). Even though ciliates are very important components in soil food webs, little information exist on the impact that agricultural practices has on community structures (Gossner et al. 2016; Soliveres et al. 2016) for ciliates in particular (Díaz et al. 2006).

Agricultural practice covers large geographical areas and terrestrial soils are, like the benthos of the deep sea, important protist reservoirs concentrating up to 15,000 individuals per gram of dry soil weight (Massana et al. 2015; Rogerson and Detwiler 1999). Even though morphological identification still is the gold standard for describing new species and their associated habitats (Chao et al. 2006; Foissner et al. 2014), new genetic barcoding methods employing next-generation sequencing (NGS) technology are culture independent and high throughput sample processing allows studies of large geographies (Grossmann et al. 2016; Pawlowski et al. 2012). Some researchers argue that taxonomic identification is not even necessary to draw conclusions from operational taxonomic units (OTUs) as markers for protists (Pawlowski et al. 2016). For some morphological species, genetic strain differences detected using NGS can be responsible for very different ecological properties within the same taxon (Díaz et al. 2006). Therefore the already existing discrepancies between morphologically collected data and the genetic sequence data in public databases like GenBank may be exacerbated (Chao et al. 2006; Forster et al. 2016). A good communication of morphological work and relatable molecular possibilities (e.g. metabarcoding studies) is needed, because studies using both methods correct for the biases of the other (Schoenle et al. 2016). Reference sequences for protists are collected in databases and is well curated up to 8 taxonomic levels (Guillou et al. 2013), setting the stage for combined phylogenetic and taxonomic classification (Forster et al. 2012; Forster et al. 2016; Pawlowski et al. 2012). A

discrepancy between the molecular data obtained for unknown environmental samples and that available in molecular databases as well as the morphologically described species exists (Chao et al. 2006), but the question remains, how great the differences for known soil taxa are.

We set out to sample 150 soil samples, all from the grassland biome, covering three sampling sites across Germany; which may serve as a reasonable representation of central European grassland soils. We hypothesize that using high-throughput sequencing (HTS), many sequences with high pairwise identities to already known and described species in the database should be found for central European soils, e.g. especially flagship species. Any deviation from this statement may be due to the discrepancy between molecular and morphological databases. We aimed to provide an overview of the morphological to molecular discrepancy that exists for ciliate taxa in databases.

Results

NGS data and taxa coverage

To study the group-specific richness of ciliates in soil by sequence barcoding, we used a molecular dataset obtained by using universal eukaryotic primers. A modified pipeline was applied to evaluate next-generation sequencing (NGS) results for ciliate taxa at unique individual reads (UIRs) level and Blast based operational taxonomic units (bbOTUs) level (UIRs grouped to the same accession number; see Methods). These two microbial taxonomic unit definitions ultimately made an upper and lower species richness evaluation possible. Based on this definition and filtering the data to evaluate ciliate taxa only, between 70 (bbOTUs - minimum number of taxa) at the most conservative (≥99.7%) pairwise identity level and 8407 (UIRs - maximum number of taxa) taxa at the all-inclusive ≥80% pairwise identity level were discovered in 150 grassland sampling sites from an initial ~1.2 million raw reads. Discovered bbOTUs fell within all seven known ciliate classes (Adl et al. 2012). Taxonomic units are dependent on the level of pairwise identity resolution (≥99.7%-, ≥97%- and ≥80% pairwise identity; Table S1) and, when converted to presence-absence matrixes per site, should present a statistically sound dataset for analyses.

Taxonomic coverage evaluation. An *in silico* PCR test of the universal primers against the reference sequences in the Protist Ribosomal Reference (PR^2) database (GenBank, version 203, downloaded June 2016) and SILVA (SSU r117) databases (www.arb-silva.de; accessed June 2016) was performed to estimate their coverage of the major taxonomic groups. A primer Blast indicated that more than 90% of the 135,216 PR^2 database reference sequences for all taxonomic groups were covered by the primer set. Compared to the SILVA database, our eukaryotic primer pairs returned only eukaryotic supergroups, covering 74.5% of the alveolate and 66.4% of all ciliophoran reference sequences, predominantly for the class Spirotrichea (88.8%).

Species richness. Twenty-four percent of the universal primer detected protist assigned sequences could be identified within one of the ciliate class taxa (Table S1). Supplementing the primer test, no single sequence pairwise identity hit for a UIR to the PR^2 database fell below the 80% cut-off for inclusion into the analyses. Despite the high similarity between query sequences and database reference sequences, at ≥97% pairwise identity, no single taxon occurred across all 150 grassland sites. The community mainly consisted of rare taxa with limited distribution, where >42% of the identified taxa occurred at less than 10 of all 150 sites (Fig. 1A). Even the most abundant ciliate taxon (an unknown *Stichotrichia* sp.) occurred in no more than 139 of the 150 grassland plots. Species discovery (sampling intensity) was highly dependent on the level of taxonomic resolution (≥99.7%, ≥97% and ≥80% pairwise identity; Fig. 1B) and how one defines taxonomic units (bbOTUs vs. UIRs; Fig. 1C). At the lower pairwise identity cut-offs (<97% pairwise identity) a number of additional UIRs with unknown lineages were added to the already numerous number of unresolved taxa (>47%) at ≥97% pairwise identity (Table S1).

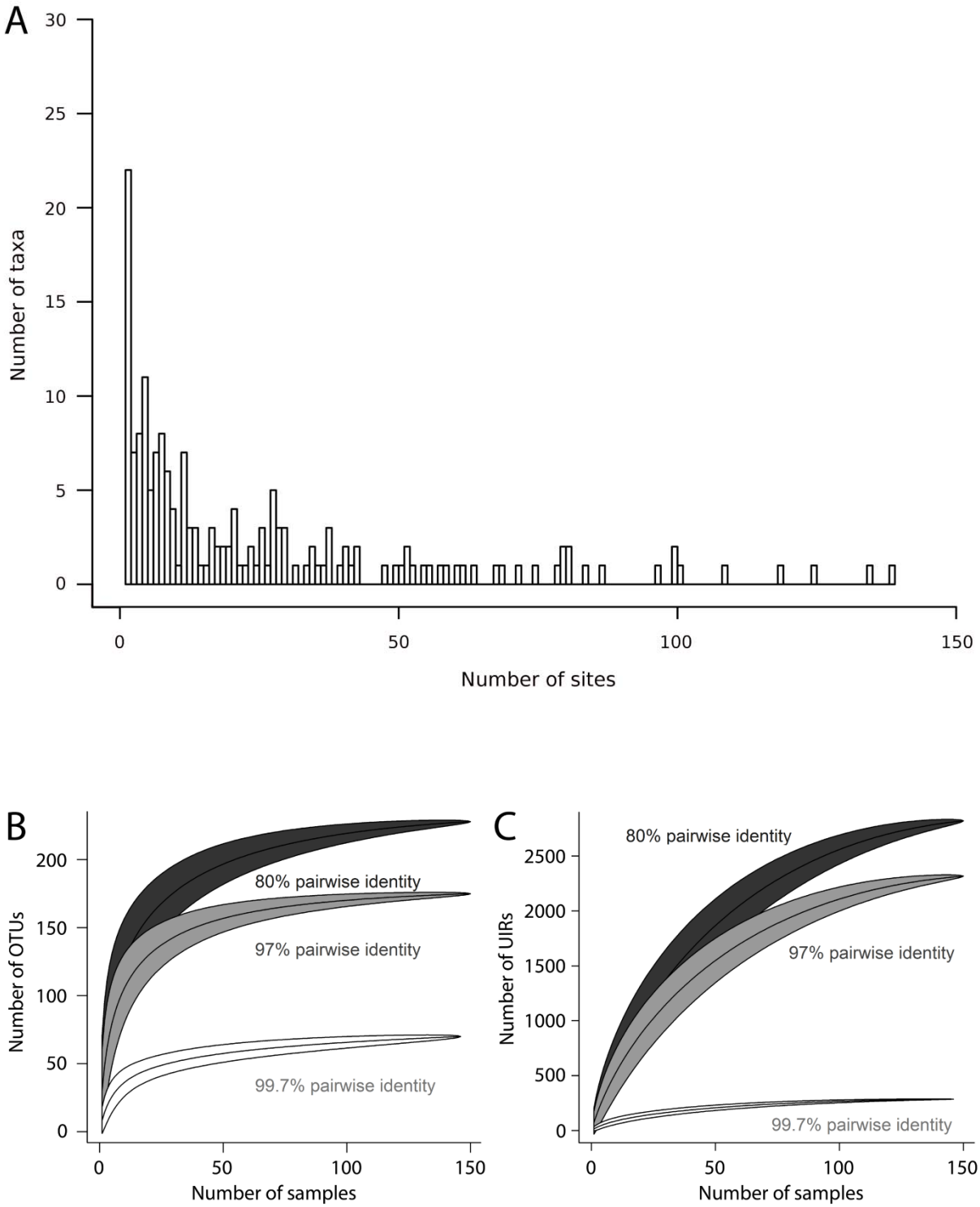
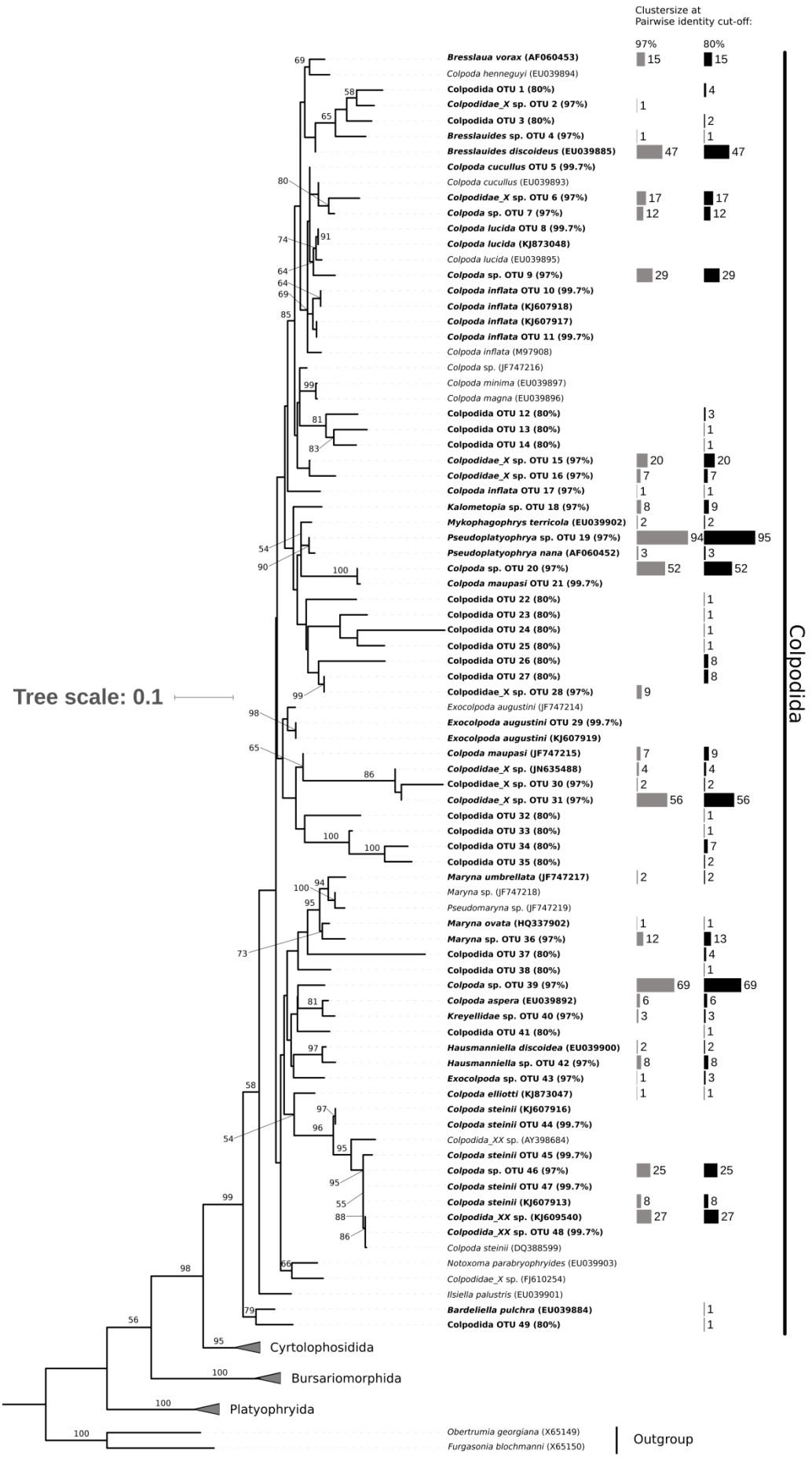


Figure 1. Frequency of taxa distribution and species accumulation curve for all 150 grassland sample sites. The frequency of Blast based operational taxonomic units (bbOTUs) occurring per the number of sampling sites (**A**). The number of discovered OTUs added per additional sampling site (**B**), compared to the number of discovered unique individual reads (UIRs) added per additional sampling site (**C**), at three sequence similarity cut-offs.

Species accumulation curves based on ciliate-specific unique sequences (UIRs) derived from all 150 soil samples did not approach an asymptote, which indicated that the full extent of the ciliate diversity associated with the mesoscale was not reflected and hence not available for analysis (Fig. 1B). Decreasing rather than increasing the reference sequence similarity cut-off to ≥80%, not only resolved putative class taxonomic level bbOTUs, but also added 20% more sequences and three ciliate classes (Armophorea, Cariacotrichea, Phyllopharyngea) for analysis not observed at the ≥97% pairwise identity resolution. bbOTUs were most frequently matched to reference sequences for the aerobic ciliate groups Colpodea, Litostomatea, Spirotrichea, Oligohymenophorea and to a lesser extent Heterotrichea, Nassophorea, and Prostomatea.

Figure 2 (Following page). Phylogenetic placement of unique individual read (UIR) clusters within the known diversity for the colpodean class of Colpodida. UIRs identified at ≥99.7% -, ≥97%-, and ≥80% sequence pairwise identity within the class group was inferred by the V4 SSU gene alignment to protist ribosomal reference (PR^2) database sequences and then clustered together with the returned reference sequences at ≥97% global similarity. Known and unknown diversity was included into the basic maximum likelihood (ML) tree structure adopted from Foissner (2011) and Foissner et al. (2014) originally containing the 51 reference taxa. Barcharts at the bold fonted leaves indicate the number of UIRs associated to the respective cluster at ≥97% - and ≥80% sequence pairwise identity level resolution to reference sequences.


Clustersize at
Pairwise identity cut-off:
97%
80%
Bresslaua vorax (AF060453) 15 15
Colpoda henneguyi (EU039894)
Colpodida OTU 1 (80%) 4
Colpodidae_X sp. OTU 2 (97%) 1
Colpodida OTU 3 (80%) 2
Bresslauides sp. OTU 4 (97%) 1 1
Bresslauides discoideus (EU039885) 47 47
Colpoda cucullus OTU 5 (99.7%)
Colpoda cucullus (EU039893)
Colpodidae_X sp. OTU 6 (97%) 17 17
Colpoda sp. OTU 7 (97%) 12 12
Colpoda lucida OTU 8 (99.7%)
Colpoda lucida (KJ873048)
Colpoda lucida (EU039895)
Colpoda sp. OTU 9 (97%) 29 29
Colpoda inflata OTU 10 (99.7%)
Colpoda inflata (KJ607918)
Colpoda inflata (KJ607917)
Colpoda inflata OTU 11 (99.7%)
Colpoda inflata (M97908)
Colpoda sp. (JF747216)
Colpoda minima (EU039897)
Colpoda magna (EU039896)
Colpodida OTU 12 (80%) 3
Colpodida OTU 13 (80%) 1
Colpodida OTU 14 (80%) 1
Colpodidae_X sp. OTU 15 (97%) 20 20
Colpodidae_X sp. OTU 16 (97%) 7 7
Colpoda inflata OTU 17 (97%) 1 1
Kalometopia sp. OTU 18 (97%) 8 9
Mykophagophrys terricola (EU039902) 2 2
Pseudoplatyophrya sp. OTU 19 (97%) 94 95
Pseudoplatyophrya nana (AF060452) 3 3
Colpoda sp. OTU 20 (97%) 52 52
Colpoda maupasi OTU 21 (99.7%)
Colpodida OTU 22 (80%) 1
Colpodida OTU 23 (80%) 1
Colpodida OTU 24 (80%) 1
Colpodida OTU 25 (80%) 1
Colpodida OTU 26 (80%) 8
Colpodida OTU 27 (80%) 8
Colpodidae_X sp. OTU 28 (97%) 9
Exocolpoda augustini (JF747214)
Exocolpoda augustini OTU 29 (99.7%)
Exocolpoda augustini (KJ607919)
Colpoda maupasi (JF747215) 7 9
Colpodidae_X sp. (JN635488) 4 4
Colpodidae_X sp. OTU 30 (97%) 2 2
Colpodidae_X sp. OTU 31 (97%) 56 56
Colpodida OTU 32 (80%) 1
Colpodida OTU 33 (80%) 1
Colpodida OTU 34 (80%) 7
Colpodida OTU 35 (80%) 2
Maryna umbrellata (JF747217) 2 2
Maryna sp. (JF747218)
Pseudomaryna sp. (JF747219)
Maryna ovata (HQ337902) 1 1
Maryna sp. OTU 36 (97%) 12 13
Colpodida OTU 37 (80%) 4
Colpodida OTU 38 (80%) 1
Colpoda sp. OTU 39 (97%) 69 69
Colpoda aspera (EU039892) 6 6
Kreyellidae sp. OTU 40 (97%) 3 3
Colpodida OTU 41 (80%) 1
Hausmanniella discoidea (EU039900) 2 2
Hausmanniella sp. OTU 42 (97%) 8 8
Exocolpoda sp. OTU 43 (97%) 1 3
Colpoda elliotti (KJ873047) 1 1
Colpoda steinii (KJ607916)
Colpoda steinii OTU 44 (99.7%)
Colpodida_XX sp. (AY398684)
Colpoda steinii OTU 45 (99.7%)
Colpoda sp. OTU 46 (97%) 25 25
Colpoda steinii OTU 47 (99.7%)
Colpoda steinii (KJ607913) 8 8
Colpodida_XX sp. (KJ609540) 27 27
Colpodida_XX sp. OTU 48 (99.7%)
Colpoda steinii (DQ388599)
Notoxoma parabryophryides (EU039903)
Colpodidae_X sp. (FJ610254)
Ilsiella palustris (EU039901)
Bardeliella pulchra (EU039884) 1
Colpodida OTU 49 (80%) 1
Colpodida
Tree scale: 0.1
Cyrtolophosidida
Bursariomorphida
Platyophryida
Obertrumia georgiana (X65149)
Furgasonia blochmanni (X65150)
Outgroup

The presence of novel diversity. To further investigate the extent of novel diversity among UIRs, we clustered environmental UIRs to reference sequences in the database and evaluated this within a phylogenetic framework of known taxa. To make sequence clustered operational taxonomic unit (scOTUs) results comparable with other studies, we used the ≥97% pairwise identity level cut-off in a sequence clustering analyses step and present data for one well described taxon – the Colpodea (Figs. 2, S1), but also for two more taxa discovered in soil (Figs. S2-S3). Multiple gene variants were present in the database for some morphospecies taxa, especially the dominant taxa in soil (e.g. Colpodida; Fig. 2 and Table S2), e.g. the gene strain sequence variants for the dominant taxa *Colpoda inflata* (KJ607917 and KJ607918) and *Colpoda steinii* (KJ607913, KJ607916 and DQ388599) are well visible within the phylogenetic context, where pairwise alignments between these sequences deliver <97% sequence similarity. Several*Colpoda* strains indicated different incidence patterns across sites and associated land-use regimes (Fig. S4).

Some branches for other taxa were grossly underrepresented in databases. In the Colpodea tree (Figs. 2, S1), possible gene variants for *Bresslauides* sp. (scOTUs 1-3) and two already unresolved Colpodidae_X sp. (scOTUs 22-27; scOTUs 30-31) could be detected. Also present were novel orphan branches for which no reference sequence could be identified by global alignment with other colpodid reference taxa (scOTUs 12-14; scOTUs 32-35). Good resolution was possible for the other colpodean orders (Fig. S1) with similar UIR alignment tendencies. However, for hypotrich UIRs (Fig.S2), phylogenetic evolutionary rates were low and clusters were unstable within the large tree containing many gene variants for unresolved taxa. This hypotrich tree topology (Fig.S2) stood in sharp contrast to that for Euplotia (Fig. S3), where the few UIR-only clusters with low pairwise similarity (except scOTU 1) associated near known reference sequences in the database. At least three UIR clusters (scOTUs1-3) associated with the genus *Aspidisca* with high bootstrap values indicating the lack of molecular data for morphologically described species (Table S4). *Aspidisca* sequences

were Blast aligned to environmental sequences in the NCBI database [accessed: 30.06.2017] and no matching uncultured environmental sequence from other studies could be found with >94% sequence similarity.

Morphological vs. molecular data. To analyse possible discrepancies between the molecular data and morphologically available data for the discovered taxa (UIRs and bbOTUs) a tabulated inventory for one of the most abundant and well studied ciliate groups in grassland soils, the Colpodea (Table S2), and two related ciliate groups (Hypotrichia and Euplotia; Table S3-S4) were made. Data (UIRs and bbOTUs) for the above mentioned taxa were compared (Fig. 3A-F) to available molecular (PR^2 database) and morphological database data available online (EOL; www.eol.org). Additionally, the comprehensive monographs (Berger (2001) for Hypotrichia and Euplotia; Foissner (1993) for Colpodea) were considered to compare OTUs with the number of morphospecies at a certain time point. UIRs matching PR^2 database accession numbers as well as the related bbOTUs were grouped at three levels of pairwise identity (≥80%, ≥97% and ≥99.7%). When considering UIRs at each pairwise identity cut-off for colpodean orders, the detected unknown diversity among colpodid ciliates using NGS laid well in excess of all species mentioned in databases (Fig. 3A-D). The number of taxa we recorded based on the grouping of Blast results for UIRs into bbOTUs was about two thirds of the number of PR^2 database entries as well as morphologically described colpodean species. This ratio was observed across all orders within the class Colpodea (Fig. 3A-D), but changed for groups overrepresented (Hypotrichia) and underrepresented (Euplotia) by HTS in soil samples (Fig. 3E-F). Of all colpodean orders, the most dominant orders (e.g. Colpodida) also contained the most database entries. Apart from this, a detailed analysis of the PR^2 database (used to identify the UIRs) binomial annotations indicated that only about one third of the reference sequences in the PR^2 database contained a unique genus with species annotation.

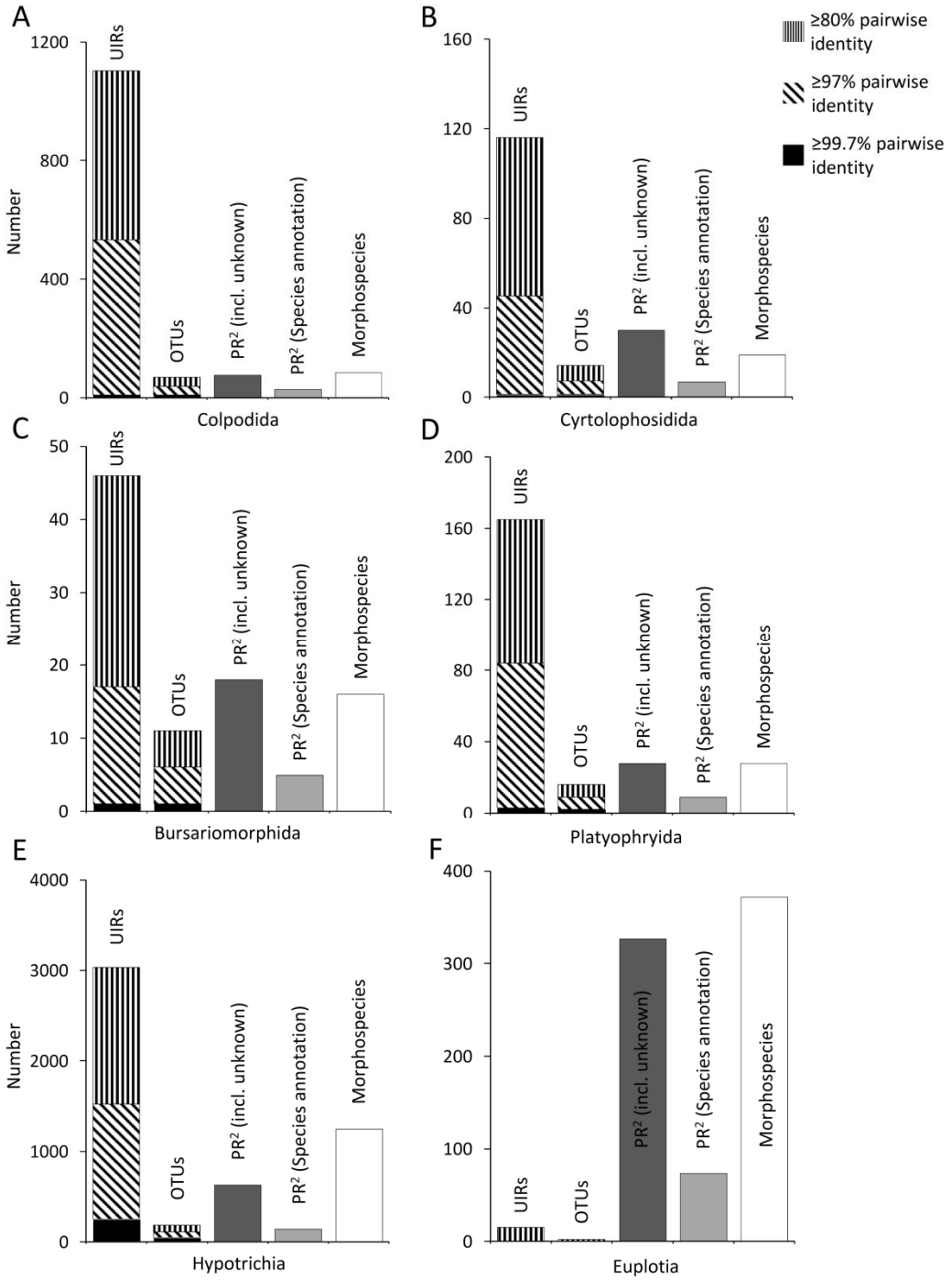


Figure 3. An inventory of the data available for ciliate colpodean UIRs and Blast based OTUs identified by means of molecular methods using a sequence database compared to morphological described data available. A barchart illustrating the relevant sequencing data available in the protist ribosomal reference database (PR^2) as measured up to species specific data (sequences containing a lineage with genus and species) and compared to morphological data collected (see Suppl.) for each of the Colpodea classes Colpodida (**A**), Cyrtolophosidida (**B**), Bursariomorphida (**C**) and Plathyophryida (**D**). The number of unique individual reads (UIRs) and operational taxonomic units (OTUs) returned after a next-generation sequencing (NGS) effort of 150 grassland soil samples are given by the last two bars.Further resolving this information at three cut-offs levels for UIRs and OTUs identification (at ≥99.7 -, ≥97% - and ≥80% sequence pairwise identity) to reference sequences in the PR^2 database are given be inter-bar chart resolution.

The same morphological and genetic inventory research was repeated for two more groups, the Hypotrichia and Euplotia (Fig. 3E, F, Table S3, S4). Both taxonomic groups belong to the ciliate class Spirotrichea, the most dominant taxonomic group across all grassland sites in the present study (Table S1) containing many taxa (Table S5). For Hypotrichia, less than half as many reference sequence data could be found in the PR^2 database when compared to the number of morphological described hypotrich species available in the monograph (Fig. 3E, Table S3). On the other hand, more hypotrich UIRs than described species were discovered (1300 UIRs), but clustered to only a tenth (63 bbOTUs) of the number of morphologically described hypotrich species. For Euplotia, the difference between genetic and morphological data was not as drastic as for the hypotrichs, although we retrieved only few taxa (3 bbOTUs and 17 UIRs) from soil data (Fig. 3F, Table S4) with low pairwise identity distances to known taxa in the PR^2 database (Table S1). On the other hand, we observed stable phylogenetic positions to reference sequences (UIRs distances <97% sequence similarity to known references, Fig. S3). When comparing individual genera within the Euplotia, the gaps between morphological and genetic database data became more pronounced. This was especially the case for the genus *Aspidisca* (Table S4, Fig. S3). Altogether 78 morphologically described *Aspidisca* species are in a clear contrast to only seven species contained in the PR^2 database (Table S4). A similar situation occurred for the genera *Euplotes* (Order Euplotia) and *Diophrys* (Order Hypotrichia) (Table S4, S5).

Dominant taxa and the factors influencing species richness. Of the 175 ciliate bbOTUs identified at ≥97% pairwise identity (Table S5), about one third were redundancies or most likely genevariants for the same species (Fig. 3A-F). This was not only the case for sequence entries with uncertain lineages at the ≥97% and ≥99.7% sequence similarity level in the PR^2 database, but especially for well-known taxa like *Colpoda* (*C. aspera, C. inflata, C. steinii*), *Cyrtolophosis mucicola*, *Exocolpoda augustini*, *Gonostomum strenuum*, *Halteria grandinella* (3 bbOTUs), *Urostyla grandis*, and *Vorticella microstoma*. Furthermore, these

species were also among the most dominant in our soil samples (e.g. *Colpoda steinii* occurring at 81 of the 150 sampled plots). Summating the presence-absence totals for all bbOTUs within each class group, species for the class Spirotrichea collectively indicated the highest relative abundance across all sites (with high presence-absence bbOTU entries: N=2185; 48.0%) followed by Colpodea (N=1580; 34.7%). Other class groups listed in decreasing order of (collective) relative abundances across sites included Litostomatea (442; 9.7%), Oligohymenophorea (243; 5.3%), Nassophorea (90; 2.0%), Heterotrichea (9; <0.2%) and Prostomatea (2; <0.05%). The number of UIRs and bbOTUs was quantified for the eleven ciliate classes (TableS1).

The most dominant ciliate taxa discovered in our dataset were Blast results for UIRs associated with unknown lineages (Table S2-S4). Grouping for database lineages, the ten most widely distributed bbOTUs in order of increasing distribution dominance were: eight ciliate taxa with uncertain lineages ([EF024684] Oxytrichidae_X sp., [JN635488] Colpodidae_X sp., [EF100341] *Platyophrya* sp., [EF024585] Orchitophryidae_X sp., [EF024004] *Colpoda* sp., [EF024903] Oxytrichidae_X sp., [KJ509196] *Bistichella* FG-2014, [AB449362] *Stichotrichia* sp.) and two with known lineage ([AF060452] *Pseudoplatyophrya nana* and [KJ873047] *Colpoda elliotti*). Two gene variants (two accession numbers for the same taxon) of the Oxytrichidae_X sp. were present; each with an own incidence pattern during our investigation (Fig. S4).

Analysis of spatial scaling for contiguous grassland habitats were performed to identify underlying mechanisms that influence ecological community structure in soil and how results may be database dependent (OTUs vs. UIRs, Fig. S5A-C). The community structure of the morphologically larger ciliates (generally 20-300µm length) was less influenced by geographic separation when comparing UIR data of non-ciliophoran protist assemblage to the much smaller flagellate species (generally 2-20 µm,Venter et al. in press) (Fig. S6). In all cases for ciliate data, taxa-area curves indicated increased *z*-values ($z > 0.2$) when based on

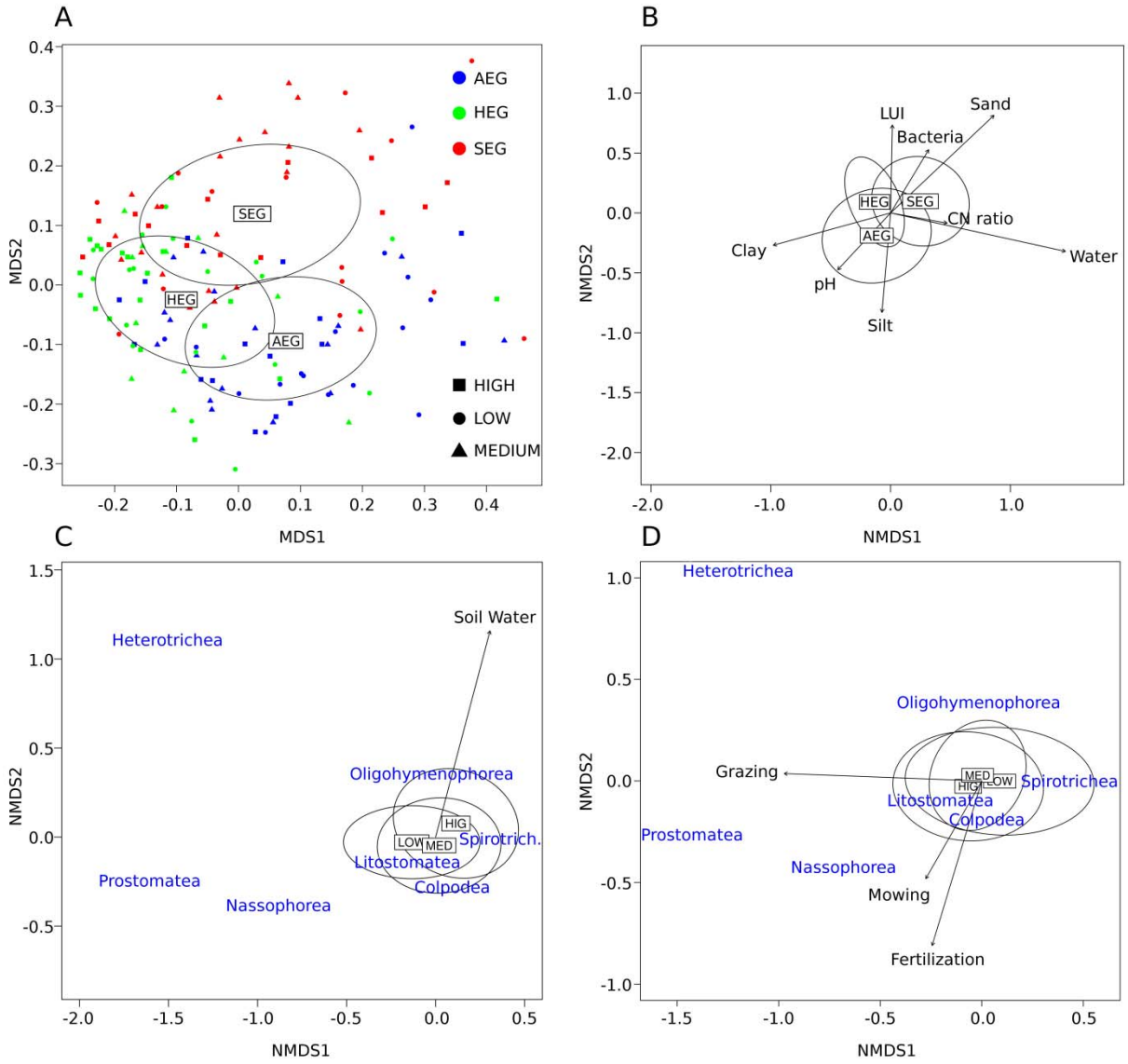


Figure 4. The relationship between soil ciliate communities and soil variables. A multidimentional scaling (MDS) plot was created resolving the community dissimilarity for the three Biodiversity Exploratories (ALB – Schwäbische Alb, HAI – Hainich-Dün and SCH – Schorfheide-Chorin) and land-use intensities (LUI – HIGH, MEDIUM and LOW) as separated along the first two eigenvector axes (**A**). A variables factor map indicated the ordination of ciliate data to soil sample properties by means of a visual non-metric multidimentional scaling (NMDS) plot at each of the Biodiversity Exploratories (ellipse: ALB, SCH and HAI) sites (soil data were obtained from Kaiser et al. 2016) (**B**). This analysis was repeated for the percentage soil water at three levels (ellipse: high, medium and low) of water content (**C**) and LUI at three levels (ellipse: high, medium and low) of land-use (**D**). We illustrated constrained community ordination and random starting configurations using Bray-Curtis dissimilarity matrixes for all figures. Species scores were added as weighted averages. Ordinations for vectors based on continuous variables and centroids were fitted after applying 999 permutations. Gradients are illustrated as arrows, where the length of the arrow is proportional to the strength of correlations.

UIRs compared to z-values based on UIRs grouped for bbOTUs, (z< 0.1, Fig. S2A-C) which is generally used as one indicator of beta diversity. The level of taxonomic identification influenced the overlap of species similar between sites in the mesoscale (Fig. S5C), where species identified at ≥99.7% pairwise identity to reference sequences showed the highest z-value (at a lower species richness).

Multidimensional scaling at ≥97% pairwise identity indicated that geographic distance rather than land use separated communities (Fig. 4A). Water (*rho* = -0.5; p< 0.01; Figs. 4B, C) and very fine sand (clay being an independent variable), especially silt (*rho* = -0.3; p< 0.01), among other edaphic variables (Fig. 4B), was responsible for significant gradients influencing the beta-diversity. Increased land use (*rho* = 0.24; p< 0.01), especially grazing intensity (*rho* = 0.3; p< 0.01), was associated with changes in ciliate species richness, where the vector was parallel with the first NMDS axis (Fig. 4D). Environmental factors associated mostly with the community structure of dominant taxa, where e.g. Spirotrichea (with the highest relative abundance) largely associated with low land use and sampling plots containing high soil moisture (Fig. 4C).

Discussion

Ciliates include many dominant "flagship" species with distinct morphology (Foissner 2006), making them easy to find in a territory like Europe where they have been well studied for a long time (Foissner 2006; Chao et al. 2006). Thus, using the sensitivity of High-Throughput Sequencing (HTS), at least the well-studied ciliate flagship taxa present in the German mesoscaleshould easily be recoverable. We hypothesized that any deviation from this statement may be due to the discrepancy between molecular and morphological databases.

Flagship species recovered. Comparing our molecular data set for central European soils to another European study (Scotland, UK) using morphological identification methods (Esteban et al. 2006) and the same number of soil samples, revealed the interesting "rediscovery of thirds" (Foissner et al. 2008). Only a third of the discovered UK-grassland taxa could be rediscovered in our middle European grassland samples (Table S5), with similar dominance patterns for well described species like *Colpoda steinii*, *Colpoda inflata* and *Platyophrya vorax* (Fig. S4). This fact could be used to motivate the ubiquitousness hypothesis (Esteban et al. 2006) or strengthen the semi-ubiquitous distribution idea of ciliates in soil (Foissner et al. 2008). But, the absence of UIRs cannot be used as proof of absence of the organism in the mesoscale, because undersampling for mostly unknown taxa (UIRs) cannot be excluded (Fig. 1C). If molecular results closely resembles typical morphological rediscovery of a third of the taxa: how biased is HTS sample identification because of the discrepancy between molecular and morphological known diversity?

In a quest for answers, we compared our molecular survey data with a list of flagship morphospecies data hypothesized by Foissner (2008) to have restricted geographic distribution and recovered the name of only one of the 52 ciliate species. This was *Bresslauides discoideus*, which is known to occur in soil and moss of Europe, Japan and Central America. This was unspectacular; however, we also found that reference sequences for only 9 of the 52 flagship species were present in the PR^2 database. E.g. one of these

flagship species (*Neokeronopsis spectabilis*) mentioned by Foissner et al. (2008) is big (300-400 μm) and already known since the 1930s to occur in central Europe, but no reference sequence for it occurred in GenBank or the dedicated PR^2 database [Accessed: June 2017]. This implied that the molecular to morphological species data gap may be so big, that sequences for most morphologically described flagship ciliate species with restricted geographic distribution are simply not available in reference databases.

Ciliate novelty and richness in soil. All discovered unique individual reads (UIRs) in our study had a more than 80% sequence similarity to reference sequences in the Protist Ribosomal Reference (PR^2) database (Fig. 1B). This stands in sharp contrast to neotropical soil samples where more than two thirds of the query sequences were less than 80% similar to reference sequences in the same database (Mahè et al. 2017). This may confirm that ciliates are more well studied in the temperate terrestrial systems of central Europe (Chao et al. 2006), even though exact reference matches were few.Most UIRs related to reference sequences up to the 20% sequence dissimilarity cut-off (Table S1). We, however, found that this was not the case in all taxonomic groups. Well studied ciliate taxa included the Colpodea and Hypotrichea (Dunthorn et al. 2014; Foissner et al. 2014; Gao et al. 2016), for which many reference sequences existed in the PR^2 database. This might explain why these taxa were predominantly recovered from soil samples (Table S1). High-throughput sequencing (HTS) UIRs revealed the presence of well known taxa for both free-living (most of these protists are known to be capable of cyst formation) and true parasites of humans (e.g. a *Balantidium* sp.), fish (*Ichtyophthirius* sp.) and garden slugs (*Tetrahymena rostrata*), indicating the broad range of ciliates covered (estimated 8,000 morphospecies) using our universal primers (Table S5).

Exact sequence matches of UIRs were sparse (~1%), therefore taxonomic binning was applied (Tables S1, S2-S4). Binning similar UIR Blast hits within pairwise similarity thresholds into Blast based operational taxonomic units (bbOTUs) allowed for related higher taxonomical and functional assignment (Urich et al. 2008). The bbOTUs grouped at ≥97%

pairwise identity cut-off (~82% of UIRs) allowed for the comparison of results with other studies of microbes. UIRs or grouped bbOTUs with ≥99.7% reference sequence identity probably indicated the same species based on potential PCR and pyrosequencing error rates (Venter et al. 2017). Based on the above mentioned UIR and bbOTU definitions, presence-absence transformed data (at ≥97% pairwise similarity) indicated a mean ciliate bbOTUs richness of 30.3 g^{-1}dry weight soil (SD: 15.5, CI: 0.85, min. 1, max. 73 bbOTU) in the 150 agricultural grassland sites. The total richness (175 bbOTUs) and mean value was within the same range as previously obtained from morphological studies using 150 soil samples at one hectare European grassland site (91 ciliates), worldwide (95 ciliates, Esteban et al. 2006) and in soils from Puerto Rico (77 ciliates from 36 sampling sites, 5-67 ciliates per g dry weight of soil, Acosta-Mercado & Lynn 2004). The total ciliate richness, we obtained with this method, is also in the same range as a molecular study using ciliate group specific primers sampling four river systems (176 clones in 50 µl size samples covering 130 cm^2 of biofilm, Dopheide et al. 2008). At greater dissimilarity from reference sequences (<97% to 80% pairwise similarity), we obtained more unique reads and more taxa per site (Table S1). This additional unknown hidden diversity increased the total richness (228 ciliate bbOTUs per g dry weight of soil), though our estimates were significantly lower than the estimated taxa richness for a worldwide sampling of different soils (1014 OTUs from 40 x 0,25g soil samples, Bates et al. 2013) and marine coastal sediments (5,616 OTUs, Forster et al. 2016), based on HTS molecular methods.

Phylogeny and species variance. One of the best ways to analyse unknown high-throughput sequences and understand database discrepancies is to put query UIRs into a closed-reference phylogenetic context with sequences from already known lineages (Figs. 2, S1-S3). Novel sequences should, by implication, have low sequence similarity to referenced taxa (Dunthorn et al. 2014; Rideout et al. 2014; Forster et al. 2016). Because ciliates are considered a monophyletic group based on synapomorphies and phylogeny (Dopheide et al.

2008; Baroin-Tourancheau et al. 1992), unknown taxa can be resolved by its association within monophyletic descendent clades, like the class Colpodea (Dunthorn et al. 2014). In our study, the order Colpodida still remained monophyletic among the other orders within the taxonomic group of Colpodea (Figs. 2, S1), even when many environmental partial sequences were added to full length sequences (Table S2). UIRs that clustered to known taxa as well as environmental clusters (UIRs clustering only with other UIRs) within each pairwise identity cut-off group, were added to a standard tree for Colpodea by Foissner et al. (2011). Sequence clustering of UIRs (defining scOTUs) using USEARCH (Edgar 2010) within Blast determined pairwise identity cut-off distances (80%, 97% and 99.7%) allowed a phylogenetic global similarity distance association with known lineages (e.g. *Colpodasteinii*and scOTU44-48; Fig. 2). Well studied taxa, for which more than one gene-variant was available in the database, were detected to have different occurance patterns across sites (Fig. S4). While the analysis for *Colpoda steinii* showed no difference in the occurrence of different OTUs, there was a clear geographical separation in the distribution of the two OTUs available for *Colpoda inflata* and *Platyophrya vorax* (Fig. S4). These three examples may indicate possible conspecific (*Colpoda steinii*) and congeneric (*Colpoda inflata*) diversity at ≥99.7% pairwise identity. It is not uncommon for well described, large (~300µm) flagship genera, with high endemic dominance to remain undiscovered (Esteban et al. 2006).

In our study (Fig. 2), the exact number of species was difficult to determine using phylogeny, because it is commonly said that more than one taxon-specific pairwise identity cut-off value may be needed to differentiate individual species (Caron et al. 2009). Some taxa with conserved 18S rRNA gene sequences could be between 99.5 and 99.7% pairwise similar (He and Xu 2011). In such cases other molecular markers, such as cytochrome C oxydase or mitochondrial SSU-rDNA markers (Dunthorn et al. 2014), are required to confirm assignments. In other cases, a 3% dissimilarity, a common OTU clustering threshold, can uncover gene variants for important and well described taxa. This was indeed the case for

Colpoda steinii, *Colpoda inflata* and *Platyophrya vorax* even at ≥99.7% pairwise identity (Fig. S4). The mechanisms for intraspecific evolutionary change may be similar to the long discussed microdiversity acknowledged since decades in prokaryotic identification (Larkin and Martiny 2017).

Molecular vs. morphological data for ciliates in soil. In our survey, most bbOTUs associated with unknown or unresolved family and superorder clades (clades with ending "_X" in Fig. 3A-F and Table S2-S4). The PR^2 database contained 9,360 ciliate sequences: Spirotrichea (3,934), Litostomatea (2,352), Oligohymenophorea (1,404), Colpodea (481) and Phyllopharyngea sequences (235); the typically dominant and described species (Dopheide et al. 2008). Fewer representative sequences were available for Heterotrichea (144), Prostomatea (117), Armophorea (76), Plagiopylea (67), Karyorelictea (65) and Nassophorea (24), which could have limited their identification because fewer gene matches were possible (Dopheide et al. 2008). Directly comparing a list of the UIRs and bbOTUs discovered within each pairwise identity group with information in molecular and morphological databases (Table S2-S4) indicated that three to eight times more UIRs were found than bbOTUs related to lineage morphospecies (Fig. 3A-D). The UIR to bbOTU-ratio is within the range of the number of cryptic morphospecies detected in globally distributed pelagic taxa using molecular data (De Vargas et al. 2015). Comparing our cluster sizes (UIRs) for OTUs withthe number of reference sequences available in the phylogenetic tree (Fig. 2) suggests that most UIRs in fact indicate microdiversityfor known taxa that are not in molecular databases (e.g.*Colpodainflata).*

Of the bbOTUs results, only 51.1% of the Colpodea, Hypotrichia and Euplotia lineages discovered in our study contained binomial nomenclature information in the PR^2 database, both containing a genus and species annotation (Table S2-S4). For the entire Colpodea, the molecular database itself only contained about a third or less of the genus and species entries contained in the morphospecies database (Fig. 3A-D), because two thirds of the molecular

database was inflated with more than one sequence for the same species. This causes much uncertainty with regard to a well-studied territory like Europe (Foissner 2006), within a class with many flagship species(Foissner et al. 2011; Foissner et al. 2014) and experts work on putting environmental sequences in a phylogenetic context with reference sequences (Dunthorn et al. 2014).This discrepancy may even be exacerbated by an average synonymy rate of 20% for ciliates (Foissner et al. 2008).

When relating the inventory data for the ciliate groups mentioned above (Table S2-S4) to phylogenetic analyses (Figs. 2, S1-S3), database discrepancies become clearer. Three novel paraphyletic branches with >97% global dissimilarity were detected for the genus *Aspidisca* in the order Euplotia (Fig. S3). These three scOTUs (OTU1 to 3) could very well have been one of the 70 morphologically described species in the databases for which only 22 reference sequences (15 species) exist in the molecular databases (PR^2). At least 6 such branches (OTU1-4, OTU10-11, OTU12-14, OTU22-27, OTU 30-31, OTU44-48) could be detected in the well studied order of Colpodida alone (Fig 2). HTS studies often discover deep branching novel diversity among well described clades (Dunthorn et al. 2014; Šlapeta et al. 2005) and yet high hidden diversity (Foissner et al. 2008). We suggest that the very divergent sequence clusters within monophyletic groups from environmental study samples are orphaned flagship species. These flagship species are often not found in molecular databases. This could very well be the case in the order Colpodida containing many new phylogenetic scOTUs entries among stable and slow-evolving taxa within the phylogenetic tree (Fig. 2).

Gradients detected usingbbOTUs. Environmental gradients and spatial scaling for 150 grassland habitats indicated several mechanisms that influenced protist community structure (Figs. 4A-D). The environmental gradient influencing ciliate communities' dispersion (Figs. 4B-D) was comparatively less obvious for database defined bbOTUs (Figs. S5A, C) than for database independent UIRs (Fig. S5B). Because gene variants existed for the same species, especially for dominant taxa in soil (Figs. 2, S1-S3), we assumed them to

representimportant ecological units (Díaz et al. 2006; Pawlowski et al. 2016). In our data, the presence-absence patterns per site were different for some variants of the same dominant species (Fig. S4), which suggested differences in ecological parameter preferences among some variants.

We detected the influence of geographical separation on OTU communities, especially between exploratory sites located further apart (AEG and SEG, Fig. 4A) as also seen in the taxa-area relationships (Fig. S5). We also discovered that no one single factor influenced OTU separation, even though some strong gradients were detected (soil water content and soil type) to influence ordination graphs of species by site (Fig. 4B-C). The effect of pH might directly influence ciliate distribution but might also indirectly affect ciliate community structure due to its strong effect on bacterial communities (Fig. 4B) as determined in the same sampling plots (Gossner et al. 2016; Kaiser et al. 2016). Despite this fact, database shortcomings are apparent among taxa with few reference sequences leading to less conclusive indications for the influence of environmental parameters, e.g. Heterotrichea, Prostomatea and Nassophorea (Fig. 4C). Abundant taxa were more influenced by edaphic factors and land-use (Fig. 4D), which made them important candidates for biomonitoring. Because ellipses in the NMDS graphs (Fig. 4A-D) indicated Bray-Curtis distances between sites and species communities to the centroid (Oksanen 2017), we assumed that taxonomic groups located further away indicated less important candidates for biomonitoring. Since the first study of ciliates in environmental samples using 18S rRNA gene sequences (Dopheide et al. 2008), it has become clear that group specific primers (Geisen et al. 2015; Esteban et al. 2006) and a combination of molecular markers (Dunthorn et al. 2014) increase the chance at detecting micro-eukaryotes and therefore their importance as biomonitoring taxa. HTS information becomes most useful when related to environmental data (Pawlowski et al. 2016), but because only a small number of all ciliates have been described thus far (Foissner et al. 2008), only a small part of known taxa are rediscovered in environmental HTS surveys (De Vargas et al. 2015).

High-throughput sequencing is a powerful tool in detecting ciliates in soil. Well studied taxa in molecular databases may contain multiple gene variants for single dominant species within well studied taxa (e.g.Spirotrichea and Colpodea) where gene variants can present own biogeography. This otherwise called microdiversitycan be important to indicate gradients (e.g., land-use intensity, soil water) influencing community structure. The discrepancy between molecular and morphological databases may cause partial or inaccurate identification and even cause inconclusive biomonitoring results for taxa not well presented in these databases.

Methods

Data collection and soil sampling. Soil samples were collected in May 2011 from three different topo-geographic regions representing large parts (~44 000 km) of Central Europe (Germany, see Venter et al., in press). These samples cover 150 grasslands from the German Biodiversity Exploratories initiative (http://www.biodiversity-exploratories.de/) (Fischer et al. 2010) encompassing three temporally and spatially scaled geo-referenced study plots: the UNESCO biosphere reserve Schorfheide-Chorin (SEG) in north-eastern Germany, Hainich-Duen national park (HEG) in central Germany and the Schwaebische Alb UNESCO biosphere reserve (AEG) in south-western Germany. Standardized field sampling (Fischer et al. 2010; Brabender et al. 2012) was performed on samples taken 1 – 1000 km apart (mesoscale). To summarize the procedure, 20 x 20 m size grassland plots were selected representing a range of land-use intensities (LUI). At each site, 14 soil sub-samples were cored out (diameter, 8.3 cm) from the upper most 10 cm of the A horizon and the top most 5 cm root-layer was removed, excluding particles >2 cm in diameter. Cores were combined, homogenized and composite samples stored at 4°C while still at field moisture content. LUI index for the year 2011 was calculated from fertilization intensity (organic and mineral fertilization excluding livestock dunging), mowing frequency and grazing intensity (livestocks). Data was obtained from land owners by questionnaires (Blüthgen et al. 2012) and applied in this study to test the effect of land management on species richness and composition.

DNA isolation, PCR amplification and NGS. Whole genomic DNA was extracted from 1 g of each composite sample using the PowerSoil® DNA isolation kit (Mo Bio Laboratories, Inc., Carlsbad, CA). DNA concentration was measured using the Nanodrop 100 spectrophotometer (Thermo Fisher Scientific, Germany) and adjusted to 100 ng/µl using ddH_2O. The V4 region of the 18S rRNA gene was directly amplified from the samples using eukaryotic specific primers 590F (5'-3':CGGTAATTCCAGCTCCAATAGC) and 1300R (5'-3':CACCAACTAAGAACGGCCATGC). To separate the sequences, the Titanium primer

design and the recommended multiplex identifier (MID) adapter complex design (Roche, Germany) method was used. The pre-454 sequencing PCR reaction mixture (25 µl) contained: 2µl (100 ng/µl) DNA, 2 µl 10x DNA polymerase buffer with 20 mM $MgSO_4$, 2µl (1 µM) 590F primer + Adaptor A, 2µl (1 µM) 1300R primer, 2µl (2 mM) dNTP each and 0.4 µl (2.5 U/µl) Pfu DNA polymerase (Fermentas, Germany). Cycling conditions were: initial denaturation at 95°C for 3 min; 30 cycles at 95°C for 30 s, 55°C for 30 s, 72°C for 2 min; and a final extension at 72°C for 10 min. Each sample was amplified in triplicate and pooled to a final concentration of 20 ng/µl to eliminate possible PCR biases. NGS using the GS-FLX sequencer and Titanium sequencing kit XLR70 (Roche, Germany) was performed by GATC Biotech AG, Germany. Sequencing was done from the forward primer (adaptor A).

Bioinformatics. DNA sequences were demultiplexed and filtered for (1.) 100% forward primer match to remove false positive PCR amplifications of non-rRNA genes; (2.) minimum sequence length of 200bp and (3.) a maximum sequence length of 710bp - to remove possible sequencing artefacts; and (4.) ambiguities (N's), to exclude sequences containing uncertain base pairs. Sequences were scanned for chimeric sequences against the Protist Ribosomal (PR^2 v203) reference database (Guillou et al. 2013) using the uchime_refalgorithm in the USEARCH v. 7.0.1090 package(Edgar et al. 2011) and trimmed to a maximum length of 530bp to avoid terminally located read errors and focus downstream analyses on the V4 region of interest (Niklas et al. 2013). Trimmed reads were dereplicated into 100% identical unique individual reads (UIRs) using the VSEARCH script (Rognes et al. 2016) to identify singletons and align sequences to the database.

Taxonomic analysis of sequencing data. UIRs with an initial abundance of 1 were termed singletons and removed to circumvent the dangers of pyrosequencing related artifacts (Tedersoo et al. 2010). UIRs were annotated by its closest matching reference sequence (accession number) in the Protist Ribosomal Reference (PR^2) database (GenBank, version 203, downloaded June 2016) for 18S rDNA sequences (Guillou et al. 2013). This applied a

form of closed-reference-clustering, where the accession number was used to identify operational taxonomic units (bbOTUs). Because multiple individual UIRs could be linked to a single accession number, a more stabilized and accurate information on complex communities could be ascertained. This compensated for pyrosequencing errors and inferred an upper (UIRs) and lower (bbOTUs) estimate of the actual species richness of protist ciliates in soil (see Venter et al., in press). Default Blastn parameters (open gap penalty 5, cost gap extension penalty 2, nucleic match 2, nucleic penalty mismatch -3 and word size 11) were applied and single hits were retained if E-value $\leq e^{-100}$. UIRs with matching accession numbers were grouped into single bbOTUs. Lineages for UIRs were inspected for ambiguous identification using the metagenome analyzer (MEGAN v. 5) program. Using 50 Blastn hits per query sequence, conserved sequences were correctly identified to the high-order taxa in the database, based on the lowest common ancestor (LCA) algorithm in MEGAN (Huson et al. 2007). bbOTUs with 100% coverage and annotations to taxa for which the primer region is not a suitable (Pawlowski et al. 2012) as well as non-protistan taxa (Metazoa, Fungi and Embryophyta) were removed.

OTUs vs. morphospecies richness. Determining species richness in environmental samples using NGS presented the added difficulty of determining species units (Caron et al. 2009). For this purpose UIRs were further binned based on pairwise identity cut-off values to determine species richness at taxonomic levels. Low pairwise identity values might have binned congeneric UIRs (Caron et al. 2009; Nebel et al. 2011) and too high values might have excluded too many UIRs. Due to evolutionary rate differences within the V4 region of the SSU across taxa (Nebel et al. 2011) as well as sequencing and PCR error rates (Huse et al. 2007; Niklas et al. 2013; Stoeck et al. 2010), UIRs with single base differences (≥99.7% sequence similarity) were binned to the species taxonomic level. The commonly used ≥97% pairwise identity level could be used as a conservative proxy level to separate genera (Caron et al. 2009), but was used in this study to compare results with other studies. The ≥80%

pairwise identity and last level for inclusion, stood as proxy for class taxonomic level (Stoeck et al. 2010). UIRs with<80% pairwise identity were excluded. We compared the OTUs found with this method with species described in Foissner (1999) and Berger (2001) and the webseite of Encyclopaedia of Life (www.eol.com).

Phylogenetic analysis of sequencing data. Complementary to local pairwise alignment, we placed amplicons into a phylogenetic context with reference sequences for known diversity. Query sequences (UIRs) together with their hit reference sequences (bbOTUs) from the PR^2 v203 database were compiled into an input FastA file for clustering using the global alignment USEARCH v9.0.2132 program (Edgar 2010). Using the-*cluster smallmem* and -*sortedby size* options, the USEARCH package algorithm matched all sequences to all seeds until all hits to a centroid were found at ≥97% sequence identity. Centroids were aligned to sequences for standardized reference trees (Colpodea, Spirotrichea and Litostomatea) selected from literature (Foissner et al. 2014; Gao et al. 2016). Multiple sequence alignment was performed using the MAFFT v7.123b (Guindon and Gascuel 2003) and MUSCLE v3.8.31 algorithm(Edgar 2004). Ambiguously aligned columns were corrected by hand. Maximum likelihood (ML) analyses were carried out in RAxML-HPC v7.2.8 using GTR + I + G model for distribution of evolutionary rates across sites as determined using the MrAIC.pl 1.4.6 modeltest(Nylander 2004) and the following parameters: raxmlHPC -f a -x 12345 -p 12345. Molecular protist ribosomal reference (PR^2) database data for the taxa Colpodea, Hypotrichia and Euplotia was compared to next-generation sequencing data and morphologically described genusses and species using the lineage (order, family/superorder, genus and species) according to Adl et al. (2012). An inventory for morphologically described species were retrieved from literature (Foissner 1993; Berger 2001;) and the online Encyclopaedia Of Life (EOL – http://www.eol.org) using the respective genera names as keywords.

Statistical analysis. bbOTU data was converted to presence-absence matrices to avoid

the effect of inflated abundances. Species richness could be derived from the number of bbOTUs/UIRs per site and within each region using the collective binary data at taxonomic levels. Statistical analysis was conducted in R software version 3.3.1 (Core Team R, 2014), using the Vegan 2.0-7 (Oksanen 2016) package. Rank abundance analysis (Fisher log-series) was performed using the "fisherfit" function to determine the number of bbOTUs (taxa) across 150 sites.Species accumulation curves using the Kindt's exact method under the "specaccum()" function was performed from random permutations of the presence-absence data to calculate the standard deviation and compute species richness. An asymptote indicates saturation for the taxa richness uncovered during the sampling effort. Correlations between taxa richness and global LUI values (Blüthgen et al. 2012) were performed using the "cor.test()" function in R by method of the Spearman rank test to produce significance values.Dependencies among variables were tested for by means of the "chisq.test()" (chi-squared test) function in R.Multidimensional scaling (MDS) and non-metric multidimensional scaling (NMDS) using Bray-Curtis dissimilarity matrixes, were carried out using the "cmdscale()" and "metaMDS()" functions in the 'Vegan' package to illustrate constrained community ordination and random starting configurations, finally adding species scores as weighted averages. Ordinations for vectors based on continuous variables and centroids were fitted using the "envfit()" function, applying 999 permutations.Illustrations were done using the 'gplots' package (http://cran.r-project.org/web/packages/gplots/index.html).

Acknowledgements

We acknowledge bioinformatical and programming help by Mr. Johannes Schoeneich and Dr. Peter Heger. The work was supported by grants from the German Research Foundation (DFG) to H.A. (grant number AR 288/16-1,2, INST 216/862-1), a stipendium of the German Academic Exchange Service (DAAD) to P.V. (grant number 91525927). We thank the managers of the three Exploratories, SwenRenner, Sonja Gockel, Martin Gorke and all former managers for their work in maintaining the plot and project infrastructure; Simone Pfeiffer for giving support through the central office, Jens Nieschulze for managing the central database, and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project. The work has been partly funded by the DFG Priority Program 1374 "Infrastructure-Biodiversity-Exploratories". Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg (according to § 72 BbgNatSchG).

References

Acosta-Mercado D, Lynn DH (2004) Soil ciliate species richness and abundance associated with the rhizosphere of different subtropical plant species. J Eukar Microbiol **51**:582–588

Adl SM, Simpson AGB, Lane CE, Lukes J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EAD, Mozley-Stanridge SE, Parfrey LW, Pawlowski J,

Rueckert S, Shadwick L, Schoch CL, Smirnov A, Spiegel FW (2012) The revised classification of eukaryotes. J Eukar Microbiol **59**:429–493

Baroin-Tourancheau A, Delgado P, Perasso R, Adoutte A (1992) A broad molecular phylogeny of ciliates: identification of major evolutionary trends and radiations within the phylum. Proc Natl Acad Sci U S A **89**:9764–9768

Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R, Fierer N (2013) Global biogeography of highly diverse protistan communities in soil. ISME J**7**:652–659

Berger H (2001) Catalogue of ciliate names 1. Hypotrichs. Verlag Helmut Berger, Salzburg, pp 1-206

Blüthgen N, Dormann CF, Prati D, Klaus VH, Kleinebecker T, Hölzel N, Alt F, Boch S, Gockel S, Hemp A, Müller J, Nieschulze J, Renner SC, Schöning I, Schumacher U, Socher SA, Wells K, Birkhofer K, Buscot F, Oelmann Y, Rothenwöhrer C, Scherber C, Tscharntke T, Weiner CN, Fischer M, Kalko EKV, Linsenmair KE, Schulze E-D, Weisser WW (2012) A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. Basic ApplEcol **13**:207–220

Bonkowski M (2004) Protozoa and plant growth: The microbial loop in soil revisited. New Phytologist **162**:617–631

Brabender M, Kiss ÁK, Domonell A, Nitsche F, Arndt H (2012) Phylogenetic and Morphological Diversity of Novel Soil Cercomonad Species with a Description of Two New Genera (Nucleocercomonas and Metabolomonas). Protist **163**:495–528

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD, Dennett MR, Moran DM, Jones AC (2009) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. Appl Environ Microbiol **75:**5797–5808

Chao A, Li PC, Agatha S, Foissner W (2006) A statistical approach to estimate soil ciliate diversity and distribution based on data from five continents. Oikos **114**:479–493

Core Team R (2013) R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna).

De Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J-M, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horák A, Jaillon O, Lima-Mendez G, Lukeš J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans Coordinators: Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E (2015) Eukaryotic plankton diversity in the sunlit ocean. Science 348:1261605

Díaz S, Martín-González A, Gutiérrez JC (2006) Evaluation of heavy metal acute toxicity and bioaccumulation in soil ciliated protozoa. Environment International **32**:711–717

Dopheide A, Lear G, Stott R, Lewis G (2008) Molecular characterization of ciliate diversity in stream biofilms. Appl Environ Microbiol **74**:1740–1747

Dunthorn M, Hall M, Foissner W, Stoeck T, Katz LA (2014)Broad taxon sampling of ciliates using mitochondrial small subunit ribosomal DNA. Acta Protozoologica **53**:207–213

Dunthorn M, Otto J, Berger SA, Stamatakis A, Mahé F, Romac S, de Vargas C, Audic S, Consortium B, Stock A, Kauff F, Stoeck T (2014) Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. Mol Biol Evol **31**:993–1009

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**:1792–1797

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26**:2460–2461

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**:2194–2200

Ekelund F, Frederiksen HB, Rønn R (2002) Population dynamics of active and total ciliate populations in arable soil amended with wheat . Appl Environ Microbiol **68**:1096–1101

Encyclopedia of Life. Available from http://www.eol.org. Accessed Dec 2016.

Esteban GF, Clarke KJ, Olmo JL, Finlay BJ (2006) Soil protozoa - An intensive study of population dynamics and community structure in an upland grassland. Appl Soil Ecol **33**:137–151

Fischer M, Bossdorf O, Gockel S, Hänsel F, Hemp A, Hessenmöller D, Korte G, Nieschulze J, Pfeiffer S, Prati D, Renner S, Schöning I, Schumacher U, Wells K, Buscot F, Kalko EKV, Linsenmair KE, Schulze E-D, Weisser WW (2010) Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic Appl Ecol **11**:473–485

Foissner W (1993) Protozoenfauna Volume 4/1. Colpodea (Ciliophora) MD (ed) Gustav Fischer Verlag, Stuttgard, pp 1-782

Foissner W (1999) Soil protozoa as bioindicators: Pros and cons, methods, diversity, representative examples. Agriculture, Ecosystems and Environment **74**:95–112

Foissner W (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta Protozoologica **45**:111–136

Foissner W (2008) Notes on soil ciliates from Singapore, with description of Suturothrix monoarmata nov. gen., nov. spec. (Protozoa, Ciliophora). Soil Organisms **80**:81–97

Foissner W (2016) Protists as bioindicators in activated sludge: Identification, ecology and

future needs. Eur J Protistol **55**:75–94

Foissner W, Bourland WA, Wolf KW, Stoeck T, Dunthorn M (2014) New SSU-rDNA sequences for eleven colpodeans (Ciliophora, Colpodea) and description of Apocyrtolophosis nov. gen. Eur J Protistol **50**:40–46

Foissner W, Chao A,Katz LA (2008) Diversity and geographic distribution of ciliates (Protista: Ciliophora). Biodivers Conserv **17**:345–363

Foissner W, Quintela-Alonso P, Al-Rasheid K (2008) Soil ciliates from Saudi Arabia, including descriptions of two new genera and six new species. Acta Protozoologica **47**:317–352

Foissner W, Stoeck T, Agatha S, Dunthorn M (2011) Intraclass Evolution and Classification of the Colpodea (Ciliophora). J Eukar Microbiol **58**:397–415

Forster D, Behnke A, Stoeck T (2012) Meta-analyses of environmental sequence data identify anoxia and salinity as parameters shaping ciliate communities. Syst Biodivers **10**:277–288

Forster D, Dunthorn M, Mahé F, Dolan JR, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie J-M, Decelle J, Edvardsen B, Egge E, Eikrem W, Gobet A, Kooistra WHCF, Logares R, Massana R, Montresor M, Not F, Ogata H, Pawlowski J, Pernice MC, Romac S, Shalchian-Tabrizi K, Simon N, Richards TA, Santini S, Sarno D, Siano R, Vaulot D, Wincker P, Zingone A, de Vargas C, Stoeck T (2016) Benthic protists: The under-charted majority. FEMS Microbiol Ecol **92**:1–11

Gao F, Warren A, Zhang Q, Gong J, Miao M, Sun P, Xu D, Huang J, Yi Z, Song W (2016) The All-Data-Based Evolutionary Hypothesis of Ciliated Protists with a Revised Classification of the Phylum Ciliophora (Eukaryota, Alveolata). Sci Rep **6**:24874

Geisen S, Tveit AT, Clark IM, Richter A, Svenning MM, Bonkowski M, Urich T (2015) Metatranscriptomic census of active protists in soils. ISME J **9**:1–13

Gossner MM, Lewinsohn TM, Kahl T, Grassein F, Boch S, Prati D, Birkhofer K, Renner

SC, Sikorski J, Wubet T, Arndt H, Baumgartner V,Blaser S, Blüthgen N, Börschig C, Buscot F, Diekötter T, Ré Jorge L, Jung K, Keyel AC, Klein A-M, Klemmer S, Krauss J, Lange M, Müller J, Overmann J, Pašalić E, Penone C, Perović DJ, Purschke O, Schall P, Socher SA, Sonnemann I, Tschapka M, Tscharntke T, Türke M, Venter PC, Weiner CN, Werner M, Wolters V, Wurst S, Westphal C, Fischer M, Weisser WW, Allan E (2016) Land-use intensification causes multitrophic homogenization of grassland communities. Nature **540**:266–269

Grattepanche J-D, McManus GB,Katz LA (2016) Patchiness of ciliate communities sampled at varying spatial scales along the New England Shelf. Plos One 11:e0167659

Grossmann L, Jensen M, Heider D, Jost S, Glücksman E, Hartikainen H, Mahamdallie SS, Gardner M, Hoffmann D, Bass D, Boenigk J (2016) Protistan community analysis: key findings of a large-scale molecular sampling. ISME J **10**:2269-2279

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet A-L, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R (2013) The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res **41**:597–604

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52**:696–704

He Y, Xu K (2011) Morphology and small subunit rDNA phylogeny of a new soil ciliate, bistichella variabilis n. sp. (Ciliophora, Stichotrichia). J Eukar Microbiol **58**:332–338

Huse SM, Hober JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively-parallel DNA pyrosequencing. Genome Biol **8**:R143

Huson D, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenome data.

Gennome Res **17**:377–386

Kaiser K, Wemheuer B, Korolkow V, Wemheuer F, Nacke H, Schöning I, Schrumpf M, Daniel R (2016) Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forrests. Sci Rep 6:33696

Lara E, Acosta-Mercado D (2012) A molecular perspective on ciliates as soil bioindicators. Eur J Soil Biol **49**:107–111

Larkin AA, Martiny AC (2017) Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. Environ Microbiol Rep **9**:55-70

Mahé F, de Vargas C, Bass D, Czech L, Stamatakis A, Lara E, Singer D, Mayor J, Bunge J, Sernaker S, Siemensmeyer T, Trautmann I, Romac S, Berney C, Kozlov A, Mitchell E, Seppey C, Egge E, Lentendu G, Wirth R, Trueba G, Dunthorn M (2017) Soil protists in three neotropical rainforests are hyperdiverse and dominated by parasites. bioRxiv 1–17

Massana R, Gobet A, Audic S, Bass D, Bittner L, Boutte C, Chambouvet A, Christen R, Claverie J-M, Decelle J, Dolan JR, Dunthorn M, Edvardsen B, Forn I, Forster D, Guillou L, Jaillon O, Kooistra WHCF, Logares R, Mahé F, Not F, Ogata H, Pawlowski J, Pernice MC, Probert I, Romac S, Richards T, Santini S, Shalchian-Tabrizi K, Siano R, Simon N, Stoeck T, Vaulot D, Zingone A, de Vargas C (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. Environ Microbiol **17**:4035–4049

Nebel M, Pfabel C, Stock A, Dunthorn M, Stoeck T (2011) Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. Environ Microbiol Rep **3**:154–158

Niklas N, Pröll J, Danzer M, Stabentheiner S, Hofer K, Gabriel C (2013) Routine performance and errors of 454 HLA exon sequencing in diagnostics. BMC Bioinformatics **14**:176

Nylander JAA (2004) MrAIC.pl. Program distributed by the author. Evolutionary Biology

Centre, Uppsala University.

Nylander JAA. MrModeltest 2.0 Program distributed by the author. Evolutionary Biology Centre, (2004) Uppsala Univ. (http://www.ebc.uu.se/systzoo/staff/nylander.html)

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin R, O'Hare RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2013) Vegan: Community ecology package. R package version 2.0-7. http://CRAN.R-project.org/package=vegan

Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M, Fiore-Donno AM, Gile GH, Holzmann M, Jahn R, Jirků M, Keeling PJ, Kostka M, Kudryavtsev A, Lara E, Lukeš J, Mann DG, Mitchell EAD, Nitsche F, Romeralo M, Saunders GW, Simpson AGB, Smirnov AV, Spouge JL, Stern RF, Stoeck T, Zimmermann J, Schindel D, de Vargas C (2012) CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. PLoS Biol **10**:e1001419

Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA (2014) Environmental monitoring through protist next□generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. Mol Ecol Res **14**:1129–1140

Pawlowski J, Lejzerowicz F, Apotheloz-Perret-Gentil L, Visco J, Eslingg P (2016) Protist metabarcoding and environmental biomonitoring: Time for change. EuropJ Protistol **55**:12–25

Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, Chase J, McDonald D, Gonzalez A, Robbins-Pianka A, Clemente JC, Gilbert JA, Huse SM, Zhou H-W, Knight R, Caporaso JG (2014) Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. PeerJ **2**:e545

Rogerson A, Detwiler A (1999) Abundance of airborne heterotrophic protists in ground level air of South Dakota. Atmos Res **51**:35–44

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ **4**:e2409v1

Schoenle A, Jeuck A, Nitsche F, Venter P, Prausse D, Arndt H (2016) Methodological studies on estimates of abundance and diversity of heterotrophic flagellates from the deep-sea floor. J Mar Sci Eng **4**:1-11

Šlapeta J, Moreira D, López-García P (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. Proc Biol Sci **272**:2073–2081

Soliveres S, Manning P, Prati D, Gossner MM, Alt F, Arndt H, Baumgartner V, Binkenstein J, Birkhofer K, Blaser S, Blüthgen N, Boch S, Böhm S, Börschig C, Buscot F, Diekötter T, Heinze J, Hölzel N, Jung K, Klaus VH, Klein A-M, Kleinebecker T, Klemmer S, Krauss J, Lange M, Morris EK, Müller J, Oelmann Y, Overmann J, Pašalić E, Renner SC, Rillig MC, Schaefer HM, Schloter M, Schmitt B, Schöning I, Schrumpf M, Sikorski J,Socher SA, Solly EF, Sonnemann I, Sorkau E, Steckel J, Steffan-Dewenter I, Stempfhuber B, Tschapka M, Türke M, Venter P, Weiner CN, Weisser WW, Werner M, Westphal C, Wilcke W, Wolters V, Wubet T, Wurst S, Fischer M, Allan E (2016) Locally rare species influence grassland ecosystem multifunctionality. Philos TransR Soc LondB Biol Sci **371**:20150269

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, Richards TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. MolEcol **19**:21–31

Tedersoo L, Nilsson RH, Aberenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G, Kõljalq U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. New Phytologist **188**:291–301

Urich T, Lanzèn A, Qi J, Huson DH, Schleper C, Schuster SC (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PloS one **3**:e2527.

Venter PC, Nitsche F, Domonell A, Heger P, Arndt H (2017) The Protistan Microbiome of

Grassland Soil: Diversity in the Mesoscale. Protist **168**:546-564

Supplementary Material

Discrepancies between molecular and morphological databases of soil ciliates studied for temperate grasslands of central Europe by Paul Christiaan Venter, Frank Nitsche, Anja Scherwass, Hartmut Arndt.

Table S1. Summary of the V4 SSU gene sequences discovered in 150 grassland soil samples from the mesoscale (central Europe - Germany). Class and order identification for ciliate taxa and non-ciliate supergroups are given in tally for unique individual read (UIRs) sequences and Blast based operational taxonomic units (bbOTUs) matched to taxa in the protist ribosomal reference (PR^2) database. Matches at three levels of sequence pairwise identity (≥80%, ≥97% and ≥99.7%) are given. Taxa percentages are given in brackets.

Sequence source (supergroup, class or order[a])	≥80 Sequence identity		≥97 Sequence identity		≥99.7 Sequence identity	
	No. of UIR[b]	No. of OTUs[c]	No. of UIR[b]	No. of OTUs[c]	No. of UIR[b]	No. of OTUs[c]
Armophorea						
Armophorida	1	1				
Cariacotrichea						
Cariacotrichida	1	1				
Uncultured ciliate						
Ciliophora-6_X	3	1				
Colpodea						
Bursariomorphida	29	5	16	5	1	1
Colpodida	570	31	523	29	10	9
Cyrtolophosidida	71	7	44	6	1	1
Platyophryida	81	7	81	7	3	2
Heterotrichea						
Heterotrichida	7	2	7	2	1	1
Litostomatea						
Cyclotrichia	7	2	3	1		
Haptoria	230	33	203	30	12	4
Trichostomatia	3	1	3	1		
Nassophorea						
Nassophorea_X	35	7	27	5		
Oligohymenophorea						
Astomatia	68	7	2	1		
Hymenostomatia	8	1				
Peniculia	6	4	1	1		
Peritrichia	40	8	14	4		
Scuticociliatia	68	10	47	3	2	1
Sessilida	24	10	18	9	4	3
Tetrahymenida	17	4	17	4	3	3
Phyllopharyngea						
Suctoria	3	2				
Synhymenia	2	2				
Prostomatea						
Prostomatea_X	2	2				
Prostomatea-3	1	1	1	1		
Spirotrichea						
Choreotrichia	5	1	5	1		
Euplotia	15	2				
Hypotrichia	1519	74	1300	63	248	44
Oligotrichia	2	1	2	1	1	1
Spirotrichea_X	3	1	3	1		
Total ciliates	2821	228	2317	175	286	70

[a] Classes and orders for the Ciliophora supergroup.

[b] Unique individual read (UIR) sequences which were identified by Blast to reference sequences in the PR2 database.

[c] Operational taxonomic units (OTUs) or accession numbers identified according to PR2 Blast matches. Percentages of ciliates are given in brackets.

Table S2. Molecular protist ribosomal reference (PR^2) v203 database curated sequence inventory compared to next-generation sequencing (NGS) data discovered in grassland soil and morphologically described genusses and species within the class Colpodea. Lineage breakdown (Oder, Family/Superorder, Genus and Species) follows the revised classification of Eukaryotes by Adl et al. (2012), where unknown or unresolved family and superorder clade lineages are indicated ending with "_X". A breakdown of NGS data indicates the number of unique individual reads (UIRs) and operational taxonomic units (OTUs) - UIRs grouped to the same accession number was counted as an OTU. Empty blocks indicate that OTUs or UIRs were not detected or respective morphologically described species were not listed. Morphologically described species are given for by Foissner et al. (1993) and the online encyclopaedia of life (EOL). #Database Acc. Num. - Number of database accountable accession numbers; #Species described - Number of morphologically described genusses and species catalogued by Foissner et al. (1993) and EOL. *Foissner W (1993) Protozoenfauna Volume 4/1. Edited by Pro. Dr. Dieter Matthes, Colpodea (Ciliophora), W. Foissner 1993. Stuttgart (Germany): Gustav Fischer Verlag. **Number of Morphospecies described and catalogued in the Encyclopaedia of Life (EOL - www.eol.org).

	Order	Family/Superorder	Genus	Species listed in PR^2 and found in soil	#Database Acc. Num.	UIRs	OTUs	#Species described	Described morphospecies - Foissner et al. 1993*	Described morphospecies - EOL**
Unknown Colpodea	Colpodea_X	Colpodea_XX	Colpodea_XXX	Colpodea_XXX sp.	1			167	7 Orders: Colpodida, Grossglockneriida, Bursariomorphida, Sorogenida, Bryophryida, Cyrtolophosidida, Bryometopida	7 Orders: Colpodida, Grossglockneriida, Bursariomorphida, Sorogenida, Bryophryida, Cyrtolophosidida, Bryometopida
			Discotricha	*Discotricha papillifera*	1					
			Lopezoterenia	*Lopezoterenia* sp.	1					
	Colpodea-1	Colpodea-1_X	Colpodea-1_XX	Colpodea-1_XX sp.	326					
					Total: 329	0	0	167		
	Bursario-morphida	Bryometopidae	*Bryometopus*	*Bryometopus* sp.	3			9	*B. viridis, B. chlorelligerus, B. balantidioides, B. edaphonus*	*B. alekperovi, B. balantidioides, B. chlorelligerus, B. edaphonus, B. hawaiiensis, B. magnus, B. muscicola, B. triquetus, B. viridis*
				Bryometopus atypicus	1			1	*B. atypicus*	*B. atypicus*
				Bryometopus pseudochilodon	2	7	1	1	*B. pseudochilodon*	*B. pseudochilodon*
				Bryometopus sphagni	1	3	1	1	*B. sphagni*	*B. sphagni*
		Bursariomorphida_X	*Bryometopus*	*Bryometopus triquetrus*	1	2	1	1	*B. triquetrus*	*B. triquetrus*
			Bursariomorphida_XX	Bursariomorphida_XX sp.	3	3	1			
		Bursariidae	*Bursaria*	*Bursaria* sp.	6	1	1	2	*B. caudata, B. ovata*	Genus only, no species
				Bursaria truncatella	1			1	*B. truncatella*	
					Total: 18	16	5	16		

Table S2 continued.

Order	Family/Superorder	Genus	Species listed in PR^2 and found in soil	#Database Acc. Num.	UIRs	OTUs	#Species described	Described morphospecies - Foissner et al. 1993*	Described morphospecies - EOL**
Colpodida	Colpodida_X	Colpodida_XX	Colpodida_XX sp.	2	26	1			
		Bromeliothrix	*Bromeliothrix metopoides*	1			1	Not listed as a genus under Colpodida	Unclassified Colpodida: *B. metopoides*
		Jaroschia	*Jaroschia sumptuosa*	1			1	*J. sumptuosa*	*J. sumptuosa*
		Kalometopia	*Kalometopia duplicata*	1	7	1	2	*K. duplicata, K. perronnei*	*K. duplicata, K. perronnei*
		Pseudomaryna	*Pseudomaryna* sp.	1			1	*P. australiensis* (Foissner, 2003)	*P. australiensis*
		Repoma	*Repoma cavicola*	1			1	1970 by Novotny for Colpoda)	*R. cavicula*
	Bardeliellidae	*Bardeliella*	*Bardeliella pulchra*	1			1	*B. pulchra*	*B. pulchra*
	Bryophryidae	*Bryophrya*	*Bryophrya gemmea*	1			5	*B. bavariensis, B. bavariensis minor, B. rubescens, B. flexilis*	*B. gemmea*
		Bryophryoides	*Bryophryoides ocellatus*	1			1	Not listed	*B. ocellatus*
		Notoxoma	*Notoxoma parabryophryides*	1			2	*N. pararyophryides, N. sigmoides*	*N. parabryophryides*
	Colpodidae	*Bresslaua*	*Bresslaua vorax*	1	8	1	4	*B. vorax, B. insidiatrix, B. sicaria, B. dissimilis*	Only genusses indicated. No species
		Colpoda	*Colpoda* sp.	13	55	2	20	*C. colpidiopsis, C. augustini, C. simulans C. californica, C. edaphoni, C. gigantea, C. grandis, C. irregularis, C. reniformis, C. ovinucleata, C. orientalis, C. tripartita, C. cavicola, C. distincta, C. praestans, C. variabilis, C. flavicans, C. simulans, C. oblonga*	*C. spiralis*
			Colpoda aspera	2	10	2	1	*C. aspera*	
			Colpoda cucullus	2	39	1	1	*C. cucullus*	*C. cucullus*
			Colpoda ecaudata	1			1	*C. ecaudata*	
			Colpoda elliotti	1	49	1	1	*C. elliotii*	
			Colpoda henneguyi	1			1	*C. henneguyi*	*C. henneguyi*
			Colpoda inflata	3	35	2	1	*C. inflata*	*C. inflata*
			Colpoda lucida	2	3	1	1	*C. lucida*	
			Colpoda magna	1			1	*C. magna*	
			Colpoda maupasi	4	20	1	1	*C. maupasi*	*C. maupasi*
			Colpoda minima	1			1	*C. minima*	
			Colpoda sp. strain1	1					
			Colpoda steinii	8	31	2	1	*C. steinii*	*C. steinii*
		Exocolpoda	*Exocolpoda augustini*	2	16	2		see *Colpoda augustini*	
		Colpodidae_X	Colpodidae_X sp.	11	137	4			
		Kreyellidae	*Kreyellidae* sp.	1	9	1	5	*Two Genera (Kreyella, Orthokreyella) K. minuta, K. muscicola, K. duplicata, K. Perronnei, O. Schimanni*	*K. minuta, K. muscicola*

Table S2 continued.

Order	Family/Superorder	Genus	Species listed in PR^2 and found in soil	#Database Acc. Num.	UIRs	OTUs	#Species described	Described morphospecies - Foissner et al. 1993*	Described morphospecies - EOL**
	Grossglockneriidae	*Mykophagophrys*	*Mykophagophrys terricola*	1	2	1	2	*M. armata, M. terricola* (Foissner, 1995)	*M. terricola*
		Pseudoplatyophrya	*Pseudoplatyophrya nana*	1	53	1	2	*P. saltans*, *P.terricola* (Foissner, 1985, 1988)	Class: Pseudoplatyophrya only.
	Hausmanniellidae	*Bresslauides*	*Bresslauides discoideus*	2	4	1	3	*B. discoideus, B. australis, B.terricola*	Only genusses indicated. No species
		Hausmanniella	*Hausmanniella discoidea*	1	7	1	3	*H. discoidea, H. patella, H.quinquecirrata*	*H. discoidea, H. patella, H. quinquecirrata*
	Marynidae	*Maryna*	*Maryna* sp.	1	3	1	16	*M. acuminata, M. antarcia, M. atra, M. cardioides, M. galeata, M. grisea, M. hyalina, M. lichenicola, M. longinucleata, M. magna, M. minima, M. pinguis, M. rottalensis, M. rotunda, M. socialis*	Genus: *Maryna* only.
			Maryna ovata	1	7	1	1	*M. ovata*	
			Maryna umbrellata	1	2	1	1	*M. umbrellata*	
		Ilsiella	*Ilsiella palustris*	2			2	*I. palustris, I. venusta*	Genus: *Ilsiella* only.
				Total: 76	523	29	85		
Cyrtolo-phosidida	Cyrtolophosididae	Cyrtolophosididae_X	Cyrtolophosididae_X sp.	13	29	2	7	*Cyrtolophosis bivacuolata, C. acuta, C. elongata, C. bursaria, C. colpidiformis, C. major*	*Cyrtolophosis bivacuolata, C. acuta, C. elongata, C. bursaria, C. colpidiformis, C. major, C. bivacuolatus*
		Apocyrtolophosis	*Apocyrtolophosis minor*	1			1		*A. minor*
		Aristerostoma	*Aristerostoma* sp.	4			4	*A. minutum*	*Aristerostoma* Assp1, *Aristerostoma* sp. ATCC 50986, *Aristerostoma* sp. CATS 2/II
			Aristerostoma marinum	1			1	*A. marinum*	*A. marinum*
		Cyrtolophosis	*Cyrtolophosis minor*	1			1	*C. minor*	*C. minor*
			Cyrtolophosis mucicola	5	3	2	1	*C. mucicola*	*C. mucicola*
		Prorocentrum	*Prorocentrum lima*	1					
		Pseudocyrtolophosis	*Pseudocyrtolophosis alpestris*	3	11	1	2	*P. alpestris* *P. terricola*	
	Kreyellidae	*Cyrtolophosididae*	*Microdiaphanosoma arcuatum*	1	1	1	2	*M. arcuatum, M. terricola*	*M. arcuatum*
				Total: 30	44	6	19		

Table S2 continued.

Order	Family/Superorder	Genus	Species listed in PR^2 and found in soil	#Database Acc. Num.	UIRs	OTUs	#Species described	Described morphospecies - Foissner et al. 1993*	Described morphospecies - EOL**
Platyophryida	Platyophryidae	*Platyophrya*	*Platyophrya* sp.	6	43	2	11	*P. armata, P. binucleata, P. citrina, P. dubia, P. hyalina, P. macrostoma, P. similis, P. sphagni*	*P. angusta, P. binucleata, P. citrina, P. dubia, P. hyalina, P. macrostoma, P. paoletti, P. similis, P.sphagni, P. terricola*
			Platyophrya bromelicola	1	3	1			
			Platyophrya spumacola	1			1	*P. spumacola*	*P. spumacola*
			Platyophrya vorax	1	24	1	1	*P. vorax*	*P. vorax*
	Platyophryida_X	*Etoschophrya*	*Etoschophrya inornata*	1			1	Not mentioned (only *E. oscillatoriophagain* VÖ: Foissner et al. 2002)	*E. inornata*
		Kuklikophrya	*Kuklikophrya ougandae*	1			1	*K. ougandae*	*K. ougandae*
		Platyophryides	*Platyophryides magnus*	1					
		Platyophryida_XX	Platyophryida_XX sp.	4					
	Sagittariidae	*Sagittaria*	*Sagittaria* sp.	1			3	*S. poligonalis, S. hyalina, S. australis*	*S. poligonalis, S. hyalina, S. australis*
	Sorogenidae	*Ottowphyra*	*Ottowphyra dragescoi*	1			1	O. dragescoi (Foissner, 1987)	*O. dragescoi*
		Platyophryides	*Platyophryides* sp.	1	8	1			
		Sorogena	*Sorogena stoianovitchae*	4	2	1	1	*S. stoianovitchae*	*S.stoianovitchae*
	Woodruffiidae	*Rostrophrya*	*Rostrophrya* sp.	1			5	*R. regis, R. camerounensis, R. terricola*	*Rostrophrya* sp. MSD-2007, *Rostrophrya* sp. Rosp1
		Woodruffides	*Woodruffides metabolicus*	3	1	1	1	*W. metabolicus*	*W. metabolicus*
		Woodruffiidae	*Woodruffiidae* sp.	1			2	*W. metabolicus, W. terricola*	*W. metabolicus, W. terricola*
				Total: 28	81	7	28		

Table S3. Molecular protist ribosomal reference (PR^2) v203 database curated sequence inventory for colpodean ciliates (Alveolata) compared to next-generation sequencing (NGS) data discovered in grassland soil and morphologically described genusses and species within the class Hypotrichia. Lineage breakdown (Oder, Family/Superorder, Genus and Species) follows the revised classification of Eukaryotes by Adl et al. (2012), where unknown or unresolved family and superorder clade lineages are indicated ending with "_X". A breakdown of NGS data indicates the number of unique individual reads (UIRs) and Blast based operational taxonomic units (OTUs) - UIRs grouped to the same accession number was counted as an OTU. Empty blocks indicate that OTUs or UIRs were not detected or respective morphologically described species were not listed. Morphologically described species are given for by Berger (2001)* and the online encyclopaedia of life (EOL). #Database Acc. Num. - Number of database accountable accession numbers; #Species described - Number of morphologically described genusses and species catalogued by Berger (2001) and EOL. *Berger H (2001) Catalogue of ciliate names 1. Hypotrichs. Salzburg (Austria): Verlag Helmut Berger. i-viii and 206 pp. **Number of Morphospecies described and catalogued in the Encyclopaedia of Life (EOL - www.eol.org).

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Cladotrichidae	*Engelmanniella*	*Engelmanniella mobilis*	2	9	1	3	Combined with *Uroleptus mobilis*	*Engelmanniella halseyi, Engelmanniella kahli, Engelmanniella mobilis*
Epiclintidae	*Amphisiella*	*Amphisiella milnei*	1			16	Combined with *Holosticha* (*Amphisiella*) *milnei*	*Amphisiella annulata, Amphisiella arenicola, Amphisiella australis, Amphisiella capitata, Amphisiella illuvialis, Amphisiella kihni, Amphisiella lithophora, Amphisiella marioni, Amphisiella Milnei, Amphisiella n.sp., Amphisiella Oblonga, Amphisiella Oscensis, Amphisiella Ovalis, Amphisiella Strenua, Amphisiella Thiophaga, Amphisiella Turanica*
	Epiclintes	*Epiclintes auricularis*	2			4	Combined with *Oxytricha auricularis*	*Epiclintes ambiguus, Epiclintes feli, Epiclintes pluvialis, Epiclintes radios*
Halteriidae	*Halteria*	*Halteria grandinella*	11	48	3	1	Not found	*Halteria grandinella*
	Halteriidae_X	Halteriidae_X sp.	12			3	Not found	2 Genera: *Halteria* and *Pelagohalteria* (P. Cirrifera, P. hyalina, P. viridis)
	Meseres	*Meseres corlissi*	8			2	Not found	*Meseres cordiformis, Meseres corlissi*

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Holostichidae	*Anteholosticha*	*Anteholosticha gracilis*		1		1	Not found	*Anteholosticha gracilis*
		Anteholosticha manca		1		1	Not found	*Anteholosticha manca*
		Anteholosticha marimonilata		1		1	Not found	*Anteholosticha marimonilata*
		Anteholosticha monilata		1		1	Not found	*Anteholosticha monilata*
		Anteholosticha petzi		1		1	Not found	*Anteholosticha petzi*
		Anteholosticha pulchra		1		1	Not found	*Anteholosticha pulchra*
		Anteholosticha scutellum		1		1	Not found	*Anteholosticha scutellum*
		Anteholosticha sp.		3		8	Not found	*Anteholosticha* cf. *azerbaijanica, Anteholosticha multicirrata, Anteholosticha paramanca, Anteholosticha parawarreni, Anteholosticha pseudomonilata, Anteholosticha* sp. (JJM08052001), *Anteholosticha* sp. (LLQ2007111302), *Anteholosticha* sp. (QDSC04082401)
		Anteholosticha warreni		1		1	Not found	*Anteholosticha warreni*
	Holosticha	*Holosticha bradburyae*		1		1	Not found	*Holosticha bradburyae*
		Holosticha diademata		2		1	Combined with Holosticha (Amphisia) diademata	*Holosticha diademata*
		Holosticha heterofoissneri		1		1	Not found	*Holosticha heterofoissneri*
		Holosticha multistylata		1		1	Not found	*Holosticha multistylata*
		Holosticha polystylata		1		1	*Holosticha polystylata*	*Holosticha polystylata*
		Holosticha sp.		1		7	Genus type for Holostichidae	*Holosticha* cf. *bradburyae, Holosticha* cf. *heterofoissneri, Holosticha* sp., *Holosticha* sp. B05, *Holosticha* sp. HL-2004, *Holosticha* sp. LFL-2004, *Holosticha* sp. WJC-2003
	Metaurostylopsis	*Metaurostylopsis antarctica*		1		1	Not found	*Metaurostylopsis antarctica*
		Metaurostylopsis cheni		4		1	Not found	*Metaurostylopsis cheni*
		Metaurostylopsis salina		2		1	Not found	*Metaurostylopsis salina*
		Metaurostylopsis sinica		1		1	Not found	*Metaurostylopsis sinica*
		Metaurostylopsis sp.		1		1	*Metaurostylopsis marina*	*Metaurostylopsis marina*
		Metaurostylopsis struederkypkeae		4		1	Not found	*Metaurostylopsis struederkypkeae*
	Pseudokeronopsis	*Pseudokeronopsis* sp.		1	15	1	20	*Pseudokeronopsis carnea,Pseudokeronopsis decolor,Pseudokeronopsis flava,Pseudokeronopsis flavicans,Pseudokeronopsis ignea,Pseudokeronopsis multinucleata,Pseudokeronopsis multiplex,Pseudokeronopsis ovalis,Pseudokeronopsis ovata,Pseudokeronopsis pernix,Pseudokeronopsis pulchra,Pseudokeronopsis qingdaoensis,Pseudokeronopsis retrovacuolata,Pseudokeronopsis rubra,Pseudokeronopsis similis,Pseudokeronopsis spectabilis,Pseudokeronopsis trisenestra*	*Pseudokeronopsis carnea, Pseudokeronopsis decolor, Pseudokeronopsis flava, Pseudokeronopsis flavicans, Pseudokeronopsis multinucleata, Pseudokeronopsis ovalis, Pseudokeronopsis ovata, Pseudokeronopsis pararubra, Pseudokeronopsis pernix, Pseudokeronopsis pulchra, Pseudokeronopsis quindaoensis, Pseudokeronopsis retrovacuolata, Pseudokeronopsis rubra, Pseudokeronopsis septibenis, Pseudokeronopsis similis, Pseudokeronopsis trisenestra*

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Hypotrichia_X	*Anteholosticha*	*Anteholosticha paramanca*		1		36	Not found	*Anteholosticha adami, Anteholosticha alpestris, Anteholosticha antecirrata, Anteholosticha arenicola, Anteholosticha australis, Anteholosticha azerbaijanica, Anteholosticha bergeri, Anteholosticha brachysticha, Anteholosticha brevis, Anteholosticha camerounensis, Anteholosticha distyla, Anteholosticha estuarii, Anteholosticha extensa, Anteholosticha gracilis, Anteholosticha grisea, Anteholosticha heterocirrata, Anteholosticha intermedia, Anteholosticha longissima, Anteholosticha manca, Anteholosticha mancoidea, Anteholosticha monilata, Anteholosticha multistilata, Anteholosticha muscicola, Anteholosticha oculata, Anteholosticha plurinucleata, Anteholosticha pulchra, Anteholosticha randani, Anteholosticha scutellum, Anteholosticha sigmoidea, Anteholosticha sphagni, Anteholosticha thononensis, Anteholosticha verrucosa, Anteholosticha violacea, Anteholosticha vuxgracilis, Anteholosticha warreni, Anteholosticha xanthichroma*
	Apobakuella	*Apobakuella fusca*		1		1	Not found	*Apobakuella fusca*
	Bakuella	*Bakuella* sp.		1		2	*Bakuella salinarum, Bakuella marine*	*Bakuella agamalievi*, *Bakuella* sp. 1 WS-2013
	Bergeriella	*Bergeriella ovata*		3	8	1	1	Not found	*Bergeriella ovata*
	Deviata	*Deviata bacilliformis*		1	6	1	2	*Deviata bacilliformis, Deviata abbrevescens*	*Deviata bacilliformis, Deviata Parabaciliformis*

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Hypotrichia_X	*Holosticha*	*Holosticha* sp.		1		148	*Holosticha adami, H. (Amphisiella) annulata, H. (Amphisiella) capitata, H. (Amphisiella) marioni, H. (Amphisiella) milnei, H. (Amphisiella) oblonga, H. (Amphisiella) thiophaga, H. (Keronopsis) alpestris, H. (Keronopsis) coronata, H. (Keronopsis) flavicans, H. (Keronopsis) fontinalis, H. (Keronopsis) globulifera, H. (Keronopsis) gracilis, H. (Keronopsis) monilata, H. (Keronopsis) multinucleata, H. (Keronopsis) multistilata, H. (Keronopsis) muscorum, H. (Keronopsis) ovalis, H. (Keronopsis) ovalis arenivora, H. (Keronopsis) pernix, H. (Keronopsis) pulchra, H. (Keronopsis) rubra, H. (Keronopsis) rubra heptasti-cha, H. (Keronopsis) rubra pentasti-cha, H. (Keronopsis) similis, H. (Keronopsis) spectabilis, H. (Paruroleptus) caudatus, H. (Paruroleptus) Kahl, H. (Paruroleptus) lacteus, H. (Paruroleptus) lepisma, H. (Paruroleptus) magnificus, H. (Paruroleptus) musculus, H. (Paruroleptus) musculus simplex, H. (Paruroleptus) piscis, H. (Parurosoma) dubium, H. (Trichotaxis) crassa, H. (Trichotaxis) fossicola, H. (Trichotaxis) velox, H. algivora, H. alpestris, H. alveolata, H. aquarumdulcium, H. arenicola, H. arenivorus, H. australis, Holosticha begoniensis, H. bergeri, H. binucleata, H. brevis, H. camerounensis, H. caudata, H. contractilis, H. corlissi, H. coronata, H. danubialis, H. decolor, H. diademata, H. discocephalus, H. distyla, H. dragescoi, H. estuarii, H. extensa, H. fasciola, H. flavorubra, H. flavorubra flava, H. flavorubra rubra, H. foissneri, H. fontinalis, H. geleii, H. gibba, H. gracilis, H. grisea, H. holomilnei, H. hymenophora, H. intermedia, H. interrupta, H. islandica, H. kessleri, H. kessleri aquae-dulcis, H. lacazei, H. longiseta, H. macronucleata, H. macrostoma, H. manca, H. manca mononucleata, H. manca plurinucleata, H. mancoidea, H. micans, H. milnei, H. minima, H. monilata, H. mononucleata, H. multicaudicirrus, H. multinucleata, H. multiplex, H. multistilata, H. muscicola, H. mystacea, H. mystacina, H. navicularum, H. obliqua, H. oculata, H. oxytrichoidea, H. Paruroleptus, H. pernix, H. polystylata, H. pseudorubra, H. pullaster, H. punctata, H. randani, H. retrovacuolata, H. rhomboedrica, H. rostrata, H. rostrata mononucleata, H. rubra, H. rubra flava, H. salina, H. scutellum, H. setifera, H. setigera, H. sigmoidea, H. similis, H. simplicis, H. sphagni, H. spindleri, H. stueberi, H. sylvatica, H. tannaensis, H. tenuiformis, H. teredorum, H. tetracirrata, H. thiophaga, H. thononensis, H. Trichotaxis, H. Trichototaxis, H. Trichototaxis stagnatilis, H. velox, H. vernalis, H. vesiculata, H. violacea, H. viridis, H. warreni, H. wetzeli, H. wrzesniowskii, H. wrzesniowskii punctata, H. xanthichroma*	*Holosticha bradburyae, H. diademata, H. heterofoissneri*

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Hypotrichia_X	Hypotrichia_XX	Hypotrichia_XX sp.	110	48	5	1	No Species or taxon type. Is a subclass	*Hypotrichia spissipes*
	Hypotrichida	*Hypotrichida* sp.	2			6	No Species or taxon type. Is a subclass	*Hypotrichida* sp. AL, *Hypotrichida* sp. Florida, *Hypotrichida* sp. KEC2002, *Hypotrichida* sp. LPJ-2005, *Hypotrichida* sp. Misty, *Hypotrichida* sp. OrrK1999, Class name-bearing genus. *Hypotrichia* sp.
	Metaurostylopsis	*Metaurostylopsis flavicana*	1			2	*Metaurostylopsis marina*	*Metaurostylopsis marina, Metaurostylopsis rubra*
	Notohymena	*Notohymena apoaustralis*	1			6	*Notohymena rubescens, Notohymena antarctica, Notohymena australis, Notohymena selvatica*	*Notohymena antarctica, Notohymena australis, Notohymena* sp., *Notohymena rubescens, Notohymena selvatica, Notohymena apoaustralis*
	Onychodromus	*Onychodromus grandis*	1			2	*Onychodromus grandis, Onychodromus grandis simplex*	*Onychodromus grandis*
		Onychodromus quadricornutus	1			1	*Onychodromus quadricornutus*	*Onychodromus quadricornutus*
		Onychodromus sp.				2	*Onychodromus acuminatus, Onychodromus indica*	
	Oxytricha	*Oxytricha* sp.	1			1	See *Oxytricha* sp. Under Oxytrichidae	*Oxytricha fallax*
	Paraurostyla	*Paraurostyla weissei*	4			12	See genus *Paraurostyla* under Family/Superorder Oxytrichidae.	*Paraurostyla brachytona, Paraurostyla Caudata, Paraurostyla Coronata, Paraurostyla Fossicola, Paraurostyla Gracilis, Paraurostyla Granulifera, Paraurostyla Hologama, Paraurostyla Polynucleata, Paraurostyla Pulchra, Paraurostyla Terricola, Paraurostyla Vernalis, Paraurostyla Weissei*
	Perisincirra	*Perisincirra paucicirrata*	1	2	1	11	*Perisincirra buitkampi, Perisincirra filiformis, Perisincirra gellerti, Perisincirra gracilis, Perisincirra interrupta, Perisincirra kahli, Perisincirra pori, Perisincirra similis, Perisincirra viridis*	*Perisincirra jankowski, Perisincirra paucicirrata*
	Plagiotoma	*Plagiotoma lumbrici*	1			12	Not found	*Plagiotoma acuminata, Plagiotoma blattarum, Plagiotoma coli, Plagiotoma concharum, Plagiotoma cordiformis, Plagiotoma difforme, Plagiotoma difformis, Plagiotoma györyana, Plagiotoma kempi, Plagiotoma lateritia, Plagiotoma lumbrici, Plagiotoma pellogaster*
	Ponturostyla	*Ponturostyla enigmatica*	1			1	*Ponturostyla enigmatica*	*Ponturostyla enigmatica*
	Psammomitra	*Psammomitra retractilis*	1			3	*Psammomitra brevicauda, Psammomitra radiosa, Psammomitra retractilis*	*Psammomitra retractilis*
	Pseudokeronopsis	*Pseudokeronopsis* sp.	1			8	See *Pseudokeronopsis* under Holostichidae.	*Pseudokeronopsis carnea, Pseudokeronopsis* cf. *Flava, Pseudokeronopsis Erythrina, Pseudokeronopsis Flava, Pseudokeronopsis Qingdaoensis, Pseudokeronopsis rubra, Pseudokeronopsis sp.* A08*, Pseudokeronopsis* sp. WS-2012

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.		UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Hypotrichia_X	*Pseudourostyla*	*Pseudourostyla franzi*		3	5	1		See *Pseudourostyla* sp. Under Pseudourostylidae.	*Pseudourostyla cristata*
		Pseudourostyla nova		1			1	*Pseudourostyla nova*	*Pseudourostyla nova*
		Pseudourostyla sp.		2			4	*Pseudourostyla cristata, Pseudourostyla levis, Pseudourostyla muscorum, Pseudourostyla urostyla*	*Pseudourostyla cristatoides, Pseudourostyla franzi, Pseudourostyla* sp. CXM08112801
	Rubrioxytricha	*Rubrioxytricha haematoplasma*		1			3	*Rubrioxytricha ferruginea, Rubrioxytricha haematoplasma*	*Rubrioxytricha haematoplasma, Rubrioxytricha ferruginea, Rubrioxytricha indica*
	Stichotrichia	*Stichotrichia* sp.		4	13	2		Not found	*Stichotrichia* incertae sedis?
	Urostylida	*Urostylida* sp.		1				Rank Family, genus name *Urostyla* (e.g. *Urostyla cristata*), See *Urostyla* under Urostylidae.	*Urostylida* incertae sedis?
Metaurostylopsis	Metaurostylopsis	*Metaurostylopsis* sp.		1			6	*Metaurostylopsis marina*	*Metaurostylopsis marina, Metaurostylopsis antarctica, Metaurostylopsis cheni, Metaurostylopsis salina, Metaurostylopsis sinica. Metaurostylopsis struederkypkeae*
Oxytrichidae	Amphisiella	*Amphisiella candida*		1			25	*Amphisiella acuta, A. annulata, A. arenicola, A. australis, A. binucleata, A. capitata, A. dorsicirrata, A. edaphoni, A. faurei, A. gellerti, A. kihni, A. lithophora, A. marioni, A. milnei, A. oblonga, A. oscensis, A. ovalis, A. perisincirra, A. polycirrata, A. quadrinucleata, A. raptans, A. terricola, A. thiophaga, A. thiophaga, A. vitiphila*	*Amphisiella candida*
		Amphisiella magnigranulosa		1	20	1	2	*Amphisiella magnigranulosa*	*Amphisiella magnigranulosa, Amphisiella pulchra*
	Apogastrostyla	*Apogastrostyla rigescens*		1			1	Not found	*Apogastrostyla rigescens*
	Bistichella	*Bistichella* FG-2014		1	225	1	1	Not found	*Bistichella* FG-2014
		Bistichella variabilis		1	8	1	1	Not found	*Bistichella variabilis*
	Cyrtohymena	*Cyrtohymena citrina*		4	24	1	1	*Cyrtohymena citrina*	*Cyrtohymena citrina*
		Cyrtohymena sp.		1			25	*Cyrtohymena aestuarii, C. australis, C. balladynula, C. candens, C. candens depressa, C. dubia, C. fenestrata, C. ferruginea, C. gracilis, C. granulata, C. inquieta, C. macrostoma, C. marina, C. muscorum, C. ovalis, C. primicirrata, C. quadrinucleata, C. sapropelica, C. simplex, C. sphagnicola, C. tetracirrata, C. torrenticola*	*Cyrtohymena aestuarii, Cyrtohymena aspoecki, Cyrtohymena australis, Cyrtohymena candens, Cyrtohymena fenestrata, Cyrtohymena ferruginea, Cyrtohymena fusiformis, Cyrtohymena gracilis, Cyrtohymena granulata, Cyrtohymena inquieta, Cyrtohymena marina, Cyrtohymena muscorum, Cyrtohymena ovalis, Cyrtohymena primicirrata, Cyrtohymena quadrinucleata, Cyrtohymena sapropelica, Cyrtohymena tetracirrata, Cyrtohymena torrenticola*

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Oxytrichidae	*Deviata*	*Deviata parabacilliformis*	1			8	*Deviata abbrevescens, Deviata bacilliformis*	*Deviata abbrevescens, Deviata parabacilliformis, Deviata brasiliensis, Deviata estevesi, Deviata polycirrata, Deviata quadrinucleata, Deviata rositae, Deviata spirostoma*
	Gastrostyla	*Gastrostyla* sp.	1	4	1	2	*Gastrostyla affine, Gastrostyla dorsicirrata, Gastrostyla dubia, Gastrostyla hebbalica, Gastrostyla minima, Gastrostyla muscorum, Gastrostyla mystacea, Gastrostyla opisthoclada, Gastrostyla parasteinii, Gastrostyla philippinensis, Gastrostyla pulchra, Gastrostyla setifera, Gastrostyla stenocephala, Gastrostyla sterkii, Gastrostyla vorax*	*Gastrostyla* sp. 1 MD-2012, *Gastrostyla* sp. 2 MD-2012
		Gastrostyla steinii	12	2	1	1	*Gastrostyla steinii*	*Gastrostyla steinii*
	Gonostomum	*Gonostomum namibiense*	1	29	1	1	Not found	*Gonostomum namibiense*
		Gonostomum sp.	2	24	2	4	*Gonostomum affine, Gonostomum algicola, Gonostomum ambiguum, Gonostomum andoi, Gonostomum bryonicolum, Gonostomum ciliophorum, Gonostomum franzi, Gonostomum geleii, Gonostomum gonostomoida, Gonostomum kuehnelti, Gonostomum mereschkowskii, Gonostomum parvum, Gonostomum pediculiforme, Gonostomum spirotrichoides*	*Gonostomum* sp. 1 TS-2013, *Gonostomum* sp. HL-2004, *Gonostomum* sp. JS-2012, *Gonostomum* sp. MD-2012,
		Gonostomum strenuum	3	37	2	1	*Gonostomum strenuum*	*Gonostomum strenuum*
	Hemigastrostyla	*Hemigastrostyla enigmatica*	2			2	*Hemigastrostyla enigmatica, Hemigastrostyla stenocephala*	*Hemigastrostyla enigmatica*
	Hemiurosoma	*Hemiurosoma terricola*	1			1	*Hemiurosoma terricola*	*Hemiurosoma terricola*
	Histriculus	*Histriculus histrio*	1			17	*Histriculus acuminatus, Histriculus admirabilis, Histriculus cavicola, Histriculus complanatus, Histriculus erethesticus, Histriculus erethisticus, Histriculus histrio, Histriculus hyalinus, Histriculus lemani, Histriculus minimus, Histriculus muscorum, Histriculus polycirratus, Histriculus similis, Histriculus similis tricirratus, Histriculus sphagni, Histriculus tricirratus, Histriculus vorax*	*Histriculus histrio, Histriculus concolor*
	Kahliella	*Kahliella* sp.	1	29	1	11	*Kahliella acrobates, Kahliella bacilliformis, Kahliella costata, Kahliella franzi, Kahliella marina, Kahliella microstoma, Kahliella multiseta, Kahliella simplex, Kahliella spirostoma, Kahliella zignis*	*Kahliella acrobates, Kahliella costata, Kahliella microstoma, Kahliella multiseta, Kahliella quadrinucleata, Kahliella simplex, Kahliella zignis*
	Laurentiella	*Laurentiella strenua*	3	4	1	5	*Laurentiella acuminata, Laurentiella macrostoma, Laurentiella monilata, Laurentiella strenua*	*Laurentiella bergeri, Laurentiella strenua*

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Oxytrichidae	*Neokeronopsis*	*Neokeronopsis aurea*	1			2	Not found	Neokeronopsis asiatica, Neokeronopsis aurea
	Onychodromopsis	*Onychodromopsis flexilis*	2			5	*Onychodromopsis flexilis, O. kahli, O. tihanyiensis, O. variabilis, O. viridis*	*Onychodromopsis flexilis*
	Orthamphisiella	*Orthamphisiella breviseries*	2	9	1	1	Not found	*Orthamphisiella breviseries*
	Orthoamphisiella	*Orthoamphisiella* sp.	1			4	*Orthoamphisiella franzi, Orthoamphisiella grelli, Orthoamphisiella stramenticola*	*Orthoamphisiella breviseries, Orthoamphisiella grelli, Orthoamphisiella stramenticola*
	Oxytricha	*Oxytricha elegans*	1			1	*Oxytricha elegans*	*Oxytricha elegans*
		Oxytricha ferruginea	1			1	*Oxytricha ferruginea*	*Oxytricha ferruginea*
		Oxytricha granulifera	11			2	*Oxytricha granulifera granulifera, Oxytricha granulifera*	*Oxytricha granulifera granulifera*
		Oxytricha lanceolata	1	26	1	1	*Oxytricha lanceolata*	*Oxytricha lanceolata*
		Oxytricha longa	2	6	1	1	*Oxytricha longa*	*Oxytricha longa*
		Oxytricha longigranulosa	2			1	*Oxytricha longigranulosa*	*Oxytricha longigranulosa*
		Oxytricha ottowi	2			2	*Oxytricha ottowi*	*Oxytricha ottowi* 1 MD-2012, *O. ottowi* 2 MD-2012
		Oxytricha saltans	1			1	*Oxytricha saltans,*	*Oxytricha saltans*
		Oxytricha sp.	101	83	6	282	*O. felis, O. gallina, O. gibbosa, O. joblotii, O. lepus, O. pelionella, O. pellionella, O. pulex, O. pullaster, O. pullicina, O. transfuga, O. variabilis, O. volutator, O. (Actinotricha) saltans, O. acuminata, O. aeruginosa, O. affinis, O. africana, O. agilis, O. alba, O. alfredi, O. alfredkahli, O. ambigua, O. anca, O. arcuata, O. auricularis, O. auripunctata, O. australis, O. balladyna, O. barbula, O. becciformis, O. bicirrata, O. bifaria, O. bilobata, O. bimembranata, O. bivacuolata, O. buitkampi, O. bulla, O. capitata, O. caudata, O. cavicola, O. chilensis, O. chlorelligera, O. cicada, O. cimex, O. cornipes, O. cornuta, O. crassa, O. crassistilata, O. cultriformis, O. curta, O. cypris, O. decumana, O. deformis, O. discifera, O. dragescoi, O. dubia, O. dujardiniana, O. durhamiensis, O. ehrenbergiana, O. elliptica, O. elongata, O. enigmatica, O. ephippioides, O. euglenivora, O. eurystoma, O. exociformis, O. fallax, O. faurei, O. felis, O. fimbriata, O. flava, O. flava carnea, O. formosa, O. fromenteli, O. furcatus, O. fusca, O. galeata, O. gallina, O. geleii, O. germanica, O. gibba, O. gigantea, O. granulifera quadricirrata, O. granulosa, O. gyrinioides, O. haematoplasma, O. halophila, O. henegui, O. hengshanensis, O. (Histrio) acuminatus, O. (Histrio) complanatus, O. (Histrio) erethisticus, O. (Histrio) histrio, O. (Histrio) muscorum, O. (Histrio) similis, O. (Histrio) sphagni, O. (Histrio) vorax, O. histrioides, O. histriomuscorum, O. hymenostoma, O. immemorata, O. incrassata, O. inquieta, O. islandica, O. joblotii, O. kahlovata, O. kessleri, O. labiata, O. lacerata, O. lacrimula, O. lamella, O. lata, O. lepus, O. leucoa, O. lingua, O. longi-caudata, O. longicirrata, O. longipes, O. longissima, O. ludibunda, O. lundi, O. luteolucens, O. macrostyla, O. magna, O. marina, O. matritensis, O. merula, O. micans, O. minima, O. minor, O. monspessulana, O. monstrosa, O. multipes, O. multiseta, O. muscorum, O. musculus, O. mystacea, O. nauplia, O. nova, O. oblonga, O. oblongatus, O. obtusa, O. oculata, O. oltenica, O. opisthomuscorum, O. (Opisthotricha) crassistilata, O. (Opisthotricha) elongata, O. (Opisthotricha) euglenivora, O. (Opisthotricha) faurei, O. (Opisthotricha) halophila, O. (Opisthotricha) muscorum, O. (Opisthotricha) ovata, O. (Opisthotricha) parallela, O. (Opisthotricha) parallela minor, O. (Opisthotricha) procera, O. (Opisthotricha) similis, O. (Opisthotricha) sphagni, O. ovalis, O. ovata, O. oxymarina, O. parahalophila, O. parallela, O. parvistyla, O. parvula, O. pellionella, O. pellionella chilensis, O. pernix, O. phytophaga, O. piscis, O. pisciunculiformis, O. pistilloides, O. plana, O. planctontica, O. platystoma, O. pleuronectes, O. plicata, O. praeceps, O. proboscis, O. procera, O. protensa, O. proximata, O. pseudofurcata, O. pseudofusiformis, O. pseudosimilis, O. pubes, O. pullicina, O. pustulata, O. quadricirrata, O. quadrinucleata, O. quercineti, O. radians, O. retractilis, O. rostrata, O. rubra, O. rubripuncta, O. salmastra, O. saprobia, O. scutellum, O. scutum, O. selvatica, O. setigera, O. shii, O. similis, O. siseris, O. sordida, O. sordis, O. sphagni, O. (Steinia) balladynula, O. (Steinia) candens, O. (Steinia) candens aestuarii, O. (Steinia) ferruginea, O. (Steinia) gracilis, O.(Steinia) granulata, O. (Steinia) inquieta, O. (Steinia) marina, O. (Steinia) muscorum, O. (Steinia) platystoma, O. (Steinia) sapropelica, O. stenocephala, O. stratiformis, O. strenua, O. striata, O. stylomuscorum, O. (Stylonychia) claviformis, O. (Stylonychia) curvata, O. (Stylonychia) fissiseta, O. (Stylonychia) grandis, O. (Stylonychia) macrostyla, O. (Stylonychia) muscorum, O. (Stylonychia) mytilus, O. (Stylonychia) notophora, O. (Stylonychia) pusilla, O. (Stylonychia) pustulata, O. (Stylonychia) putrina, O. (Stylonychia) vorax, O. subcylindrica, O. (Tachysoma) echinata, O. (Tachysoma) furcata, O. (Tachysoma) mirabilis, O. (Tachysoma) parvistyla, O. (Tachysoma) pellionella, O. (Tachysoma) rigescens, O. (Tachysoma) saltans, O. tenella, O. tennella, O. terrestris, O. tetracirrata, O. transfuga, O. tretranucleata, O. tricirrata, O. tricornis, O. truncata, O. truncata dilatata, O. truncata piriforme, O. tubicola, O. (Urosoma) acuminata, O. (Urosoma) caudata, O. (Urosoma) cienkowskii, O. (Urosoma) emarginata, O. (Urosoma) gigantea, O. (Urosoma) longicirrata, O. (Urosoma) macrostyla, O. (Urosoma) planctonticum, O. (Urosoma) stenocephala, O. urostyla, O. variabilis, O. velox, O. viridis, O. volutator, O. wrzesniowskii*	*O.* sp. ZL-2014, *Oxytricha* sp., *Oxytricha* sp. 1 MD-2012, *Oxytricha* sp. 2 MD-2012, *Oxytricha* sp. 3 MD-2012, *Oxytricha* sp. Aspen, *Oxytricha* sp. JS-2012, *Oxytricha* sp. LPJ-2005, *Oxytricha* sp. Misty, *Oxytricha* sp. Oc_L1, *Oxytricha* sp. Ox_L1, *Oxytricha* sp. Steamboat Hot Springs

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Oxytrichidae	*Oxytricha*	*Oxytricha trifallax*	6	9	1	1	*Oxytricha trifallax*	*Oxytricha trifallax*
	Oxytrichidae	*Oxytrichidae* sp.	1			106	Family taxonomic group with the name-bearing genus *Oxtricha*. See *Oxytricha* sp.	*Oxtricha acuminatus, Oxtricha aeruginosa, Oxtricha aestuarii, Oxtricha africana, Oxtricha alfredi, Oxtricha alfredkahli, Oxtricha ambigua, Oxtricha arabica, Oxtricha auripunctata, Oxtricha balladyna, Oxtricha balladynula, Oxtricha bimembranata, Oxtricha bulla, Oxtricha chlorelligera, Oxtricha claviformis, Oxtricha complanatus, Oxtricha crassistilata, Oxtricha cultriformis, Oxtricha curvata, Oxtricha discifera, Oxtricha durhamiensis, Oxtricha echinata, Oxtricha elongata, Oxtricha enigmatica, Oxtricha enigmatica, Oxtricha erethisticus, Oxtricha euglenivora, Oxtricha fallax, Oxtricha fastigata, Oxtricha faurei, Oxtricha felis, Oxtricha fennica, Oxtricha fissiseta, Oxtricha fromenteli, Oxtricha fusiformis, Oxtricha geleii, Oxtricha gibba, Oxtricha grandis, Oxtricha granulosa, Oxtricha halophila, Oxtricha histrioides, Oxtricha hymenostoma, Oxtricha islandica, Oxtricha kahlovata, Oxtricha lamella, Oxtricha leucoa, Oxtricha longicirrata, Oxtricha longissima, Oxtricha ludibunda, Oxtricha maeceps, Oxtricha marcili, Oxtricha maskelli, Oxtricha matritensis, Oxtricha micans, Oxtricha minor, Oxtricha mirabilis, Oxtricha mistacea, Oxtricha multiseta, Oxtricha nauplia, Oxtricha notophora, Oxtricha nuptacina, Oxtricha oblongua, Oxtricha opisthomuscorum, Oxtricha ovalis, Oxtricha oxymarina, Oxtricha parahalophila, Oxtricha parallela, Oxtricha parvistyla, Oxtricha piscis, Oxtricha praeceps, Oxtricha procera, Oxtricha proteusa, Oxtricha proximata, Oxtricha pseudofurcata, Oxtricha pseudofusiformis, Oxtricha pseudosimilis, Oxtricha pubes, Oxtricha pullicina, Oxtricha pusilla, Oxtricha putrina, Oxtricha quadricirrata, Oxtricha rigescens, Oxtricha rostrata, Oxtricha rubra, Oxtricha rubripuncta, Oxtricha saprobia, Oxtricha saprobica, Oxtricha scutum, Oxtricha setigera, Oxtricha shii, Oxtricha similis, Oxtricha similis, Oxtricha siseris, Oxtricha sordida, Oxtricha sp., Oxtricha sphagni, Oxtricha tenella, Oxtricha tennella, Oxtricha terrestris, Oxtricha torrenticola, Oxtricha transfuga, Oxtricha tricirrata, Oxtricha tricornis, Oxtricha truncata, Oxtricha variabilis, Oxtricha vorax*

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
	Oxytrichidae_X	Oxytrichidae_X sp.	86	235	12		Family taxonomic group with the name-bearing genus Oxtricha. See *Oxytricha* sp.	Contains taxonomic groups (Cyrtohymena, Hemigastrostyla, Oxytricha, Ponturostyla, Tachysoma)
	Parabistichella	*Parabistichella variabilis*	1			1	Not found	*Parabistichella variabilis*
	Paraparentocirrus	*Paraparentocirrus sibillinensis*	1	4	1	1	Not found	*Paraparentocirrus sibillinensis*
	Parasterkiella	*Parasterkiella thompsoni*	1			1	Not found	*Parasterkiella thompsoni*
	Paraurostyla	*Paraurostyla viridis*	1			20	*Paraurostyla viridis, P. buitkampi, P. dispar, P. enigmatica, P. fossicola, P. gibba, P. granulifera, P. hologama, P. hymenophora, P. latissima, P. macrostoma,P. marina, P. aumanni, P . polymicronucleata, P. polynucleata, P. pulchra, P. raikovi, P. rubra, P. terricola, P. viridis, P. weissei*	*Paraurostyla weissei, Paraurostyla* sp., *Paraurostyla viridis*
	Pattersoniella	*Pattersoniella vitiphila*	2			1	*Pattersoniella vitiphila*	*Pattersoniella vitiphila*
	Pleurotricha	*Pleurotricha lanceolata*	2			20	*Pleurotricha (Allotricha) lanceolata, P. dubium, P. echinata, P. flexilis, P. grandis, P. indica, P. lanceolata, P. macrostoma, P. mononucleata, P. planensis, P. grandis, P. setifera, P. tchadensis, P. tihanyensis, P. tihanyiensis, P. variabilis*	*Pleurotricha lanceolata, Pleurotricha curdsi, Pleurotricha flexilis, Pleurotricha grandis, Pleurotricha monilata, Pleurotricha multinucleata*
		Pleurotricha lanceolota	1				Not found	No results
	Protogastrostyla	*Protogastrostyla pulchra*	5			2	Not found	*Protogastrostyla pulchra, Protogastrostyla sterkii*
	Rigidohymena	*Rigidohymena candens*	1			1	Not found	*Rigidohymena candens*
	Schmidingerothrix	*Schmidingerothrix* sp.	2			1	Not found	*Schmidingerothrix* sp. 1 TS-2013
	Steinia	*Steinia sphagnicola*	2			28	*Steinia balladynula, S. balladynula, S. bivacuolata, S. candens, S. candens aestuarii, S. candens depressa, S. candens, S. citrina, S. dubia, S. fenestrata, S. ferruginea, S. gracilis, S. granulata, S. inquieta, S. macrostoma, S. marina, S. muscorum, S. ovalis, S. platystoma, S. primicirrata, S. quadrinucleata, S. sapropelica, S. sapropelica, S. simplex, S. sphagnicola, S. tetracirrata, S. torrenticola, S. ultricirrata*	*Steinia sphagnicola, Steinia bivacuolata, Steinia platystoma*

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
	Sterkiella	*Sterkiella cavicola*		1		12	*Sterkiella cavicola*	*Sterkiella cavicola*
		Sterkiella histriomuscorum		5		1	*Sterkiella histriomuscorum*	*Sterkiella histriomuscorum*
		Sterkiella nova		2		1	*Sterkiella nova*	*Sterkiella nova*
		Sterkiella sp.		12	28	2	16	*Sterkiella admirabilis, Sterkiella quadrinucleatus, Sterkiella similis tricirrata, Sterkiella terricola, Sterkiella thompsoni, Sterkiella tricirrata*	*Sterkiella* sp., *Sterkiella* sp. 1201-1, *Sterkiella* sp. Aspen, *Sterkiella* sp. CH55, *Sterkiella* sp. JS-2012a, *Sterkiella* sp. JS-2012b, *Sterkiella* sp. JS-2012c, *Sterkiella* sp. JS-2012d, *Sterkiella* sp. KK-2011, *Sterkiella* sp. SK-2013
	Stichotrichia	*Stichotrichia* sp.		1	##	1	6	Not found. See Hypotrichia_X.	*Stichotrichia* sp. K3-0803-6, *Stichotrichia* sp. St04, *Stichotrichia* sp. St05, *Stichotrichia* sp. St09, *Stichotrichia* sp. St13, *Stichotrichia* sp. St20
	Stylonychia	*Stylonychia ammermanni*		2		1	Not found	*Stylonychia ammermanni*
		Stylonychia bifaria		1		1	*Stylonychia bifaria*	*Stylonychia bifaria*
		Stylonychia lemnae		23	8	1	1	*Stylonychia lemnae*
		Stylonychia mytilus		18		1	*Stylonychia mytilus*	*Stylonychia mytilus*
		Stylonychia notophora		1		1	*Stylonychia notophora*	*Stylonychia notophora*
		Stylonychia sp.		3		64	*Stylonychia appendiculata, S. calva, S. cimex, S. clavata, S. claviformis, S. curvata, S. echinata, S. fissiseta, S. grandis, S. harbinensis, S. histrio, S. kahli, S. lanceolata, S. macrostyla, S. makrostyla, S. monostylus, S. muscorum, S. mytilus pusilla, S. nodulinucleata, S. ovalis, S. parallela, S. patula, S. pseudograndis, S. pulchra, S. pusilla, S. pustulata, S. putrina, S. quadrinucleata, S. regularis, S. silurus, S. similis, S. sphaerica, S. steinii, S. stylomuscorum, S. tricornis, S. virgula, S. vorax*	*Stylonychia anfracta, S. appendiculatus, S. auriformis, S. bicaudatus, S. clavata, S. claviformis, S. complanatus, S. cornifrons, S. curvata, S. dupla, S. ellpisoides, S. excavata, S. fastigata, S. frontealbens, S. frontosus, S. globifrons, S. grandis, S. harbinensis, S. indica, S. labiata, S. limbiformis, S. massula, S. membranaceus, S. microstoma, S. minimus, S. mutabunda, S. nodulinucleata, S. octonistylus, S. parallela, S. patula, S. pseudograndis, S. pusilla, S. pustulata, S. putrina, S. reclinis, S. rostrata, S. sphagni, S. stylomuscorum, S. tricornis, S. trochiformis, S. viridis, S. vorax*
	Tetmemena	*Tetmemena pustulata*		3		1	*Tetmemena bifaria, Tetmemena pustulata, Tetmemena vorax*	*Tetmemena pustulata*
	Urospinula	*Urospinula succisa*		1		5	*Urospinula bicaudata, Urospinula calciba, Urospinula simplex, Urospinula sinistrocaudata*	*Urospinula bicaudata, Urospinula simplex, Urospinula succisa*
Parabirojimidae	*Parabirojimia*	*Parabirojimia multinucleata*		1		1	Not found	*Parabirojimia multinucleata*
		Parabirojimia similis		1		1	Not found	*Parabirojimia similis*
Pseudokeronopsidae	*Antiokeronopsis*	*Antiokeronopsis flava*		1		1	Not found	*Antiokeronopsis flava*
	Apokeronopsis	*Apokeronopsis bergeri*		2		2	Not found	*Apokeronopsis bergeri, Apokeronopsis antarctica*
		Apokeronopsis crassa		2		1	Not found	*Apokeronopsis crassa*
		Apokeronopsis ovalis		2		1	Not found	*Apokeronopsis ovalis*
		Apokeronopsis sinica		1		1	Not found	*Apokeronopsis sinica*
		Apokeronopsis wrighti		1		1	Not found	*Apokeronopsis wrighti*
	Pseudokeronopsis	*Pseudokeronopsis carnea*		2		1	*Pseudokeronopsis carnea*	*Pseudokeronopsis carnea*
		Pseudokeronopsis erythrina		1		1	Not found	*Pseudokeronopsis erythrina*
		Pseudokeronopsis flava		3		2	*Pseudokeronopsis flava*	*Pseudokeronopsis flava, Pseudokeronopsis* cf. *flava*
		Pseudokeronopsis rubra		4		1	*Pseudokeronopsis rubra*	*Pseudokeronopsis rubra*
		Pseudokeronopsis sp.		1		2	See *Pseudokeronopsis* under Holostichidae.	*Pseudokeronopsis flavicans, Pseudokeronopsis ovalis, Pseudokeronopsis rubra, Pseudokeronopsis qingdaoensis, Pseudokeronopsis* sp. A08, *Pseudokeronopsis* sp. WS-2012
	Thigmokeronopsis	*Thigmokeronopsis stoecki*		1		5	*Thigmokeronopsis antarctica, Thigmokeronopsis crystallis, Thigmokeronopsis jahodai*	*Thigmokeronopsis stoecki, Thigmokeronopsis rubra*

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Pseudourostylidae	*Anteholosticha*	*Anteholosticha* sp.	1			16	Not found	*Anteholosticha* cf. *azerbaijanica, Anteholosticha gracilis, Anteholosticha manca, Anteholosticha marimonilata, Anteholosticha monilata, Anteholosticha multicirrata, Anteholosticha paramanca, Anteholosticha parawarreni, Anteholosticha petzi, Anteholosticha pseudomonilata, Anteholosticha pulchra, Anteholosticha scutellum, Anteholosticha* sp. JJM08052001, *Anteholosticha* sp. LLQ2007111302, *Anteholosticha* sp. QDSC04082401, *Anteholosticha warreni*
	Heterokeronopsis	*Heterokeronopsis pulchra*	1			1	Not found	*Heterokeronopsis pulchra*
	Metaurostylopsis	*Metaurostylopsis sinica*	1			6	*Metaurostylopsis marina*	*Metaurostylopsis sinica, Metaurostylopsis antarctica, Metaurostylopsis cheni, Metaurostylopsis salina, Metaurostylopsis struederkypkeae*
	Nothoholosticha	*Nothoholosticha fasciola*	1			1	Not found	*Nothoholosticha fasciola*
	Pseudokeronopsidae	*Pseudokeronopsidae* sp.	1			1	Family taxonomic group with the name-bearing genus *Pseudokeronopsis*. See *Pseudokeronopsis* sp.	*Pseudokeronopsidae* sp. 1 BNB-2013
	Pseudokeronopsis	*Pseudokeronopsis* sp.	1				See *Pseudokeronopsis* under Holostichidae.	See *Pseudokeronopsis* under *Holostichidae*
	Pseudourostyla	*Pseudourostyla franzi*	1	23	1	8	*Pseudourostyla cristata, Pseudourostyla franzi, Pseudourostyla levis, Pseudourostyla muscorum, Pseudourostyla nova, Pseudourostyla urostyla*	*Pseudourostyla franzi, Pseudourostyla Cristata, Pseudourostyla Cristatoides, Pseudourostyla Nova, Pseudourostyla* sp. CXM08112801
Rigidotrichidae	*Rigidothrix*	*Rigidothrix goisery*	1			1	Not found	*Rigidothrix goisery*
Spirofilidae	*Pseudouroleptus*	*Pseudouroleptus caudatus*	1			1	*Pseudouroleptus caudatus*	*Pseudouroleptus caudatus* (*Pseudouroleptus caudatus caudatus*)
		Pseudouroleptus jejuensis	1			1	Not found	*P. jejuensis*
		Pseudouroleptus sp.	1	25	1	5	*Pseudouroleptus buitkampi, Pseudouroleptus humicola, Pseudouroleptus procerus, Pseudouroleptus terrestris*	*Pseudouroleptus* sp. ALT-2014
	Strongylidium	*Strongylidium orientale*	1			25	*Strongylidium arenicola, S. arenicolus, S. bacilliforme, S. californicum, S. contortum, S. contortus, S. coronatum, S. crassum, S. crepidatum, S. deflectum, S. granuliferum, S. labiatum, S. lanceolatum, S. maritimum, S. microstoma, S. mucicola, S. muscorum, S. packii, S. polystichum, S. suis, S. (Urostrongylum) caudatum, S. (Urostrongylum) contorta, S. wilberti*	*Strongylidium orientale*
		Strongylidium pseudocrassum	1			1	*Strongylidium pseudocrassum*	*Strongylidium pseudocrassum*
Trachelostylidae	*Spirotrachelostyla*	*Spirotrachelostyla tani*	1			3	Not found	*Spirotrachelostyla tani, Spirotrachelostyla simplex, Spirotrachelostyla spiralis*
	Trachelostyla	*Trachelostyla pediculiformis*	1			14	*Trachelostyla affine, T. bryoniculum, T. canadensis, T. caudata, T. ciliophorum, T. dubia, T. geleii, T. macrostoma, T. pediculiformis, T. rostrata, T. simplex, T. spiralis, T. spirotrichoides*	*Trachelostyla pediculiformis, Trachelostyla bryoniculum, Trachelostyla caudata, Trachelostyla elongata, Trachelostyla rostrata*

Table S3 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Urostylidae	*Diaxonella*	*Diaxonella pseudorubra*		1		1	Not found	*Diaxonella pseudorubra*
		Diaxonella trimarginata		3		1	Diaxonella trimarginata	*Diaxonella trimarginata*
	Hemicycliostyla	*Hemicycliostyla sphagni*		1		4	*Hemicycliostyla lacustris, H. marina, H. sphagni, H. trichota*	*Hemicycliostyla sphagni, Hemicycliostyla Lacustris, Hemicycliostyla Marina, Hemicycliostyla Trichota*
	Paruroleptus	*Paruroleptus lepisma*		2		19	*Paruroleptus caudatus, P. dubium, P. gallina, P. gibbosus, P. interrupta, P. lacteus, P. lepisma, P. magnificus, P. muscorum, P. musculus, P. musculus minor, P. musculus simplex, P. notabilis, P. novitas, P. ophryoglena, P. pectinatus, P. piscis, P. strenuus, P. viridis*	*Paruroleptus lepisma, Paruroleptus gibbosus, Paruroleptus lacteus, magnificus, Paruroleptus musculus, Paruroleptus novitas, Paruroleptus ophryoglena, Paruroleptus pectinatus, Paruroleptus viridis*
	Tunicothrix	*Tunicothrix* sp.		1		1	Not found	*Tunicothrix rostrata*
		Tunicothrix wilberti		1		1	Not found	*Tunicothrix wilberti*
	Uroleptopsis	*Uroleptopsis citrina*		3		7	*Uroleptopsis citrina, U. ignea, U. kahli, U. multiseta, U. ovata, U. roscoviana, U. viridis*	*Uroleptus citrina*
	Uroleptus	*Uroleptus gallina*		2	11	1	1 Not found	*Uroleptus gallina*
		Uroleptus pisces		3		1	Not found	*Uroleptus pisces*
		Uroleptus sp.		6	7	1	4 Not found	*Uroleptus* sp. INHC148, *Uroleptus* sp. MD-2012, *Uroleptus* sp. Willii, *Uroleptus* sp. WJC-2003
	Urostyla	*Urostyla grandis*		5	27	2	42 *Urostyla agamalievi, U. algivora, U. brachytona, U. caudata, U. chlorelligera, U. coei, U. concha, U. cristata, U. dispar, U. elongata, U. flavicans, U. franzi, U. fulva, U. gigas, U. gracilis, U. gracilis pallida, U. gracilis sanguinea, U. grandis, U. grandis kahli, U. granids typica, U. hologama, U. intermedia, U. latissima, U. limboonkengi, U. lynchi, U. marina, U. multipes, U. muscorum, U. naumanni, U. paragrandis, U. polymicronucleata, U. pseudomuscorum, U. rubra, U. sphagni, U. thompsoni, U. trichogaster, U. trichogaster elongata, U. trichogaster fulva, U. trichogaster multipes, U. vernalis, U. viridis, U. weissei*	*Urostyla grandis, Urostyla muscorum, Urostyla viridis, Urostyla weissei*

Table S4. Molecular protist ribosomal reference (PR^2) v203 database curated sequence inventory for euplotian ciliates (Alveolata) compared to next-generation sequencing (NGS) data discovered in grassland soil and morphologically described genusses and species within the order Euplotia. Lineage breakdown (Oder, Family/Superorder, Genus and Species) follows the revised classification of Eukaryotes by Adl et al. (2012), where unknown or unresolved family and superorder clade lineages are indicated ending with "_X". A breakdown of NGS data indicates the number of unique individual reads (UIRs) and Blast based operational taxonomic units (OTUs) - UIRs grouped to the same accession number was counted as an OTU. Empty blocks indicate that OTUs or UIRs were not detected or respective morphologically described species were not listed. Morphologically described species are given for by Berger (2001)* and the online encyclopaedia of life (EOL). #Database Acc. Num. - Number of database accountable accession numbers; #Species described - Number of morphologically described genusses and species catalogued by Berger (2001) and EOL. *Berger H (2001) Catalogue of ciliate names 1. Hypotrichs. Salzburg (Austria): Verlag Helmut Berger. i-viii and 206 pp. **Number of Morphospecies described and catalogued in the Encyclopaedia of Life (EOL - www.eol.org).

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Aspidiscidae	*Aspidisca*	*Aspidisca aculeata*	1			1	*Aspidisca aculeata*	*A. aculeata*
		Aspidisca fusca	1			1	*Aspidisca fusca*	*A. fusca*
		Aspidisca hexeris	2			1	*Aspidisca hexeris*	*A. hexeris*
		Aspidisca leptaspis	1			1	*Aspidisca leptaspis*	*A. leptaspis*
		Aspidisca magna	1			1	*Aspidisca magna*	*A. magna*
		Aspidisca orthopogon	1			1	*Aspidisca orthopogon*	*A. orthopogon*
		Aspidisca sp.	22	15	2	70	*Aspidisca acerosa, A. andreewi, A. angulata, A. antarctica, A. caspica, A. lynceus, A. psammobiotica, A. bellua, A. bengalensis, A. bicornis, A. binucleata, A. bipartita, A. caspica, A. caudata, A. cicada, A. costata, A. costata tetracirrata, A. crenata, A. dentata, A. denticulata, A. (Dimaspidisca) binucleata, A. (Dimaspidisca) fusca, A. (Dimaspidisca) mutans, A. eocenica, A. fjeldi, A. fuscoides, A. glabra, A. herbicola, A. heterotrichus, A. hyalina, A. irinae, A. jugensis, A. longipes, A. lyncaster, A. lynceus, A. major faurei, A. marsupialis, A. maxima, A. minuta, A. mutans, A. nana, A. (Netaspidisca) acerosa, A. (Onychaspis) polystyla, A. pelvis, A. pertinens, A. plana, A. poljanskyi, A. polypoda, A. polystyla maxima, A. psammobiotica, A. pulcherrima, A. pulcherrima baltica, A. pulvinata, A. putrina, A. quadrilineata, A. radiata, A. robusta, A. sedigita, A. steini, A. steini major, A. sulcata, A. terranovae, A. tridentata, A. tuberosa, A. turrita, A. turrita echinata, A. turrita tricosta, A. zonata*	*Aspidisca acerosa, A. andreewi, A. angulata, A. antarctica, A. baltica, A. ellua, A. bengalensis, A. bicornis, A. binucleata, A. caudata, A. cicada, A. crenata, A. dentata, A. eocenica, A. fuscoides, A. glabra, A. herbicola, A. heterotrichu, A. hyalina, A. irinae, A. jugensis, A. longipes, A. lyncaster, A. lynceus, A. major, A. marsupialis, A. maxima, A. minuta, A. mutans, A. nana, A. pelvis, A. pertinens, A. plana, A. polypoda, A. polystyla, A. psammobiotica, A. pulcherrima, A. pulvinata, A. quadrilineata, A. radiata, A. robusta, A. sedigita, A. steini, A. tetracirrata, A. tuberosa, A. turrita, A. zonata*

Table S4 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
		Aspidisca steini	2			1	*Aspidisca steini, Aspidisca (Onychaspis) steini major*	*Aspidisca steini*
Certesiidae	*Certesia*	*Certesia quadrinucleata*	1			2	*Certesia ovata, Certesia quadrinucleata*	*Certesia quadrinucleata*
Discocephalidae	*Discocephalus*	*Discocephalus ehrenbergi*	1			4	*Discocephalus ehrenbergi, Discocephalus grandis, Discocephalus minimus, Discocephalus rotatorius*	*Discocephalus ehrenbergi*
		Discocephalus pararotatorius	1			1	Not found	*Discocephalus pararotatorius*
	Paradiscocephalus	*Paradiscocephalus elongatus*	1			1	Not found	*Paradiscocephalus elongatus*
		Paradiscocephalus sp.	1				Not found	Not found
	Prodiscocephalus	*Prodiscocephalus borrori*	1			2	*Prodiscocephalus minimus*	*Prodiscocephalus borrori*
Euplotia_X	Euplotia_XX	Euplotia_XX sp.	2				Subclass. Namebearing genus of this class is *Euplotes*.	Not found
Euplotidae	*Euplotes*	*Euplotes aediculatus*	8			1	*Euplotes aediculatus*	*Euplotes aediculatus*
		Euplotes balteatus	1			1	*Euplotes balteatus*	*Euplotes balteatus*
		Euplotes bisulcatus	1			1	*Euplotes bisulcatus*	*Euplotes bisulcatus*
		Euplotes charon	3			3	*Euplotes charon, Euplotes charon carinata, Euplotes charon marina*	*Euplotes charon*
		Euplotes cristatus	1			1	*Euplotes cristatus*	*Euplotes cristatus*
		Euplotes daidaleos	5			1	*Euplotes daidaleos*	*Euplotes daidaleos*
		Euplotes dammamensis	1			1	Not found	*Euplotes dammamensis*
		Euplotes elegans	1			2	*Euplotes elegans, Euplotes elegans litoralis*	*Euplotes elegans*
		Euplotes encysticus	2			1	*Euplotes encysticus*	*Euplotes encysticus*
		Euplotes euryhalinus	4			1	*Euplotes euryhalinus*	*Euplotes euryhalinus*
		Euplotes eurystomus	5			3	*Euplotes eurystomus, Euplotes eurystomus excavatus, Euplotes eurystomus marinus*	*Euplotes eurystomus*
		Euplotes focardii	2			1	*Euplotes focardii*	*Euplotes focardii*
		Euplotes harpa	7			3	*Euplotes harpa, Euplotes harpa baikalensis, Euplotes harpa marina*	*Euplotes harpa*
		Euplotes magnicirratus	3			1	*Euplotes magnicirratus*	*Euplotes magnicirratus*
		Euplotes minuta	14			1	*Euplotes minuta*	*Euplotes minuta*
		Euplotes muscicola	2			3	*Euplotes muscicola, Euplotes muscicola alatus, Euplotes muscicola bialatus*	*Euplotes muscicola*
		Euplotes muscorum	2			1	*Euplotes muscorum*	*Euplotes muscorum*
		Euplotes neapolitanus	1			1	*Euplotes neapolitanus*	*Euplotes neapolitanus*
		Euplotes nobilii	24			1	Not found	*Euplotes nobilii*
		Euplotes novemcarinatus	1			1	*Euplotes novemcarinatus*	*Euplotes novemcarinatus*

Table S4 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
Euplotidae	*Euplotes*	*Euplotes octocarinatus*	3			1	*Euplotes octocarinatus*	*Euplotes octocarinatus*
		Euplotes orientalis	1			1	Not found	*Euplotes orientalis*
		Euplotes parabalteatus	1			1	Not found	*Euplotes parabalteatus*
		Euplotes parawoodruffi	1			1	*Euplotes parawoodruffi*	*Euplotes parawoodruffi*
		Euplotes parkei	1			1	*Euplotes parkei*	*Euplotes parkei*
		Euplotes patella	1			10	*Euplotes patella, E. patella alatus, E. patella eurystomus, E. patella flabellata, E. patella gigas, E. patella latus, E. patella lemani, E. patella planctonica, E. patella typicus, E. patella verticillata*	*Euplotes patella*
		Euplotes plicatum	1			1	*Euplotes plicatum*	*Euplotes plicatum*
		Euplotes quinquecarinatus	1			1	*Euplotes quinquecarinatus*	*Euplotes quinquecarinatus*
		Euplotes raikovi	4			1	*Euplotes raikovi*	*Euplotes raikovi*
		Euplotes rariseta	5			1	*Euplotes rariseta*	*Euplotes rariseta*
		Euplotes sinicus	2			1	Not found	*Euplotes sinicus*
		Euplotes sp.	71			153	*Euplotes aberrans, E. acanthodus, E. aculeatus, E. affinis, E. affinis tricirrata, E. agamalievi, E. alatus, E. albus, E. algivora, E. amieti, E. antarcticus, E. appendiculatus, E. apsheronicus, E. arenularum, E. balteata, E. bicornis, E. bistylus, E. camurus, E. carinata, E. caudata, E. charonopsis, E. cimex, E. cithara, E. corsica, E. crassus, E. crassus minor, E. crenosus, E. dogieli, E. excavatus, E. extensus, E. finki, E. gabrieli, E. garabagi, E. gracilis, E. grandis, E. iliffei, E. indentatus, E. inkystans, E. kasymovi, E. kurekchayi, E. labiatus, E. latus, E. leticiensis, E. longicirratus, E. longipes, E. longiremis, E. margherensis, E. marinus, E. marioni, E. mediterraneus, E. michaellae, E. (Mideuplotes) tegulatus, E. minima, E. modunensis, E. moebiusi, E. moebiusi quadricirratus, E. (Moneuplotes) balticus, E. (Moneuplotes) crassus, E. (Moneuplotes) cristatus, E. (Moneuplotes) cristatus, E. (Moneuplotes) minutus, E. (Moneuplotes) vannus, E. monostylus, E. mutabilis, E. mutagens, E. nana, E. (Neteuplotes) elegans, E. (Neteuplotes) moebiusi, E. (Neteuplotes) muscicola, E. (Neteuplotes) muscorum, E. (Neteuplotes) quadricirratus, E. nobilis, E. novemcarinata, E. octocirratus, E. ogusi, E. oropensis, E. ouinquecarinatus, E. palustris, E. paradoxa, E. platystoma, E. plumipes, E. poljanskyi, E. polycarinatus, E. psammophilus, E. pseudocharon, E. pterotae, E. roscoffensis, E. rotunda, E. salina, E. sexcostatus, E. shanghaiensis, E. sharuri, E. sigmolateralis, E. strelkovi, E. striatus, E. subrotundus, E. taylori, E. tegulatus, E. terrestris, E. terricola, E. thononensis, E. truncata, E. truncatus, E. tuffraui, E. turritus, E. vannus balticus, E. variabilis, E. violaceus, E. viridis, E. worcesteri, E. zenkewitchi*	*Euplotes alatus*, *Euplotes* cf. *antarcticus*, *Euplotes cristatus*, *Euplotes eurystomus*, *Euplotes kahli*, *Euplotes magnicirratus*, *Euplotes muscicola*, *Euplotes muscorum*, *Euplotes neapolitanus*, *Euplotes petzi*, *Euplotes plumipes*, *Euplotes* sp. AgTo2, *Euplotes* sp. B11, *Euplotes* sp. BB-2004, *Euplotes* sp. EdPoA02, *Euplotes* sp. EdPoB02, *Euplotes* sp. EMP, *Euplotes* sp. ER-2014, *Euplotes* sp. GZJJM2009121510, *Euplotes* sp. PUHP, *Euplotes* sp. SNK-2011, *Euplotes* sp. WSea3
		Euplotes trisulcatus	1			1	*Euplotes trisulcatus*	*Euplotes trisulcatus*
		Euplotes vannus	11			1	*Euplotes vannus*	*Euplotes vannus*
		Euplotes woodruffi	13			1	*Euplotes woodruffi*	*Euplotes woodruffi*
	Euplotoides	*Euplotoides vannus*	1			11	*Euplotoides aediculatus, E. agamalievi, E. amieti, E. daidaleos, E. dogieli, E. eurystomus, E. octocarinatus, E. palustris, E. patella, E. plumipes, E. woodruffi*	*Euplotoides vannus*
	Hypotrichia	*Hypotrichia* sp.	1				Is a subclass. See Hypotrichia (Table S3)	Contains the taxonomic groups: Euplotida and Kiitrichida. *Hypotrichia* sp. I-99, *Hypotrichida* sp. AL, *Hypotrichida* sp. Florida, *Hypotrichida* sp. KEC2002, *Hypotrichida* sp. LPJ-2005, *Hypotrichida* sp. Misty, *Hypotrichida* sp. OrrK1999. Included is: *Parabistichella variabilis*

Table S4 continued.

Family / Superorder	Genus	Species	#Database Acc. No.	UIRs	OTUs	#Species described	Described morphospecies - Berger 2001*	Described morphospecies - EOL**
	Moneuplotes	*Moneuplotes crassus*	33			6	*Moneuplotes balticus, M. crassus, M. cristatus, M. minuta, M. minutus, M. mutabilis*	*Euplotes crassus* subsp. *crassus*, *Euplotes crassus* subsp. *minor*
		Moneuplotes vannus	2			2	*Moneuplotes vannus*	Moneuplotes minuta
Gastrocirrhidae	*Euplotidium*	*Euplotidium arenarium*	2			1	*Euplotidium arenarium*	Euplotidium arenarium
		Euplotidium itoi	2			1	*Euplotidium itoi*	Not found
		Euplotidium sp.	1			5	*Euplotidium agitatum, Euplotidium helgae, Euplotidium prosaltans, Euplotidium psammophilus*	*Euplotidium agitatum, Euplotidium helgae, Euplotidium prosaltans, Euplotidium psammophilum, Euplotidium smalli*
	Gastrocirrhus	*Gastrocirrhus monilifer*	1			5	*Gastrocirrhus adhaerens, G. intermedius, G. (Spongiocirrhus) adhaerens, G. stentoreus, G. trichocystus*	*Gastrocirrhus monilifer, Gastrocirrhus stentoreus*
Uronychiidae	*Apodiophrys*	*Apodiophrys ovalis*	1			1	Not found	*Apodiophrys ovalis*
	Diophryopsis	*Diophryopsis hystrix*	1			1	Genus: *Diophrys*	*Diophryopsis hystrix*
	Diophrys	*Diophrys apoligothrix*	2			1	Not found	*Diophrys apoligothrix*
		Diophrys appendiculata	2			2	*Diophrys appendiculata, Diophrys appendiculata samuchi,*	*Diophrys appendiculata*
		Diophrys japonica	1				Not found	*Diophrys japonica*
		Diophrys oligothrix	5			1	*Diophrys oligothrix*	*Diophrys oligothrix*
		Diophrys parappendiculata	2				Not found	*Diophrys parappendiculata*
		Diophrys scutum	6			1	*Diophrys scutum*	*Diophrys scutum*
		Diophrys sp.	3			22	*Diophrys bifaria, Diophrys marina, Diophrys scutum, Diophrys (Epidiophrys) quadricaudatus, Diophrys grandis, Diophrys hystrix, Diophrys irmgard, Diophrys kahli, Diophrys kasymovi, Diophrys magnus, Diophrys (Monilidiophrys) magnus, Diophrys multicirratus, Diophrys multinucleata, Diophrys peloetes, Diophrys pentacirratus, Diophrys (Polydiophrys) tetramacronucleata, Diophrys quadricaudatus, Diophrys salina, Diophrys scutoides, Diophrys tetramacronucleata, Diophrys triangulata*	*Diophrys bifaria, Diophrys grandis, Diophrys kahli, Diophrys kasymovi, Diophrys magnus, Diophrys multicirratus, Diophrys norwegicus, Diophrys peloetes, Diophrys pentacirratus, Diophrys quadricaudata, Diophrys salina, Diophrys scutoides, Diophrys tetramacronucleata*
	Heterodiophrys	*Heterodiophrys zhui*	1			1	Not found	*Heterodiophrys zhui*
	Paradiophrys	*Paradiophrys irmgard*	2			1	*Paradiophrys irmgard*	*Paradiophrys irmgard, Paradiophrys kahli*
							Paradiophrys kahli, Paradiophrys multinucleata, Paradiophrys (Lacifer) multinucleata,	*Paradiophrys irmgard, Paradiophrys kahli, Paradiophrys zhangi*
		Paradiophrys zhangi	1			5	*Paradiophrys multinucleata*	
	Pseudodiophrys	*Pseudodiophrys nigricans*	2			1	Not found	Pseudodiophrys nigricans
	Uronychia	*Uronychia binucleata*	2			1	*Uronychia binucleata*	*Uronychia binucleata*
		Uronychia multicirrus	2			1	*Uronychia multicirrus*	*Uronychia multicirrus*
		Uronychia setigera	5			1	*Uronychia setigera*	*Uronychia setigera*
		Uronychia sp.	4			9	*Uronychia antarctica, Uronychia bivalvorum, Uronychia festinans, Uronychia heinrothi, Uronychia invicta, Uronychia magna, Uronychia paupera, Uronychia transfuga, Uronychia uncinata*	*Uronychia transfuga*

Table S5. A list of organisms and their accession numbers detected at ≥97% pairwise identity in all 150 grassland soil sampling sites.

Clas	Order	Genus and species (Accession number)
Colpodea	Bursariomorphida	*Bryometopus pseudochilodon* (EU039887), *Bryometopus sphagni* (AF060455), *Bryometopus triquetrus* (KJ873045), *Bursaria* sp. (AB695450), Bursariomorphida_XX sp (GU479974)
	Colpodida	*Bresslaua vorax* (AF060453), *Bresslauides discoideus* (EU039885), *Colpoda aspera* (EU039892), *Colpoda aspera* (KF111344), *Colpoda cucullus* (EU039893), *Colpoda elliotti* (KJ873047), *Colpoda inflata* (KJ607917), *Colpoda inflata* (KJ607918), *Colpoda lucida* (KJ873048), *Colpoda maupasi* (JF747215), *Colpoda* sp. (EF024004), *Colpoda* sp. (JF747216), *Colpoda steinii* (KJ607913), *Colpoda steinii* (KJ607916), Colpodida_XX sp. (KJ609540), Colpodidae_X sp. (JN020238), Colpodidae_X sp. (JN635479), Colpodidae_X sp. (JN635481), Colpodidae_X sp. (JN635488), *Exocolpoda augustini* (JF747214), *Exocolpoda augustini* (KJ607919), *Hausmanniella discoidea* (EU039900), *Kalometopia duplicata* (KJ873050), *Kreyellidae* sp. (JQ723971), *Maryna ovata* (HQ337902), *Maryna* sp. (JF747218), *Maryna umbrellata* (JF747217), *Mykophagophrys terricola* (EU039902), *Pseudoplatyophrya nana* (AF060452)
	Cyrtolophosidida	Cyrtolophosididae_X sp. (EF023400), Cyrtolophosididae_X sp. (FJ592470), *Cyrtolophosis mucicola* (AB695454), *Cyrtolophosis mucicola* (EU039899), *Microdiaphanosoma arcuatum* (GU997633), *Pseudocyrtolophosis alpestris* (EU264564)
	Platyophryida	*Platyophrya bromelicola* (EU039905), *Platyophrya* sp. (EF024263), *Platyophrya* sp. (EF100341), *Platyophrya vorax* (AF060454), *Platyophryides* sp. (KJ873052), *Sorogena stoianovitchae* (AF300285), *Woodruffides metabolicus* (JQ356869)
Heterotrichea	Heterotrichida	Blepharisma hyalinum (AM713184), Blepharisma musculus (KJ651813)
Litostomatea	Cyclotrichia	Cyclotrichia_XX sp. (GU647177)
	Haptoria	*Acaryophrya* sp. (KF733758), Amphileptidae_X sp. (GU972138), *Arcuospathidium cultriforme* (DQ411860), *Arcuospathidium namibiense* (JF263442), *Balantidion pellucidum* (AY821917), *Cultellothrix coemeterii* (KF733755), *Cultellothrix lionotiformis* (JF263445), *Dileptus costaricanus* (HM581679), *Dileptus microstoma* (HM581676), *Dileptus mucronatus* (HM581675), *Enchelyodon* sp. (JF263446), *Enchelys polynucleata* (DQ411861), *Epispathidium papilliferum* (DQ411858), Haptoria_XX sp. (AB695503), Haptoria_XX sp. (AB725341), Haptoria_XX sp. (AB725344), Haptoria_XX sp. (AB725346), Haptoria_XX sp. (AM114813), Haptoria_XX sp. (HQ219407), Lacrymariidae_X sp. (AB695505), Lacrymariidae_X sp. (FN689997), *Loxophyllum jini* (EF123708), *Pelagodileptus trachelioides* (KJ925348), *Phialina salinarum* (HM140391), *Protospathidium muscicola* (JF263449), *Pseudomonilicaryon* sp. (HM581677), *Spathidium foissneri* (KF733756), *Spathidium rectitoratum* (KF733757), *Spathidium stammeri* (DQ411862), *Trachelophyllum* sp. (JF263452)
	Trichostomatia	Polycostidae_X sp. (EF024292)
Nassophorea	Nassophorea_X	*Colpodidiidae* sp. (EU264561), *Furgasonia blochmanni* (X65150), *Leptopharynx costatus* (EU286811), Nassophorea_XXX sp. (EF024295), *Obertrumia georgiana* (X65149)
Oligohymenophorea	Astomatia	*Anoplophrya marylandensis* (AY547546)
	Peniculia	Peniculia_XX sp. (AB725340)
	Peritrichia	*Epistylis* sp. (JQ723982), *Opisthonecta henneguyi* (JN120201), Peritrichia-1_X sp. (JN020239), *Zoothamnium duplicatum* (JX457451)
	Scuticociliatia	*Homalogastra setosa* (EF158848), Orchitophryidae_X sp. (EF024585), Scuticociliatia_XX sp. (AB695448)
	Sessilida	*Vorticella gracilis* (GQ872429), *Vorticella microstoma* (JN120200), *Vorticella microstoma* (JN120206), *Vorticella* sp. (JN120223), Vorticellidae_X sp. (AY835672), Vorticellidae_X sp. (HQ219427), Vorticellidae_X sp. (JX457442), *Vorticellides astyliformis* (GQ872427), *Vorticellides* sp. (JQ723991)
	Tetrahymenida	*Tetrahymena rostrata* (JQ045342), *Tetrahymena* sp. (KJ028502), Tetrahymenidae_X sp. (KJ028482), Tetrahymenidae_X sp. (KJ028519)
Prostomatea	Prostomatea-3	*Prorodon teres* (AY821920)
Spirotrichea	Choreotrichia	*Strobilidium caudatum* (AY143573)
	Hypotrichia	*Amphisiella magnigranulosa* (AM412774), *Bergeriella ovata* (KJ925302), *Bistichella* FG-2014 (KJ509196), *Bistichella variabilis* (HQ699895), *Cyrtohymena citrina* (KC182574), *Deviata bacilliformis* (KJ766110), *Engelmanniella mobilis* (AF164134), *Gastrostyla* sp. (KC411832), *Gastrostyla steinii* (EU647172), *Gonostomum namibiense* (AY498655), *Gonostomum* sp. (EF024472), *Gonostomum* sp. (JQ723970), *Gonostomum strenuum* (AJ310493), *Gonostomum strenuum* (FJ592429), *Halteria grandinella* (AB695451), *Halteria grandinella* (AF164137), *Halteria grandinella* (JF730812), Hypotrichia_XX sp. (AB725345), Hypotrichia_XX sp. (AF372826), Hypotrichia_XX sp. (GU479973), Hypotrichia_XX sp. (JN020236), Hypotrichia_XX sp. (KF517003), *Kahliella* sp. (EU079472), *Laurentiella strenua* (AJ310487), *Orthamphisiella breviseries* (AY498654), *Oxytricha lanceolata* (AM412773), *Oxytricha longa* (AF508763), *Oxytricha* sp. (AF508776), *Oxytricha* sp. (EF024941), *Oxytricha* sp. (EF441988), *Oxytricha* sp. (FN429124), *Oxytricha* sp. (JQ723978), *Oxytricha* sp. (JQ723979), *Oxytricha trifallax* (AC237474), Oxytrichidae_X sp. (EF023177), Oxytrichidae_X sp. (EF023337), Oxytrichidae_X sp. (EF024290), Oxytrichidae_X sp. (EF024618), Oxytrichidae_X sp. (EF024639), Oxytrichidae_X sp. (EF024684), Oxytrichidae_X sp. (EF024731), Oxytrichidae_X sp. (EF024748), Oxytrichidae_X sp. (EF024775), Oxytrichidae_X sp. (EF024903), Oxytrichidae_X sp. (EF024909), Oxytrichidae_X sp. (EF024975), *Paraparentocirrus sibillinensis* (KF184655), *Perisincirra paucicirrata* (JX012184), *Pleurozia purpurea* (AY607864), *Pseudokeronopsis* sp. (FJ775724), *Pseudouroleptus* sp. (KJ173910), *Pseudourostyla cristata* (DQ019318), *Pseudourostyla franzi* (AM412765), *Sterkiella* sp. (JX946274), *Sterkiella* sp. (KC404828), *Stichotrichia* sp. (AB449360), *Stichotrichia* sp. (AB449361), *Stichotrichia* sp. (AB449362), *Stylonychia lemnae* (AM233917), *Uroleptus gallina* (AF164130), *Uroleptus* sp. (AY294646), *Urostyla grandis* (AF508781), *Urostyla grandis* (FJ577813)
	Oligotrichia	*Strombidium* sp. (KJ759912)
	Spirotrichea_X	*Phacodinium metchnikoffi* (AJ277877)

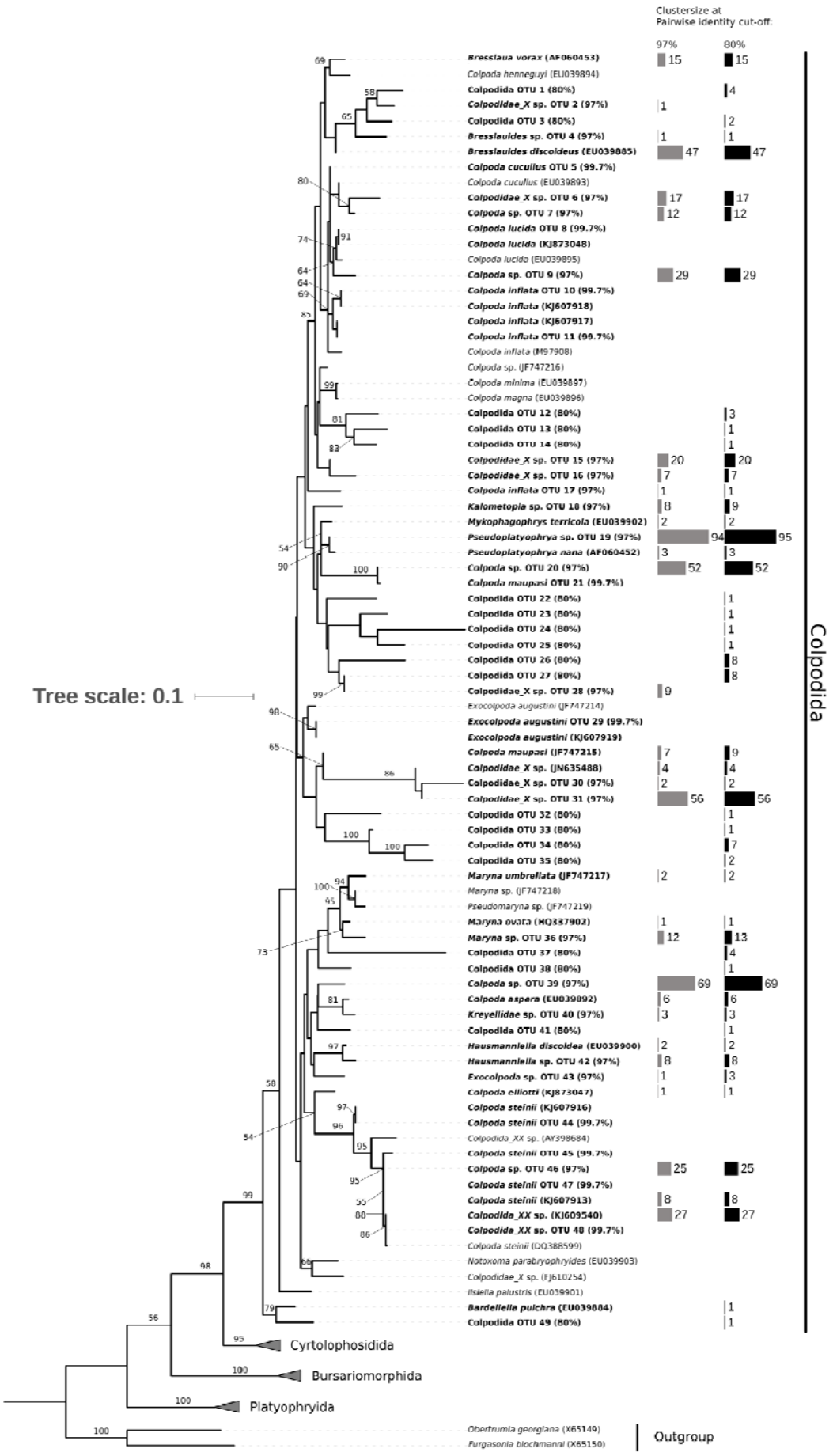

Clustersize at
Pairwise identity cut-off:
97%
80%
Bresslaua vorax (AF060453)
Colpoda henneguyi (EU039894)
Colpodida OTU 1 (80%)
Colpodidae_X sp. OTU 2 (97%)
Colpodida OTU 3 (80%)
Bresslauides sp. OTU 4 (97%)
Bresslauides discoideus (EU039885)
Colpoda cucullus OTU 5 (99.7%)
Colpoda cucullus (EU039893)
Colpodidae_X sp. OTU 6 (97%)
Colpoda sp. OTU 7 (97%)
Colpoda lucida OTU 8 (99.7%)
Colpoda lucida (KJ873048)
Colpoda lucida (EU039895)
Colpoda sp. OTU 9 (97%)
Colpoda inflata OTU 10 (99.7%)
Colpoda inflata (KJ607918)
Colpoda inflata (KJ607917)
Colpoda inflata OTU 11 (99.7%)
Colpoda inflata (M97908)
Colpoda sp. (JF747216)
Colpoda minima (EU039897)
Colpoda magna (EU039896)
Colpodida OTU 12 (80%)
Colpodida OTU 13 (80%)
Colpodida OTU 14 (80%)
Colpodidae_X sp. OTU 15 (97%)
Colpodidae_X sp. OTU 16 (97%)
Colpoda inflata OTU 17 (97%)
Kalometopia sp. OTU 18 (97%)
Mykophagophrys terricola (EU039902)
Pseudoplatyophrya sp. OTU 19 (97%)
Pseudoplatyophrya nana (AF060452)
Colpoda sp. OTU 20 (97%)
Colpoda maupasi OTU 21 (99.7%)
Colpodida OTU 22 (80%)
Colpodida OTU 23 (80%)
Colpodida OTU 24 (80%)
Colpodida OTU 25 (80%)
Colpodida OTU 26 (80%)
Colpodida OTU 27 (80%)
Colpodidae_X sp. OTU 28 (97%)
Exocolpoda augustini (JF747214)
Exocolpoda augustini OTU 29 (99.7%)
Exocolpoda augustini (KJ607919)
Colpoda maupasi (JF747215)
Colpodidae_X sp. (JN635488)
Colpodidae_X sp. OTU 30 (97%)
Colpodidae_X sp. OTU 31 (97%)
Colpodida OTU 32 (80%)
Colpodida OTU 33 (80%)
Colpodida OTU 34 (80%)
Colpodida OTU 35 (80%)
Maryna umbrellata (JF747217)
Maryna sp. (JF747218)
Pseudomaryna sp. (JF747219)
Maryna ovata (HQ337902)
Maryna sp. OTU 36 (97%)
Colpodida OTU 37 (80%)
Colpodida OTU 38 (80%)
Colpoda sp. OTU 39 (97%)
Colpoda aspera (EU039892)
Kreyellidae sp. OTU 40 (97%)
Colpodida OTU 41 (80%)
Hausmanniella discoidea (EU039900)
Hausmanniella sp. OTU 42 (97%)
Exocolpoda sp. OTU 43 (97%)
Colpoda elliotti (KJ873047)
Colpoda steinii (KJ607916)
Colpoda steinii OTU 44 (99.7%)
Colpodida_XX sp. (AY398684)
Colpoda steinii OTU 45 (99.7%)
Colpoda sp. OTU 46 (97%)
Colpoda steinii OTU 47 (99.7%)
Colpoda steinii (KJ607913)
Colpodida_XX sp. (KJ609540)
Colpodida_XX sp. OTU 48 (99.7%)
Colpoda steinii (DQ388599)
Notoxoma parabryophryides (EU039903)
Colpodidae_X sp. (FJ610254)
Ilsiella palustris (EU039901)
Bardeliella pulchra (EU039884)
Colpodida OTU 49 (80%)
Tree scale: 0.1
Colpodida
Cyrtolophosidida
Bursariomorphida
Platyophryida
Obertrumia georgiana (X65149)
Furgasonia blochmanni (X65150)
Outgroup

Figure S1 (Previous page). Phylogenetic placement of unique individual read (UIR) clusters within the known diversity for the colpodean classes of Cyrtolophosidida, Bursariomorphida and Plathyophryida. UIRs identified at ≥99.7% -, ≥97% -, and ≥80% sequence pairwise identity within the class group was inferred by the V4 SSU gene alignment to protist ribosomal reference (PR^2) database sequences and then clustered together with the returned reference sequences at ≥97% global similarity. Known and unknown diversity was included into the basic maximum likelihood (ML) tree structure adopted from Foissner (2011) and Foissner et al. (2014) originally containing the 51 reference taxa. Barcharts at the bold fonted leaves indicate the number of UIRs associated to the respective cluster at ≥97% - and ≥80% sequence pairwise identity level resolution to reference sequences.

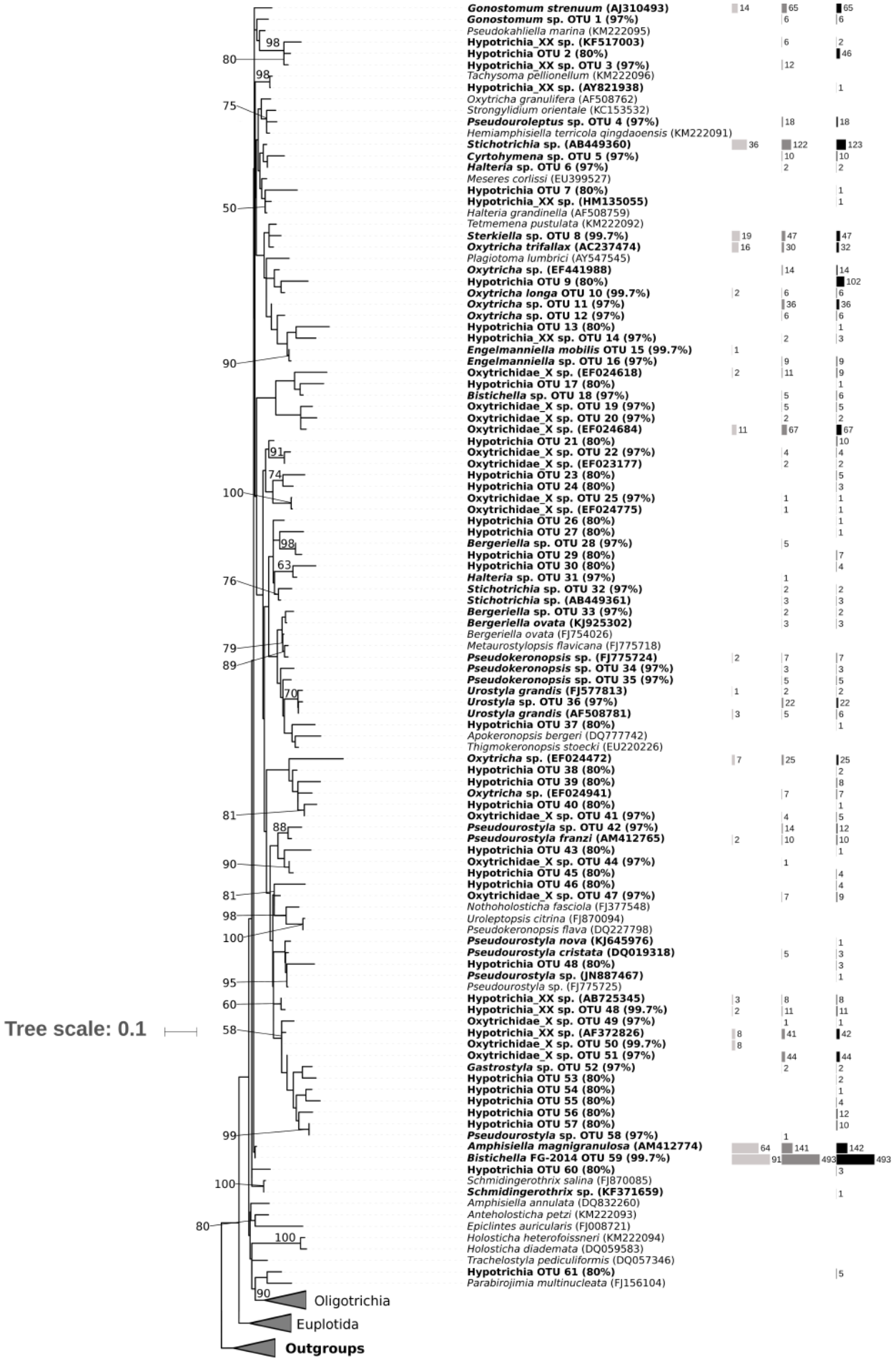

Gonostomum strenuum (AJ310493)
Gonostomum sp. OTU 1 (97%)
Pseudokahliella marina (KM222095)
Hypotrichia_XX sp. (KF517003)
Hypotrichia OTU 2 (80%)
Hypotrichia_XX sp. OTU 3 (97%)
Tachysoma pellionellum (KM222096)
Hypotrichia_XX sp. (AY821938)
Oxytricha granulifera (AF508762)
Strongylidium orientale (KC153532)
Pseudouroleptus sp. OTU 4 (97%)
Hemiamphisiella terricola qingdaoensis (KM222091)
Stichotrichia sp. (AB449360)
Cyrtohymena sp. OTU 5 (97%)
Halteria sp. OTU 6 (97%)
Meseres corlissi (EU399527)
Hypotrichia OTU 7 (80%)
Hypotrichia_XX sp. (HM135055)
Halteria grandinella (AF508759)
Tetmemena pustulata (KM222092)
Sterkiella sp. OTU 8 (99.7%)
Oxytricha trifallax (AC237474)
Plagiotoma lumbrici (AY547545)
Oxytricha sp. (EF441988)
Hypotrichia OTU 9 (80%)
Oxytricha longa OTU 10 (99.7%)
Oxytricha sp. OTU 11 (97%)
Oxytricha sp. OTU 12 (97%)
Hypotrichia OTU 13 (80%)
Hypotrichia_XX sp. OTU 14 (97%)
Engelmanniella mobilis OTU 15 (99.7%)
Engelmanniella sp. OTU 16 (97%)
Oxytrichidae_X sp. (EF024618)
Hypotrichia OTU 17 (80%)
Bistichella sp. OTU 18 (97%)
Oxytrichidae_X sp. OTU 19 (97%)
Oxytrichidae_X sp. OTU 20 (97%)
Oxytrichidae_X sp. (EF024684)
Hypotrichia OTU 21 (80%)
Oxytrichidae_X sp. OTU 22 (97%)
Oxytrichidae_X sp. (EF023177)
Hypotrichia OTU 23 (80%)
Hypotrichia OTU 24 (80%)
Oxytrichidae_X sp. OTU 25 (97%)
Oxytrichidae_X sp. (EF024775)
Hypotrichia OTU 26 (80%)
Hypotrichia OTU 27 (80%)
Bergeriella sp. OTU 28 (97%)
Hypotrichia OTU 29 (80%)
Hypotrichia OTU 30 (80%)
Halteria sp. OTU 31 (97%)
Stichotrichia sp. OTU 32 (97%)
Stichotrichia sp. (AB449361)
Bergeriella sp. OTU 33 (97%)
Bergeriella ovata (KJ925302)
Bergeriella ovata (FJ754026)
Metaurostylopsis flavicana (FJ775718)
Pseudokeronopsis sp. (FJ775724)
Pseudokeronopsis sp. OTU 34 (97%)
Pseudokeronopsis sp. OTU 35 (97%)
Urostyla grandis (FJ577813)
Urostyla sp. OTU 36 (97%)
Urostyla grandis (AF508781)
Hypotrichia OTU 37 (80%)
Apokeronopsis bergeri (DQ777742)
Thigmokeronopsis stoecki (EU220226)
Oxytricha sp. (EF024472)
Hypotrichia OTU 38 (80%)
Hypotrichia OTU 39 (80%)
Oxytricha sp. (EF024941)
Hypotrichia OTU 40 (80%)
Oxytrichidae_X sp. OTU 41 (97%)
Pseudourostyla sp. OTU 42 (97%)
Pseudourostyla franzi (AM412765)
Hypotrichia OTU 43 (80%)
Oxytrichidae_X sp. OTU 44 (97%)
Hypotrichia OTU 45 (80%)
Hypotrichia OTU 46 (80%)
Oxytrichidae_X sp. OTU 47 (97%)
Nothoholosticha fasciola (FJ377548)
Uroleptopsis citrina (FJ870094)
Pseudokeronopsis flava (DQ227798)
Pseudourostyla nova (KJ645976)
Pseudourostyla cristata (DQ019318)
Hypotrichia OTU 48 (80%)
Pseudourostyla sp. (JN887467)
Pseudourostyla sp. (FJ775725)
Hypotrichia_XX sp. (AB725345)
Hypotrichia_XX sp. OTU 48 (99.7%)
Oxytrichidae_X sp. OTU 49 (97%)
Hypotrichia_XX sp. (AF372826)
Oxytrichidae_X sp. OTU 50 (99.7%)
Oxytrichidae_X sp. OTU 51 (97%)
Gastrostyla sp. OTU 52 (97%)
Hypotrichia OTU 53 (80%)
Hypotrichia OTU 54 (80%)
Hypotrichia OTU 55 (80%)
Hypotrichia OTU 56 (80%)
Hypotrichia OTU 57 (80%)
Pseudourostyla sp. OTU 58 (97%)
Amphisiella magnigranulosa (AM412774)
Bistichella FG-2014 OTU 59 (99.7%)
Hypotrichia OTU 60 (80%)
Schmidingerothrix salina (FJ870085)
Schmidingerothrix sp. (KF371659)
Amphisiella annulata (DQ832260)
Anteholosticha petzi (KM222093)
Epiclintes auricularis (FJ008721)
Holosticha heterofoissneri (KM222094)
Holosticha diademata (DQ059583)
Trachelostyla pediculiformis (DQ057346)
Hypotrichia OTU 61 (80%)
Parabirojimia multinucleata (FJ156104)
Tree scale: 0.1
Oligotrichia
Euplotida
Outgroups

Figure S2 (Previous page). Phylogenetic placement of unique individual read (UIR) clusters within the known diversity for the Hypotrichia. UIRs identified at ≥99.7% -, ≥97% -, and ≥80% sequence pairwise identity within the class group was inferred by the V4 SSU gene alignment to protist ribosomal reference (PR^2) database sequences and then clustered together with the returned reference sequences at ≥97% global similarity. Known and unknown diversity was included into the basic maximum likelihood (ML) tree structure adopted from Dunthorn et al. (2014) originally containing the 51 reference taxa. Barcharts at the bold fonted leaves indicate the number of UIRs associated to the respective cluster at ≥97% - and ≥80% sequence pairwise identity level resolution to reference sequences.

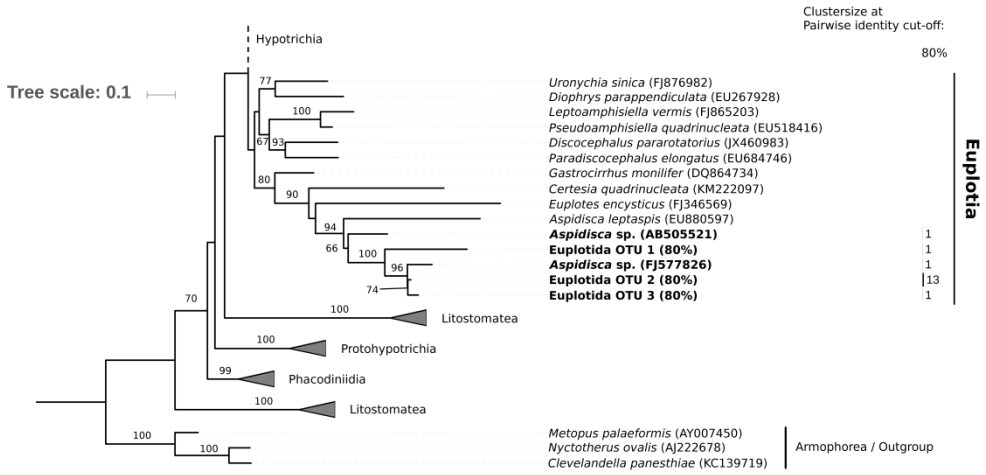
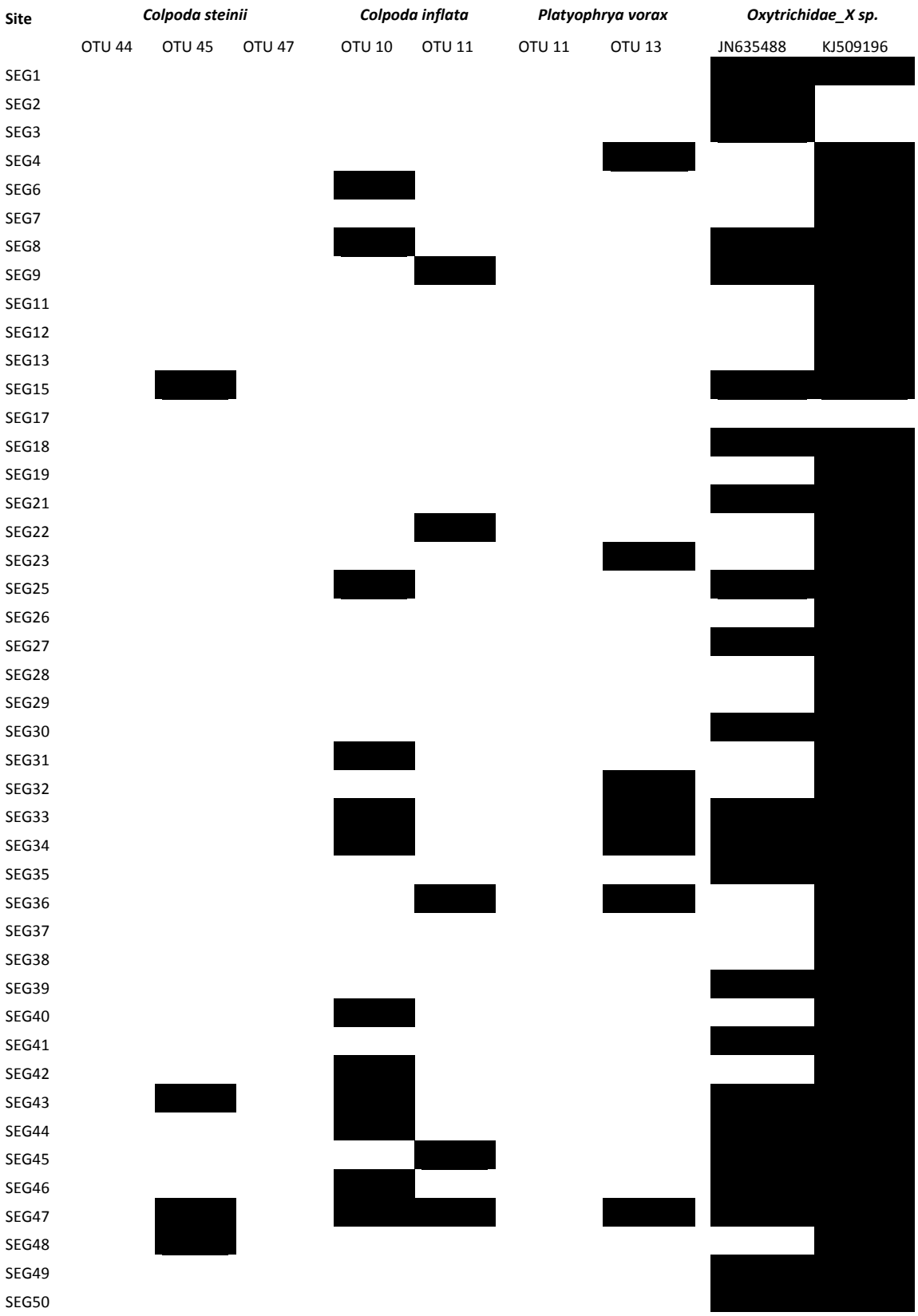


Figure S3. Phylogenetic placement of unique individual read (UIR) clusters within the known diversity for the Euplotia. UIRs identified at ≥99.7% -, ≥97% -, and ≥80% sequence pairwise identity within the class group was inferred by the V4 SSU gene alignment to protist ribosomal reference (PR^2) database sequences and then clustered together with the returned reference sequences at ≥97% global similarity. Known and unknown diversity was included into the basic maximum likelihood (ML) tree structure adopted from Dunthorn et al. (2014) originally containing the 51 reference taxa. Barcharts at the bold fonted leaves indicate the number of UIRs associated to the respective cluster at ≥97% - and ≥80% sequence pairwise identity level resolution to reference sequences.

Figure S4. Heatmaps indicating the geographic occurrences of gene variants for dominant taxa. Gene variants correspond to the different operations taxonomic units (OTUs) and their differential occurrence at the Biodiversity Exploratory sites Schwaebische Alb (AEG), Hainich-Duen (HEG) and Schorfheide-Chorin (SEG).

Figure S4. Continued.

Figure S4. Continued.



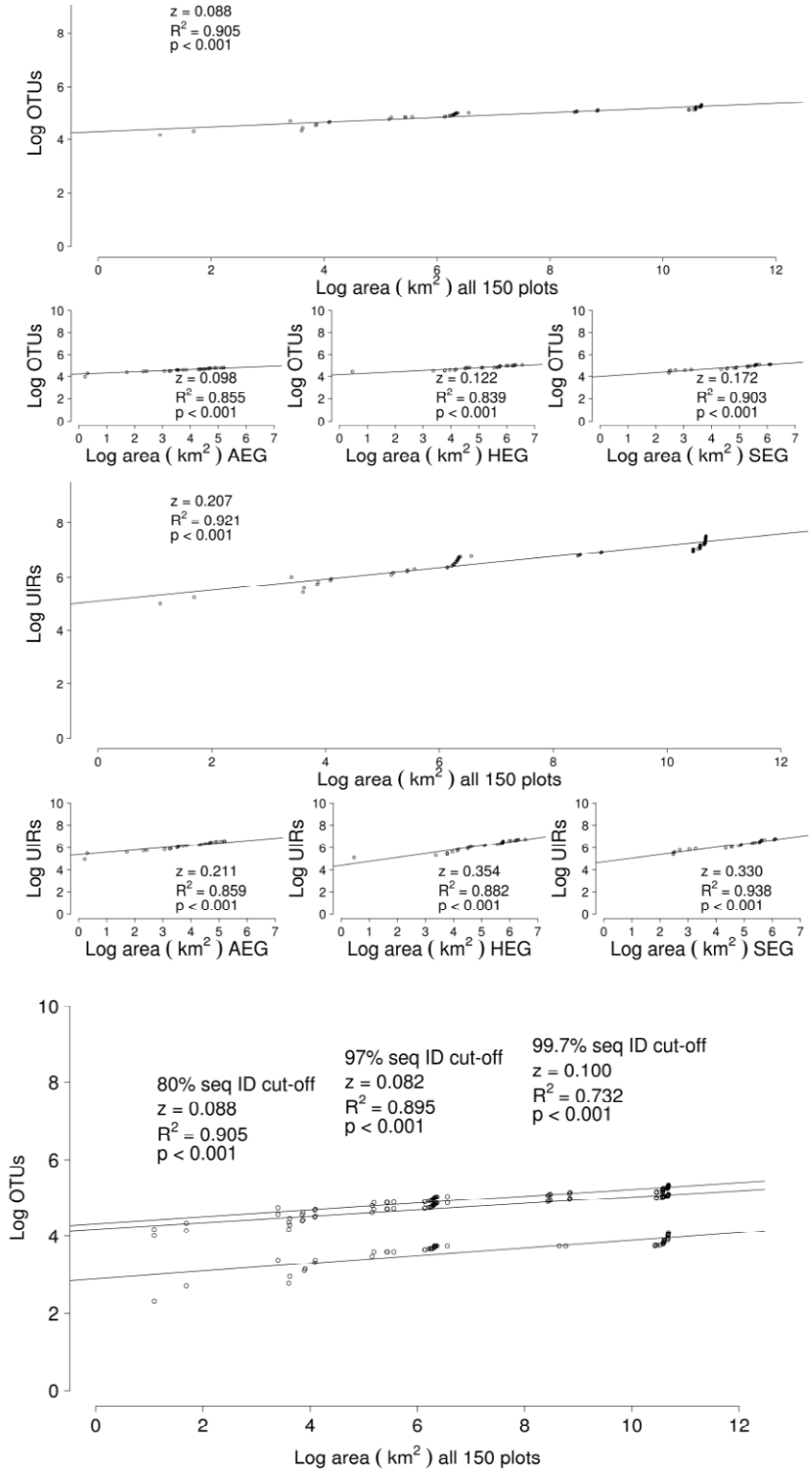


Figure S5. Log-log transformed taxa-area curves comparing all protist unique individual reads (UIRs) to ciliate and non-ciliate UIRs for all Biodiversity Exploratory samples (n=150). Geographic distance was calculated using the coordinates for all sites and presence/absence data for protist UIRs was determined at ≥80% reference sequence similarity. The y-intercept, slope of the taxa-area curve (z-value), amount of variation in species by area (R-squared) and the significance (p-value) are given (all protists = upper line, non Ciliophora = middle line, Ciliophora = lower line). Presence/absence data for UIRs indicating taxa-area gradients (*z*-value) in relation to area size (km²) were adjusted for statistical evenness (see Methods). Mean distance between Biodiversity Exploratory sites: ~311 km (<1 - 626 km).

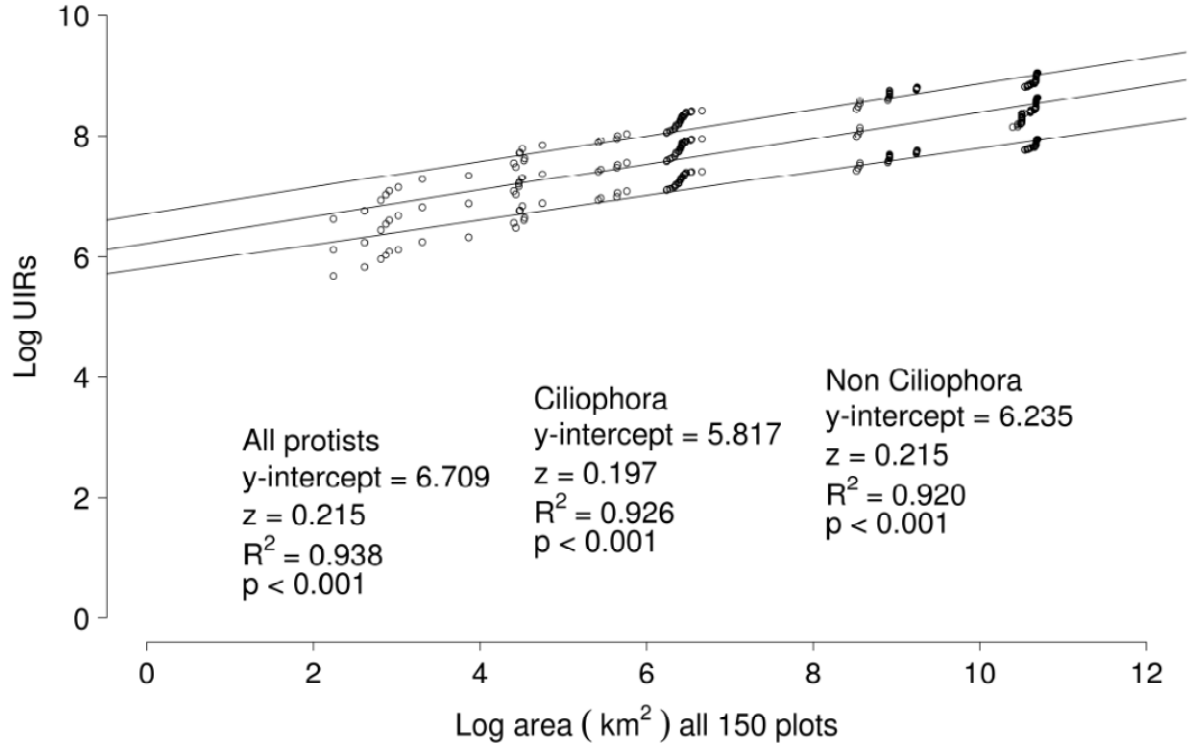


Figure S6. Log-log transformed ciliate taxa-area curves for all Biodiversity Exploratory sites (n=150) in Schwaebische Alb (AEG), Hainich-Duen (HEG) and Schorfheide-Chorin (SEG). (**A**) Taxa-area relationship of Blast based operational taxonomic units (OTUs) for all sites and individual exploratories. (**B**) Taxa-area relationship of unique individual reads (UIRs) for all sites and individual exploratories. (**C**) Taxa-area relationship of OTUs according to ≥80% (upper line), ≥97% (middle line) and ≥99.7% (lower line) sequence pairwise identity cut-offs (seq ID cut-off). Distances are based on geographic coordinates and data for OTUs/UIRs based on presence/absence data, calculated at ≥80% reference sequence similarity. The slope of the taxa-area curve (z-value), variance in OTUs/UIRs by area (R-squared) and the significance (p-value) is given (Methods).

Chapter 7 – The hidden diversity of flagellated protists in soil

The hidden diversity of flagellated protists in soil

Paul Christiaan Venter, Frank Nitsche, Hartmut Arndt*.

University of Cologne, Biocenter Cologne, Institute of Zoology, Department of General Ecology, Zuelpicher Str. 47b, D-50674 Koeln (Cologne), Germany

Running title: The hidden diversity of flagellated protists in soil

Keywords: heterotrophic flagellates; choanoflagellate; bicosoecids; cercomonads; apusozoans; high-throughput sequencing

Abstract

Protists are the most diverse and abundant eukaryotes in soil. However, the gap between described and sequenced protist morphospecies still presents a pending problem when surveying environmental samples for known species using molecular methods. The number of sequences in the database (~130,000) is limited compared to the species richness expected (~1.8 million named species) - limiting the recovery rate. This is important, since high throughput sequencing (HTS) methods are used to find associative patterns between functional traits, taxa and environmental parameters. We surveyed soil flagellates using HTS in 150 grasslands of central Europe, tested the recovery rate of ten previously isolated and cultivated cercomonad species, and compared locally found diversity. We recovered sequences for expected flagellate morphospecies, but also a great number of their genetic variants – possibly different species, among rare and dominant taxa with presumably own biogeography. We recorded dominant (cercozoans, *Sandona*), rare (apusozoans) and a large hidden diversity of predominantly aquatic protists in soil (choanoflagellates, bicosoecids) often forming novel clades associated with uncultured environmental sequences. Evaluating the UIRs, instead of the OTUs that individual reads are usually clustered into, we discovered that much of the hidden diversity may be lost due to clustering.

Introduction

Heterotrophic flagellates are a heterogeneous group of protists comprising colorless grazers of bacteria, other protists, particulate and dissolved organic matter with high diversity (Leadbeater and Green 2000). They are found widespread in soils, with abundances of up to 170,000 cells per gram of soil dry weight (Ekelund et al. 2001; Geisen et al. 2015). Functionally, they link the carbon and nutrient flow between primary producers with higher trophic levels, making them very important players in the microbial food web in soil (e.g. Geisen and Bonkowski 2017). Despite their important role in soil, they are the least studied soil protist group, compared to soil ciliates and testate amoebae, because of their small size of between 1 and 450µm and less distinctive morphology (Foissner 2006). Their accurate taxonomic placement is greatly dependent on phylogenetic analyses of their 18S rRNA genes (Geisen et al. 2015). Typical soil flagellates include cercozoans, bodonids and chrysomonads (Foissner 1991), while rare soil taxa include apusomonads, euglenids and bicosoecids (Ekelund et al. 2001; Griffith et al. 2001; Domonell et al. 2013). Because of the taxonomical uncertainty and morphological identification shortcomings, a combination of culture-independent and culture-dependent methods are necessary to accurately describe their diversity in ecological systems (e.g. Ekelund et al. 2001; Schoenle et al. 2016; Jeuck et al. 2017). This becomes obvious when clades within mainly freshwater and marine taxa (bicosoecids and choanoflagellates) are discovered in terrestrial environments, but largely associates with other environmental sequences only (e.g. del Campo and Massana 2011).

Using the V4 region of the 18S rRNA gene as a marker in high-throughput sequencing (HTS) is considered a quick and reliable method for describing flagellate communities (del Campo and Massana 2011). HTS diversity for cercozoans

correlates well with Sanger sequences for the same marker gene (Dunthorn et al. 2014; Harder et al. 2016). Even though dominant protists in soil can be identified, a comparison of morphological and molecular methods indicates that community dominance patters are not always the same between methods used (Domonell 2013; Santoferrara et al. 2014). To reliably resolve known and unknown taxa as well as to find novel clades in environmental samples, HTS metabarcoding primers need to be specific (Pawlowski et al. 2012), presence/absence data is preferred to equate relative richness within geographic scales (Venter et al. 2017; Zaiko et al. 2015) and partial sequences must be put in a phylogenetic context with full length 18S SSU gene sequences from described morphospecies (Dunthorn et al. 2014). Using the V4 region of this gene is enough to resolve species diversity for the main flagellate lineages (Pawlowski et al. 2012). Very few studies examine the recovery of described morphospecies and their genetic variability under semi-natural conditions (e.g. Glücksman et al. 2010), and there is a lack of recovery rate studies in natural environments. This is especially important, because even ecological and behavioral differences among the same species can be identified as intraspecific variation (Glücksman et al. 2010), where interspecific variation among protist species can reflect differences in ingestion- and growth-rates (Weisse 2002; Weisse et al. 2016). These clonal variations within the same species is thought to be caused by epigenetic pressures such as bacterial population shifts leading to grazing pressures on bacterivores, making protists excellent indicators of changing environmental conditions (Foissner 1999; Glücksman et al. 2010; Weisse 2002).

The majority of HTS 18S rRNA marker gene data present many novel environmental sequences, with large distances to cultured clones. Additionally, some of the smaller taxa like Apusozoa, have not yet been recovered in HTS metabarcoding and metatranscriptomic annotated data from surveys of terrestrial

systems (Bates et al. 2013; del Campo and Massana 2011; Geisen et al. 2015). Most of all, we still lack detailed information regarding the biogeography and distribution of phylotypes in soils. For most of these molecular surveys a huge rare biosphere is investigated and even sequences for flagellates previously thought to be geographically limited to specific limnic systems are discovered and unknown clades are left as “cryptic species” (Bates et al. 2013; del Campo and Massana 2011; Geisen et al. 2015). Using detailed phylogenetic techniques, some of these cryptic species may represent a closely related diversity, where possible gene-variants of similar species may fill the functional gaps in complex terrestrial ecosystems under different ecological conditions (Gossner et al. 2016; Soliveres et al. 2016; Venter et al. 2017). Even a few base pair differences between sequences for the same species may indicate a different ecotype/species (Glücksman et al. 2010).

In this study, we evaluated the recovery rate and genetic variance of HTS for ten cercozoan species previously described from the same sampling sites (Brabender et al. 2012). We hypothesize that the recovery of species described from these same sampling sites should be possible with high pairwise similarity identification. The recovery rate should be high for described species from the original site of discovery compared to sampling sites distant from the original point of discovery. This recovery rate should make it possible to describe the biogeographic dispersal and accuracy of the recovery of variants for the same morphospecies and how abundant the same species is in the mesoscale (150 sampling sites). Using the 18S rRNA gene as a HTS marker we expect many novel environmental sequences with large distances to cultured clones and a huge rare biosphere for flagellates previously thought to be geographically limited to specific habitats.

Results

The first aim was to test whether the HTS of all 150 sampling sites did recover the ten cercomonad species which were isolated and described from four grassland and from four nearby forest sites in the Hainich-Dün National Park (Brabender et al. 2012). All four grassland sites being identical to four of the metabarcoding sites. We intended to estimate how frequently the HTS unique individual reads (UIRs) for the ten cercomonads could be recovered from all 150 sites (Fig. 1). The studied species included three *Cercomonas* (*C. directa* [HFCC901], *C. pellucida* [HFCC903], *C. jendrali* [HFCC905]), two *Eocercomonas* (*E. perecta* [HFCC908], *E. exploratorii* [HFCC909]), three *Paracercomonas* (*P. proboscata* [HFCC913], *P. kuegeri* [HFCC915], *P. bassi* [HFCC918]), one *Nucleocercomonas* (*N. praelonga* [HFCC920]) and one *Metabolomonas* (*M. insania* [HFCC922]) species (see Brabender et al. 2012).

Seven of the ten species were recovered from the HTS data of the same sampling sites, where unique individual reads (UIRs) were identified by the first BLAST hit accession number in the Protist Ribosomal Reference (PR^2; Guillou et al. 2013) database (Fig. 1). Because the first BLAST hit may conceal equally valid bitscore results, we constructed a database consisting of the reference sequences for the ten reference species and performed a BLAST of all query UIRs sequences to them. BLAST results indicated that only a single base pair difference existed between UIRs (max length: 530bp) and the reference sequences for the three remaining species (*C. pellucida, C. directa, P. kuegeri*) (Fig. 1). Their sequence similarities (99.81% pairwise identity) were well within the p-distances to other congeneric species within *Cercomonas* (0.050; max. p-distance 0.183) and *Paracercomonas* (0.095; max. p-distance: 0.280; Table S1).

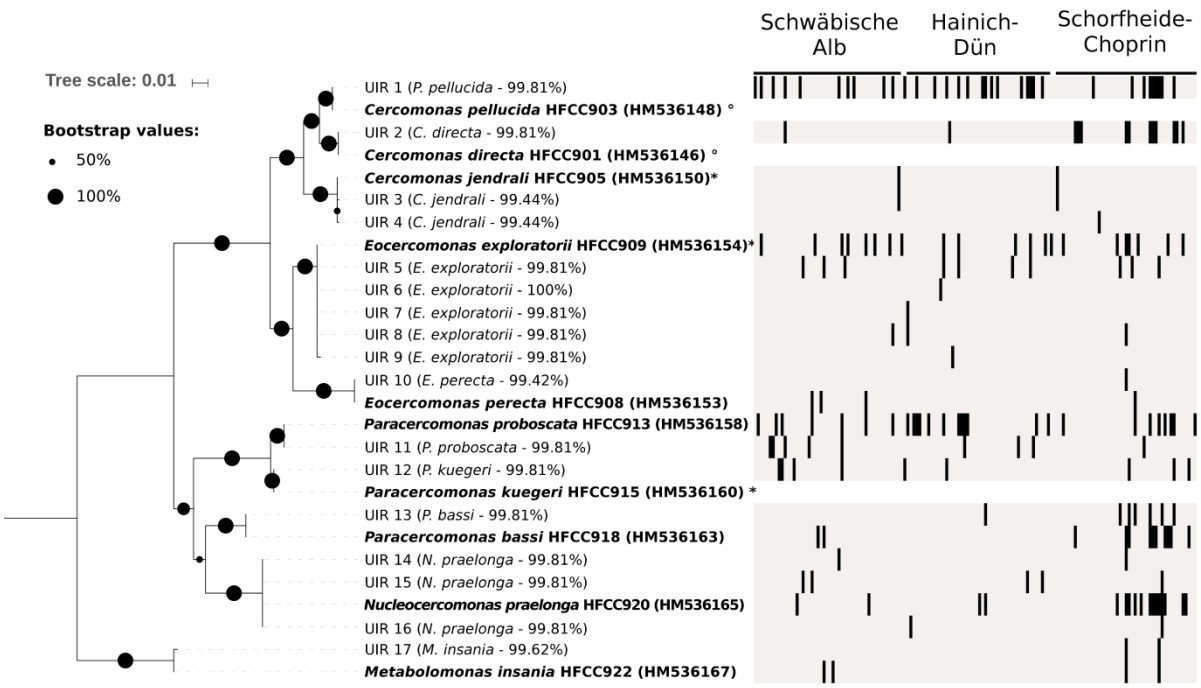


Fig. 1: Maximum likelihood (ML) phylogenetic tree showing the affiliation of the 18S rRNA gene Sanger sequences for the 10 described morphospecies (in bold) by Brabender et al. (2012) together with presence/absence distribution within the 150 grassland soil plots of the Biodiversity Exploratory sites (heatmaps). Circles (°) indicate organism not identified by first BLAST hit results of all unique individual reads (UIRs) against the Protist Ribosomal Reference (PR^2) database. Stars (*) indicate organisms initially identified in forest samples, where all others were isolated from grassland samples. Query UIRs were re-BLASTed against the 10 reference sequences isolated from Biodiversity sites in Hainich-Dün and their highest scoring hit sequences from HTS were added to the tree. When more than one high scoring hit was present above 99.7% pairwise identity, these were included. ML bootstraps, produced by 1000 replicates in RAxML, are indicated by the size of the black filled circles, all greater than 50%.

Recovery amongst other dominant cercozoans: The recovery rates for the ten isolated species were then evaluated for among all other recovered cercomonad UIRs in the 150 HTS study sites (Figs. 2 and S1). For this purpose UIRs together with PR^2 reference sequences were clustered into operational taxonomic units (OTUs) at a common and comparative threshold value (97% pairwise identity) and the OTUs were integrated into a tested backbone phylogenetic tree by Bass et al. (2009). Environmental-only OTU-clusters nested among known diversity (OTUs with accession numbers) with PR^2 database curated taxonomy, but were discovered at different biogeographic sites compared to the known-diversity clusters (e.g. *Cercomonas celer*, Fig. 2). Cluster sizes (number of UIRs) for environmental OTUs

were often larger than reference sequence-containing clusters, across similarity cut-offs for UIRs as determined by BLAST. The cercomonad backbone tree from Bass et al. (2009) containing the described and recovered species, retained its topology (see Brabender et al. 2012; Clade A and B), despite the addition of partial sequences (Figs. 2 and 1A). Monophyletic branches existed for *Cercomonas* and *Eocercomonas*, while *Nucleocercomonas* (including *Nucleocercomonas praelonga*) was paraphyletically associated with *Paracercomonas*, within the larger tree (Fig. S1; see Methods). Compared to other taxa identified, the ten isolated species occurred at fewer sites (e.g. *Cercomonas* sp. OTU 297; Fig. 2).

The dominant cercozoan taxon Sandonidae: One third (76 accession numbers) of the 288 reference sequences for *Sandona* in the PR^2 database were recovered in this study at >80% pairwise identity (Fig. 3). Assuming that distribution patterns (discovered location and number of occurrences in 150 sites) reflect differences in ecological adaptations, marker UIRs for the Sandonidae was discovered with high coverage and ubiquitous in grassland soils. Interesting was that dominant taxa were often OTUs with high pairwise identity distance to reference sequences and different biogeography to reference-taxa containing OTUs. For example, *Neoheteromita globosa* (U42447) was represented with a high coverage rate (83 of the 150 sites), but UIRs had low pairwise similarity to reference sequences (mean: 98.95%, max: 99.04%). This OTU phylogenetically associated closely with an environmental cluster (OTU 272) and an unresolved *Neoheteromita* sp. (AF411280), although occurrence patterns overlapped. In comparison to the *Neoheteromita globosa* OTU biogeography, the *Mollimonas vickermani* (HQ918173) cluster contained fewer UIRs with higher pairwise identity (mean: 99.43%, max: 99.81%), but the closest environmental cluster (OTU 85) was distantly related and its UIRs had a different distribution pattern (Fig. 3).

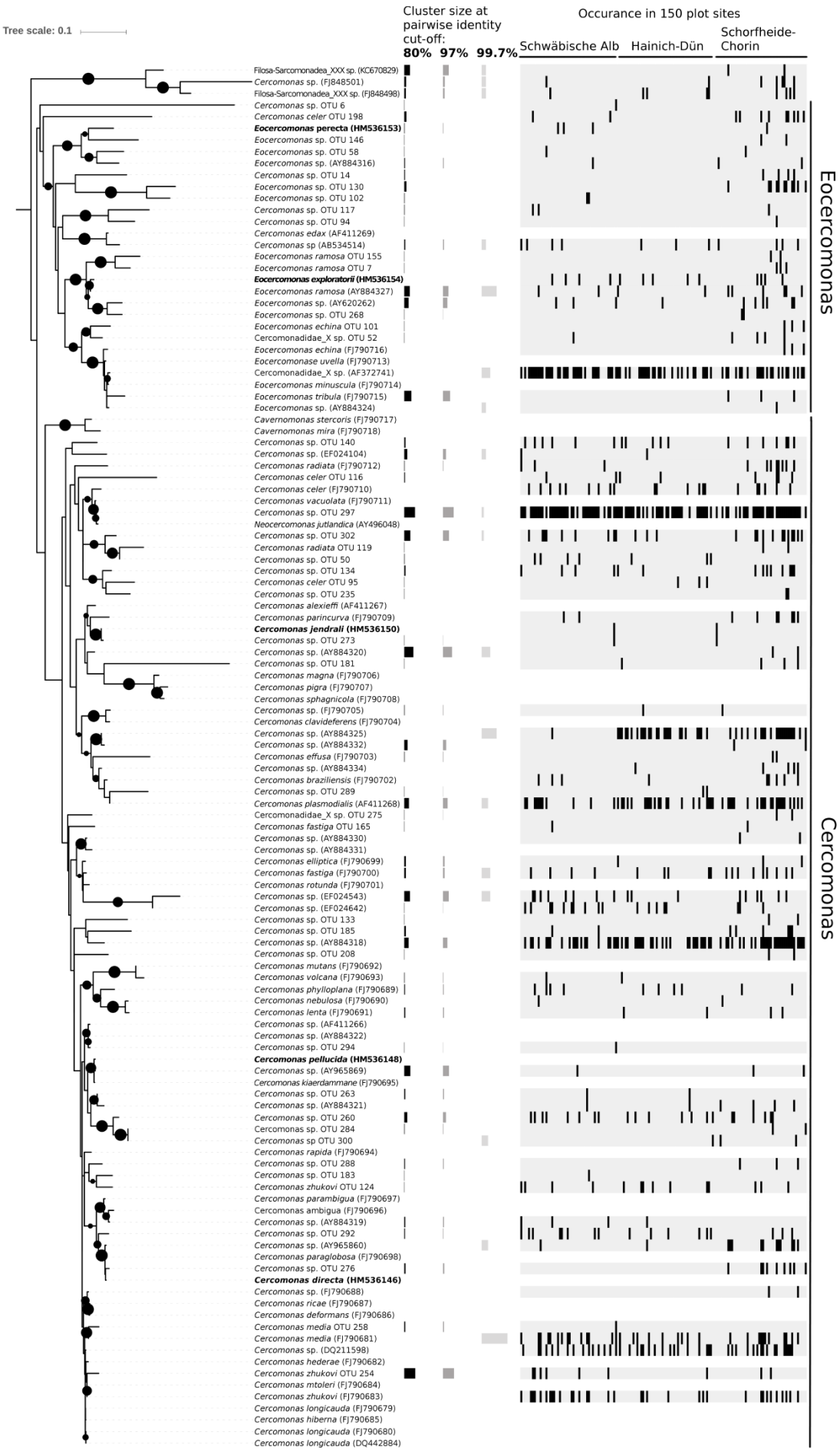

Tree scale: 0.1
Cluster size at pairwise identity cut-off:
80% 97% 99.7%
Occurance in 150 plot sites
Schwäbische Alb
Hainich-Dün
Schorfheide-Chorin
Eocercomonas
Cercomonas
Filosa-Sarcomonadea_XXX sp. (KC670829)
Cercomonas sp. (FJ848501)
Filosa-Sarcomonadea_XXX sp. (FJ848498)
Cercomonas sp. OTU 6
Cercomonas celer OTU 198
Eocercomonas perecta (HM536153)
Eocercomonas sp. OTU 146
Eocercomonas sp. OTU 58
Eocercomonas sp. (AY884316)
Cercomonas sp. OTU 14
Eocercomonas sp. OTU 130
Eocercomonas sp. OTU 102
Cercomonas sp. OTU 117
Cercomonas sp. OTU 94
Cercomonas edax (AF411269)
Cercomonas sp (AB534514)
Eocercomonas ramosa OTU 155
Eocercomonas ramosa OTU 7
Eocercomonas exploratorii (HM536154)
Eocercomonas ramosa (AY884327)
Eocercomonas sp. (AY620262)
Eocercomonas sp. OTU 268
Eocercomonas echina OTU 101
Cercomonadidae_X sp. OTU 52
Eocercomonas echina (FJ790716)
Eocercomonase uvella (FJ790713)
Cercomonadidae_X sp. (AF372741)
Eocercomonas minuscula (FJ790714)
Eocercomonas tribula (FJ790715)
Eocercomonas sp. (AY884324)
Cavernomonas stercoris (FJ790717)
Cavernomonas mira (FJ790718)
Cercomonas sp. OTU 140
Cercomonas sp. (EF024104)
Cercomonas radiata (FJ790712)
Cercomonas celer OTU 116
Cercomonas celer (FJ790710)
Cercomonas vacuolata (FJ790711)
Cercomonas sp. OTU 297
Neocercomonas jutlandica (AY496048)
Cercomonas sp. OTU 302
Cercomonas radiata OTU 119
Cercomonas sp. OTU 50
Cercomonas sp. OTU 134
Cercomonas celer OTU 95
Cercomonas sp. OTU 235
Cercomonas alexieffi (AF411267)
Cercomonas parincurva (FJ790709)
Cercomonas jendrali (HM536150)
Cercomonas sp. OTU 273
Cercomonas sp. (AY884320)
Cercomonas sp. OTU 181
Cercomonas magna (FJ790706)
Cercomonas pigra (FJ790707)
Cercomonas sphagnicola (FJ790708)
Cercomonas sp. (FJ790705)
Cercomonas clavideferens (FJ790704)
Cercomonas sp. (AY884325)
Cercomonas sp. (AY884332)
Cercomonas effusa (FJ790703)
Cercomonas sp. (AY884334)
Cercomonas braziliensis (FJ790702)
Cercomonas sp. OTU 289
Cercomonas plasmodialis (AF411268)
Cercomonadidae_X sp. OTU 275
Cercomonas fastiga OTU 165
Cercomonas sp. (AY884330)
Cercomonas sp. (AY884331)
Cercomonas elliptica (FJ790699)
Cercomonas fastiga (FJ790700)
Cercomonas rotunda (FJ790701)
Cercomonas sp. (EF024543)
Cercomonas sp. (EF024642)
Cercomonas sp. OTU 133
Cercomonas sp. OTU 185
Cercomonas sp. (AY884318)
Cercomonas sp. OTU 208
Cercomonas mutans (FJ790692)
Cercomonas volcana (FJ790693)
Cercomonas phylloplana (FJ790689)
Cercomonas nebulosa (FJ790690)
Cercomonas lenta (FJ790691)
Cercomonas sp. (AF411266)
Cercomonas sp. (AY884322)
Cercomonas sp. OTU 294
Cercomonas pellucida (HM536148)
Cercomonas sp. (AY965869)
Cercomonas kiaerdammane (FJ790695)
Cercomonas sp. OTU 263
Cercomonas sp. (AY884321)
Cercomonas sp. OTU 260
Cercomonas sp. OTU 284
Cercomonas sp OTU 300
Cercomonas rapida (FJ790694)
Cercomonas sp. OTU 288
Cercomonas sp. OTU 183
Cercomonas zhukovi OTU 124
Cercomonas parambigua (FJ790697)
Cercomonas ambigua (FJ790696)
Cercomonas sp. (AY884319)
Cercomonas sp. OTU 292
Cercomonas sp. (AY965860)
Cercomonas paraglobosa (FJ790698)
Cercomonas sp. OTU 276
Cercomonas directa (HM536146)
Cercomonas sp. (FJ790688)
Cercomonas ricae (FJ790687)
Cercomonas deformans (FJ790686)
Cercomonas media OTU 258
Cercomonas media (FJ790681)
Cercomonas sp. (DQ211598)
Cercomonas hederae (FJ790682)
Cercomonas zhukovi OTU 254
Cercomonas mtoleri (FJ790684)
Cercomonas zhukovi (FJ790683)
Cercomonas longicauda (FJ790679)
Cercomonas hiberna (FJ790685)
Cercomonas longicauda (FJ790680)
Cercomonas longicauda (DQ442884)

Fig. 2 (previous page): Maximum likelihood (ML) phylogenetic tree pruned to represent the *Cercomonas* and *Eocercomonas* clades indicating the affiliation of 266 full length reference library-derived GenBank (August 2017) and 330 partial environmental unique individual reads (UIRs) sequences for the V4 region of the 18S rRNA gene; 569 sequences; 2939 positions. The backbone tree (101 sequences) is derived from a *Cercomonas* and *Eucercomonas* gene library tree by Bass et al. (2009). Annotations for described sequences by Brabender et al. (2012) are in bold. ML bootstraps, produced by 1000 replicates in RAxML, are indicated by the size of the black filled circles, all greater than 50%. Horizontal barcharts at the tips of leaves in the ML tree indicate the cluster size at the relative pairwise identity cut-off levels (≥80%, ≥97% and ≥99.7%) included in operational taxonomic units (OTUs). Heatmaps associated with leaves in the ML tree indicate the presence/absence of UIRs centroids within the 150 Biodiversity Exploratory grassland sites.

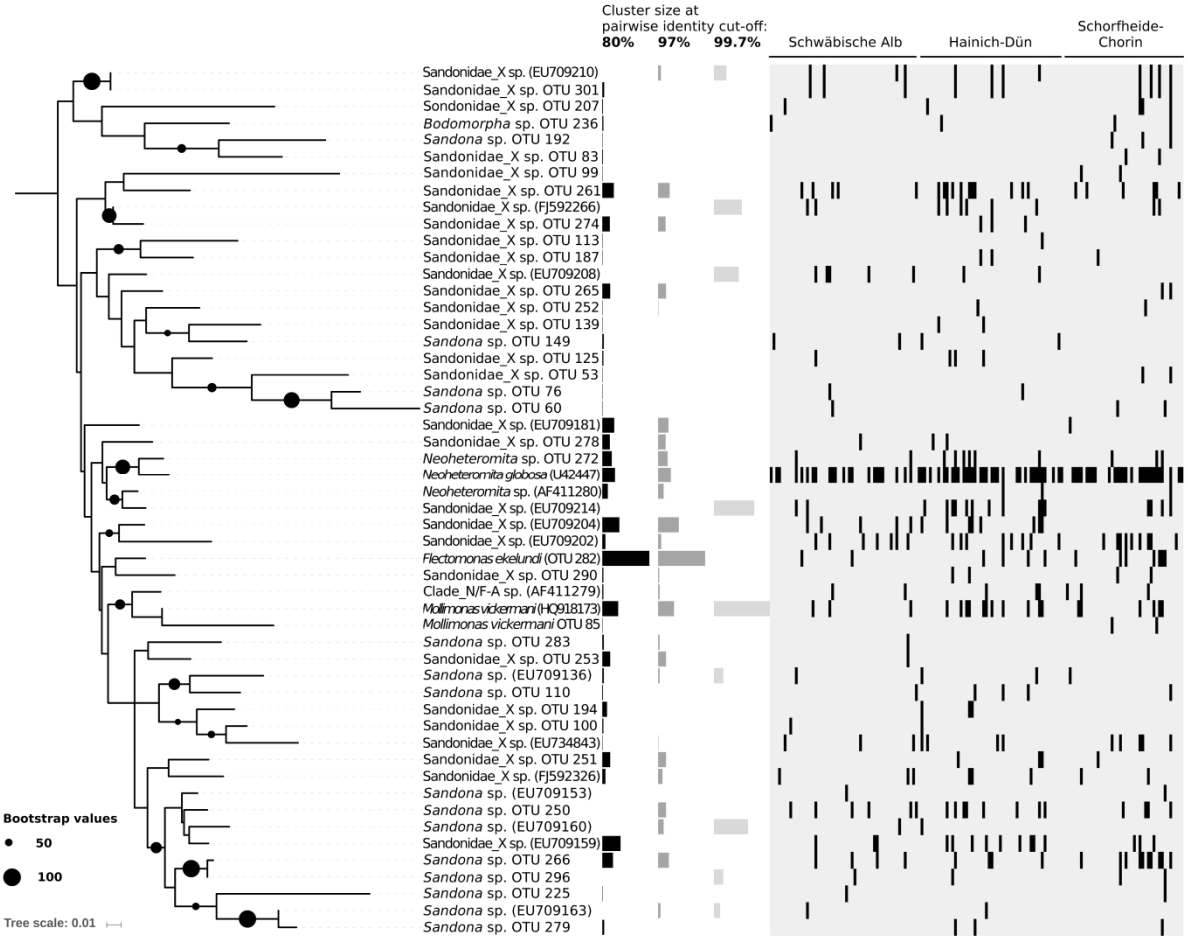


Fig. 3: Maximum likelihood (ML) phylogenetic tree pruned to represent the Sandonidae clade indicating the affiliation of 266 full length reference library-derived GenBank (August 2017) and 330 partial environmental unique individual reads (UIRs) sequences for the V4 region of the 18S rRNA gene; 569 sequences; 2939 positions. The backbone tree (101 sequences) is derived from a *Cercomonas* and *Eucercomonas* gene library tree by Bass et al. (2009). ML bootstraps, produced by 1000 replicates in RAxML, are indicated by the size of the black filled circles, all greater than 50%. Horizontal barcharts at the tips of leaves in the ML tree indicate the cluster size at the relative pairwise identity cut-off levels (≥80%, ≥97% and ≥99.7%) included in operational taxonomic units (OTUs). Heatmaps associated with leaves in the ML tree indicate the presence/absence of UIRs centroids within the 150 Biodiversity Exploratory grassland sites.

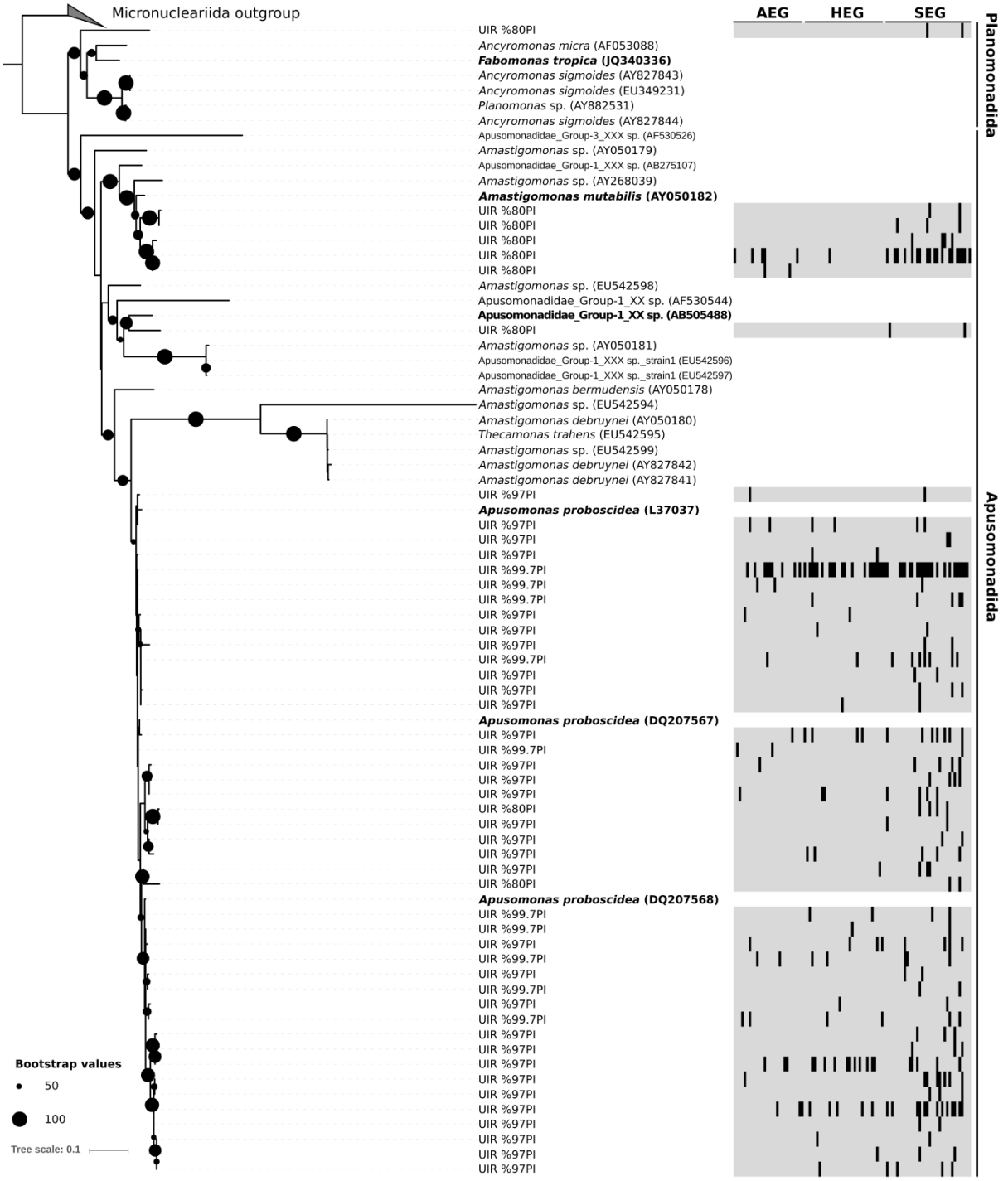


Fig. 4: Maximum likelihood (ML) phylogenetic tree representing the Apusomonadida (Protozoa: Apusozoa) indicating the affiliation of 28 full length reference library-derived GenBank (August 2017) and 50 partial environmental unique individual reads (UIRs) sequences for the V4 region of the 18S rRNA gene; 78 sequences; 2200 positions. The backbone tree (25 sequences) is derived from an apusozoan gene library tree by Cavalier-Smith and Chao (2010). ML bootstraps, produced by 1000 replicates in RAxML, are indicated by the size of the black filled circles, all greater than 50%. Heatmaps associated with leaves in the ML tree indicate the presence/absence of UIRs centroids within the 150 Biodiversity Exploratory grassland sites.

The rare cercozoan group Apusozoa: To further elucidate the distribution patterns of single UIRs clustered within OTUs, we selected another supergroup with fewer representation in our HTS dataset. Without clustering, we put all recovered UIRs in a phylogenetic relationship; including their first BLAST hit reference sequences (Fig. 4). Even when adding many environmental partial sequences to full length reference sequences, the published phylogenetic backbone tree by Cavalier-Smith and Chao (2010) retained its topography. Most UIRs clustered very clearly within shallow branches for three *Apusomonas proboscidea* (L37037, DQ207567 and DQ207568) reference sequences at the far edge of the tree (Fig. 4). Within these three groupings, some single UIRs indicated obvious ubiquitous distribution (max. 78 of 150 sampling sites) while having BLAST hit results with high pairwise sequence identity (≥99.7%) to reference sequences. Comparing geographic coverage of UIRs within phylogenetic groupings, some close clustering UIRs with lower pairwise identity to the BLAST reference sequences presented with an own individual biogeographic distribution. For example, within the *Amastigomonas* cluster, a distantly related group of five UIRs with BLAST sequence similarity below 97% clustered with *Amastigomonas mutabilis* (AY050182) could be a representative of an undiscovered species.

Choanoflagellates in soil. Analysing the barcoding V4 region of the SSU rRNA gene indicated that UIRs for these typical aquatic flagellates were present and widespread in the arable soil samples (69 of the 150 sites), but not ubiquitous (Fig. 5). Most UIRs occurred at lower altitude (10 – 140m) in the Schorfheide-Chorin (30 sites) compared to the Hainich-Dün (16 sites; elevation: 285 – 550m) or the Schwäbische Alb (23 sites; elevation: 462 – 858m). Interestingly, many of these UIRs were phylogenetically closely related to each other, forming large clusters, but many of these clusters could not be directly assigned to already sequenced species (Fig. 5;

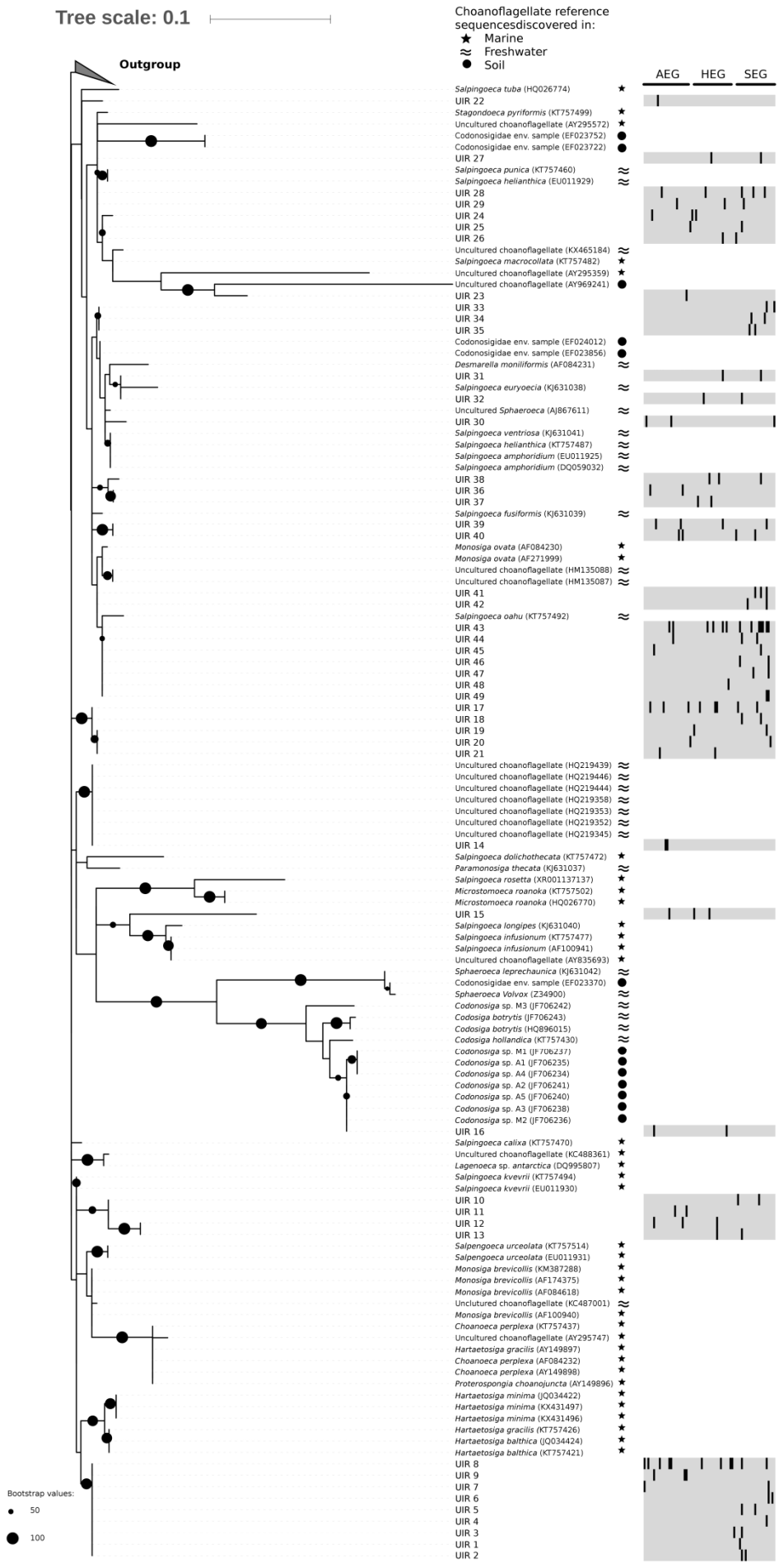

Tree scale: 0.1
Outgroup
Choanoflagellate reference sequencesdiscovered in:
Marine
Freshwater
Soil
AEG
HEG
SEG
Salpingoeca tuba (HQ026774)
UIR 22
Stagondoeca pyriformis (KT757499)
Uncultured choanoflagellate (AY295572)
Codonosigidae env. sample (EF023752)
Codonosigidae env. sample (EF023722)
UIR 27
Salpingoeca punica (KT757460)
Salpingoeca helianthica (EU011929)
UIR 28
UIR 29
UIR 24
UIR 25
UIR 26
Uncultured choanoflagellate (KX465184)
Salpingoeca macrocollata (KT757482)
Uncultured choanoflagellate (AY295359)
Uncultured choanoflagellate (AY969241)
UIR 23
UIR 33
UIR 34
UIR 35
Codonosigidae env. sample (EF024012)
Codonosigidae env. sample (EF023856)
Desmarella moniliformis (AF084231)
UIR 31
Salpingoeca euryoecia (KJ631038)
UIR 32
Uncultured Sphaeroeca (AJ867611)
UIR 30
Salpingoeca ventriosa (KJ631041)
Salpingoeca helianthica (KT757487)
Salpingoeca amphoridium (EU011925)
Salpingoeca amphoridium (DQ059032)
UIR 38
UIR 36
UIR 37
Salpingoeca fusiformis (KJ631039)
UIR 39
UIR 40
Monosiga ovata (AF084230)
Monosiga ovata (AF271999)
Uncultured choanoflagellate (HM135088)
Uncultured choanoflagellate (HM135087)
UIR 41
UIR 42
Salpingoeca oahu (KT757492)
UIR 43
UIR 44
UIR 45
UIR 46
UIR 47
UIR 48
UIR 49
UIR 17
UIR 18
UIR 19
UIR 20
UIR 21
Uncultured choanoflagellate (HQ219439)
Uncultured choanoflagellate (HQ219446)
Uncultured choanoflagellate (HQ219444)
Uncultured choanoflagellate (HQ219358)
Uncultured choanoflagellate (HQ219353)
Uncultured choanoflagellate (HQ219352)
Uncultured choanoflagellate (HQ219345)
UIR 14
Salpingoeca dolichothecata (KT757472)
Paramonosiga thecata (KJ631037)
Salpingoeca rosetta (XR001137137)
Microstomoeca roanoka (KT757502)
Microstomoeca roanoka (HQ026770)
UIR 15
Salpingoeca longipes (KJ631040)
Salpingoeca infusionum (KT757477)
Salpingoeca infusionum (AF100941)
Uncultured choanoflagellate (AY835693)
Sphaeroeca leprechaunica (KJ631042)
Codonosigidae env. sample (EF023370)
Sphaeroeca Volvox (Z34900)
Codonosiga sp. M3 (JF706242)
Codosiga botrytis (JF706243)
Codosiga botrytis (HQ896015)
Codosiga hollandica (KT757430)
Codonosiga sp. M1 (JF706237)
Codonosiga sp. A1 (JF706235)
Codonosiga sp. A4 (JF706234)
Codonosiga sp. A2 (JF706241)
Codonosiga sp. A5 (JF706240)
Codonosiga sp. A3 (JF706238)
Codonosiga sp. M2 (JF706236)
UIR 16
Salpingoeca calixa (KT757470)
Uncultured choanoflagellate (KC488361)
Lagenoeca sp. antarctica (DQ995807)
Salpingoeca kvevrii (KT757494)
Salpingoeca kvevrii (EU011930)
UIR 10
UIR 11
UIR 12
UIR 13
Salpengoeca urceolata (KT757514)
Salpengoeca urceolata (EU011931)
Monosiga brevicollis (KM387288)
Monosiga brevicollis (AF174375)
Monosiga brevicollis (AF084618)
Unclutured choanoflagellate (KC487001)
Monosiga brevicollis (AF100940)
Choanoeca perplexa (KT757437)
Uncultured choanoflagellate (AY295747)
Hartaetosiga gracilis (AY149897)
Choanoeca perplexa (AF084232)
Choanoeca perplexa (AY149898)
Proterospongia choanojuncta (AY149896)
Hartaetosiga minima (JQ034422)
Hartaetosiga minima (KX431497)
Hartaetosiga minima (KX431496)
Hartaetosiga gracilis (KT757426)
Hartaetosiga balthica (JQ034424)
Hartaetosiga balthica (KT757421)
UIR 8
UIR 9
UIR 7
UIR 6
UIR 5
UIR 4
UIR 3
UIR 1
UIR 2
Bootstrap values:
50
100

Fig. 5 (previous page): A choanoflagellate maximum likelihood (ML) phylogenetic tree indicating the affiliation of 80 full length reference library-derived GenBank (August 2017) and 50 partial environmental unique individual reads (UIRs) for the V4 region of the 18S rRNA gene; 445 positions. The backbone tree (101 sequences) is derived from a study by Carr et al. (2017). ML bootstraps, produced by 1000 replicates in RAxML, are indicated by the size of the black filled circles, all greater than 50%. All UIRs with hits greater than 80% pairwise identity were included. Heatmaps associated with leaves in the ML tree indicate the presence/absence of UIRs centroids within the 150 Biodiversity Exploratory grassland sites.

UIRs 1-9, 10-11, 12-13, 17-21, 43-49). They showed a sequence similarity to only 16 choanoflagellate reference sequences in the PR^2 database (mean 94%; min 83%; max 99%), indicating that soils are inhabited by a high number of yet unsequenced, if not undescribed choanoflagellate species. Some UIRs were considered spurious (at a single site), but in all cases the same UIR appeared at least 2 times. Among strict freshwater related UIRs and shallow branches, support values for associations were strong (Fig. 5; UIRs 17-21, UIRs 39-40 and UIRs 36-36). In addition, several soil choanoflagellate related UIR clusters were also identified (e.g. UIRs 33-35; Fig. 5).

Bicosoecids in soil. This HTS survey detected no less than 55 UIRs scattered throughout 76 of the 150 grassland sites, where most UIRs occurrences were in agriculturally non-fertilized pastures of the Schwäbische Alb (36 UIRs occurring at 10 sites; Fig. 6). Aligning these UIRs to reference sequences using BLAST, grouped all UIRs under 12 accession numbers of which only one was a described species (*Adriamonas peritocrescens,* AF243501). Putting these UIR partial sequences into a published phylogenetic tree for bicosoecids with other reference sequences (Cavalier-Smith and Chao 2006) returned monophyletic associations with strong support for the Bicoecales and Pseudodendromonadales where all UIRs clustered (Fig. 6). In other words, UIRs grouped within phylogenetic branches containing reference sequences previously isolated from freshwater and soil samples with high support values (>50% bootstrap values).

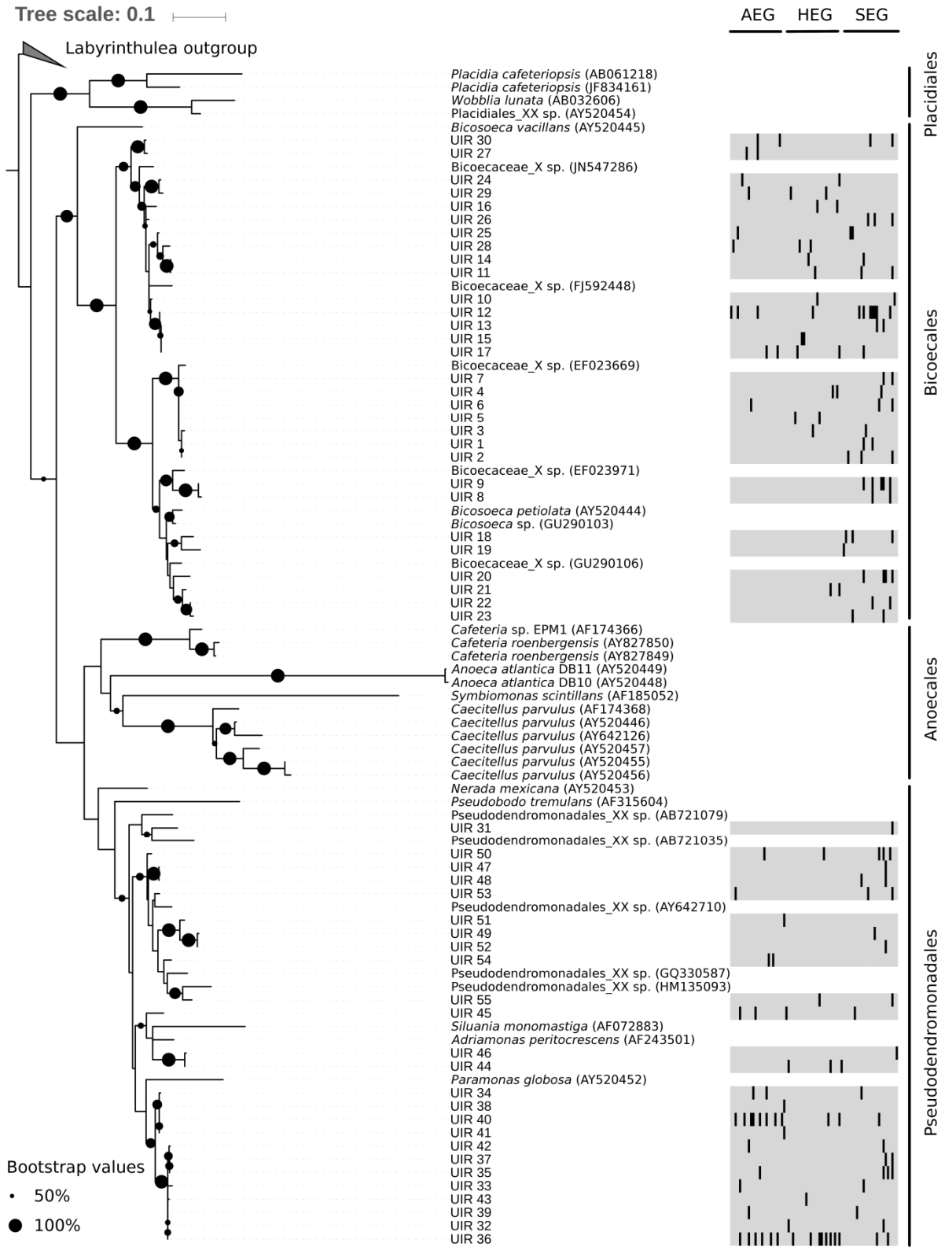


Fig. 6: A maximum likelihood (ML) phylogenetic tree of the bicosoecids indicating the affiliation of 36 full length reference library-derived GenBank (August 2017) and 55 partial environmental unique individual reads (UIRs) for the V4 region of the 18S rRNA gene; 1931 positions. The backbone tree (22 sequences) is derived from a study by Cavalier-Smith and Chao (2006). ML bootstraps, produced by 1000 replicates in RAxML, are indicated by the size of the black filled circles, all greater than 50%. All UIRs with hits greater than 80% pairwise identity were included. Heatmaps associated with leaves in the ML tree indicate the presence/absence of UIRs centroids within the 150 Biodiversity Exploratory grassland sites.

Discussion

Recovery rate. We have demonstrated that ten organisms, previously isolated and cultivated from soils in Hainich-Dün National Park and morphologically described by Brabender et al. (2012) could be recovered from environmental samples from the same sites. We could even indicate their recovery rate in 150 sampling sites across central Europe (Fig. 1). Flagellated protozoans are known to be the most dominant soil organisms next to ciliophorans (Bates et al. 2013; Domonell et al. 2013; Howe et al. 2009; Venter et al. 2017). Flagellate abundances in the same exploratory sites were previously determined to be about 28 x 10^3 ind./g dry soil weight, of which ~64% were cercozoans (Domonell et al. 2013).

We found that recovering the ten morphospecies in 150 grassland soil samples was highly dependent on pairwise-identity inclusion values (Fig. 1) and methods (see results). The first BLAST hits of all HTS reads against the entire PR^2 database only identified seven of the ten species. Only when UIRs were BLASTed to the ten reference sequences alone, was the presence of the others discovered. Secondly, when putting the recovery of the ten morphospecies into a phylogenetic context with all other cercozoans found in the 150 sampling sites, their distribution and UIR-richness in soil was among the rare species cases. If only one BLAST hit for a rare query sequence is returned, this reduces the chance of correctly identifying taxa, because of low representation. Sometimes only one UIR with low distribution would represent a single species, compared with other more dominant taxa and OTUs masking many UIRs (Figs. 1, 2 and S1). Finally, it became obvious that related UIRs, often with different sequence pairwise similarity values to reference sequences, also had complete different geographical distribution (e.g. *E. exploratorii*, Fig. 1). These may have been gene-variants or different ecotypes of the same

species or different species (Gill et al. 2015; Glücksman et al. 2010; Weisse et al. 2016). This experiment indicated an extreme case of how the hidden diversity and rare biosphere represented by species with low abundance in environmental samples could be undervalued.

The ten morphotypes in context with other cercozoans. Putting the 10 morpholgically described species from the exact same HTS sampling sites in context with all other HTS reads for the 150 sampling sites in phylogenetic trees underlined their position in the rare biosphere (Figs. 2 and S1). Clustering HTS unique individual reads (UIRs) into biologically meaningful OTUs indicated that their biogeographic distributions were sparse compared to other dominant OTUs (Figs. 1-4, S1). Despite this, HTS detected two of the ten morphologically described soil cercozoans originally described from forest plots (Brabender et al. 2012), but here recovered in grassland sites (e.g. *E. exploratorii*). This confirms the sensitivity of HTS compared to culture-dependent methods.

Cercozoans in general. Based on the number of UIRs incidences of the recovered ten species within all sites, we discovered that cercozoan communities were more dominant in the Schorfheide-Chorin (53 in Schwäbische Alb, 53 in Hainich-Dün, 90 in Schorfheide-Chorin, Fig. 1). Schorfheide-Chorin soils were expected to be more arid, because of less precipitation (520-600mm mean annual precipitation) than the Schwäbische Alb (938-963 mm mean annual precipitation(and the Hainich-Dün (500-800mm mean annual precipitation; Blüthgen et al. 2012). The ten described morphotypes were however wider distributed, than where the cercozoans which were first discovered by isolation and cultivation in the Hainich-Dün National Park (Brabender et al. 2012; Figs. 2, S1). Even though all ten morphospecies are capable of cyst formation, this geographic distribution in

temperate grasslands from central Europe is similar to HTS results for Cercozoa from other parts of the world (North- and South America and even the Antarctica; Bates et al. 2013).

With UIRs clustered into OTUs at individual pairwise identity cut-offs sometimes increased the appearance of a broader geographic distribution (Fig. 2) compared to UIRs alone (Fig. 1). OTUs with dominant occurrence at the 150 sampling sites were often found to be also UIRs with >3% distance to phylogenetically related reference sequences in the database (Figs. 1 to 3, S1). Some OTUs (clustering at 97% pairwise identity) included UIRs ranging in sequence pairwise distance from ≥99.7% to <97% to the closest reference sequence in the PR^2 database. Together, this information indicates a high hidden diversity, even within OTUs clustered at ≥97% pairwise identity across taxa. This most probably disregarded the possible differences in intraspecific sequence variation, where conspecific species were grouped into single OTUs (Caron et al. 2009). Results simultaneously indicated that a high (>50%) number of protist species are still not morphologically described, or maybe described but not sequenced (Foissner 2006). However, HTS reads were not so distantly related to sequences for uncultured organisms from soil compared to sequences originating from marine and freshwater studies (del Campo and Massana 2011).

Other dominant cercozoan taxa. The sandonids are a clade of gliding cercozoans, redescribed under the term glissomonads basically assigned to this group based on their 18S rRNA gene (Howe et al. 2009). This group is part of the free-living heterotrophic biflagellates and described to be dominant predators of bacteria in soil (Howe et al. 2009). In our study we detected one of the most common flagellated species in soil with high abundance: *Neoheteromita globosa* (Fig. 3). This

flagellate is known from soil for a long time, previously called *Heteromita globosa*, and is also very common in geographically distant soils, e.g. Australia (Ekelund and Patterson 1997). What we however discovered was that even for this species, most UIRs within the cluster was below 99.7% and even below 97% similar to the reference sequence (accession number U42447; Fig. 3). OTUs once again hide a large phylogenetic diversity and possible ecotypes (where UIRs have different biogeography), indicating that we probably found traces of a related species. Most clusters containing UIRs with higher sequence identity were associated with unresolved taxa (e.g. Sandonidae_X sp.). The only sandonid OTU with genus and species annotation, detected at higher pairwise identity levels (>97%) and clustering many UIRs with high distribution across all 150 sites, was *Mollimonas vickermani* (HQ918173). This species was also previously detected from Scottish grasslands (Howe et al. 2011) indicating that it most probably has a wide distribution in temperate soils. For the rest of the tree containing mostly unresolved taxa, we have to agree with findings of Howe et al. (2011), that most of their large diversity is still undescribed.

Examples of rare species being dominant in soil. To evaluate individual UIRs within OTUs, we examined a small group of protists – the apusozoans. For the 50 UIRs in our dataset, BLAST identified six of the 261 *Apusomonas* supergroup accession numbers in the PR^2 database. Representatives of three species groups (*Amastigomonas mutabilis*, *Apusomonas proboscidea* and *Fabomonas tropica*) and one unresolved apusomonadid were discovered. To elucidate this, UIRs from soil HTS sites were directly put into a phylogenetic context among other described SSU reference reads (Fig. 4). We uncovered at least three *Apusomonas proboscidea* and one environmental cluster for an *Amastigomonas* related species in grassland soil. This is noteworthy, because *Apusomonas* sequences from soil were not even

mentioned in some of the pioneering protist diversity studies based on HTS, where samples were taken from a wide variety of soils (Bates et al. 2013). This species has been known for a long time to be a typical soil flagellate, but less abundant and detected only sporadically in soil (Ekelund and Patterson 1997). One UIR, possibly a single gene variant of *Apusomonas proboscidea* (L37037), indeed seemed to be non-fastidious, spread highly ubiquitously (occurring at 51 of the 150 sites; Fig. 4). Put into the "high number of rare taxa in soil" perspective, this *Apusomonas* sp. in our study alone is extremely ubiquitous. HTS often presents with many rare taxa (Harder et al. 2016; Geisen et al. 2015), where >70% of OTUs are represented by <5 reads and OTUs are clustered when ≥97% sequence similarity exists between them (Bates et al. 2013). A deeper analysis of the three *Apusomonas proboscidea* (L37037, DQ207567, DQ207568) clusters indicated that the three strains are 97%, 96% and 94% sequence similar to one another. If clustered under the same name, this species would have occurred at 86 of the 150 sites. As for the five UIRs associated with *A. mutabilis* (AY050182), they possibly belong to an undescribed *Amastigomonas* species, because of the high distance (>97%). Because some UIRs were endemic to the grasslands of Schorfheide-Chorin, it may be that more than one ecotype is present within this cluster.

Apparently aquatic protists in soil (choanoflagellates and bicosoecids). Choanoflagellates are generally considered as bacteriovores, making up a considerable part of the pelagic communities in marine and freshwater systems (Arndt et al. 2000). However, several recent HTS studies have indicated the presence of choanoflagellates in soil (Bates et al. 2013; Geisen et al. 2015; Lara et al. 2011), and proof exists that this typical marine and freshwater group is also an autochthonous inhabitant of soil (Ekelund and Patterson 1997; Geisen et al. 2015). Unfortunately, above mentioned studies provide low resolution of the query

sequences and retrieved only few OTUs from soil. Furthermore, despite the PR^2 database (version) containing 486 different choanoflagellate sequences, it is very likely that many more species remain to be discovered in soil.

Very few UIRs associated with terrestrial reference sequences (5 UIRs; UIRs 16, 23, 33-35), while most recovered UIRs clustered with freshwater reference sequences (27 UIRs) with high support values (Fig. 5). As most craspedid species have not yet been studied regarding their salinity tolerance, any conclusion regarding restrictions or affiliations to a certain habitat parameter, e. g. salinity, would be highly speculative. Nevertheless, no UIRs related to acanthoecid species were found, which are predominantly restricted to marine and brackish water, with only a few exceptions (Nitsche 2014; Paul 2012). One important prerequisite for aquatic species to survive in soil is the ability to form resting stages to withstand unfavorable conditions like drought, which frequently occurs in soil. Up to now, no cyst formation has been described for acanthoecids (Leadbeater 2015 ; Nitsche 2014), while cyst formation for craspedids has been reported (Stoupin et al. 2012; Jeuck et al. 2014). Because only a single SSU gene was used for the analyses in this study, we did not attempt to elaborate on aspects of taxonomy or phylogeny. Some authors even caution the use of environmental sequences in phylogenetic context with reference sequences, because weak phylogenetic relationships between environmental freshwater and marine sequences can cause unstable or unresolved phylogenies, especially when using the SSU gene for choanoflagellates (Carr et al. 2017; Nitsche et al. 2011).

Despite the relatively short UIR read length, all choanoflagellate species detected, clustered clearly within the clade of Craspedida, supporting the monophyly of this group (Fig. 5). In general, various variants (sequences with only one basepair difference to the original sequences; Mahé et al. 2015) within a phylogenetic cluster

were detected to have a very specific distribution across the 150 sampling sites with little biogeographic overlap (Fig. 5). These microvariants are usually masked within OTUs, due to clustering methods. One example are the strain variants (UIR 43-49) related to *Salpingoeca oahu*, which show different distribution patterns.

Worth mentioning are the strain variants for *Codosiga botrytis*, previously discovered in ancient and modern soils from Morocco and Siberia (Stoupin et al. 2012) and here also discovered from two sampling sites in Central Europe (Fig. 5; UIR 16). *C. botrytis* is a morphospecies complex, including many cryptic species with the same morphology but with large genetic distances (Stoupin et al. 2012).

When comparing our results to a previous study, only one UIR (UIR 14) was >94% pairwise identical to a contig (contig 3) previously isolated by Geisen et al. (2015) from 12 mineral and organic soil samples from various vegetation and climatic zones (Fig. 5). This is probably related to the technological limitations resulting in a limited number of reads, which again does not allow drawing conclusions regarding general distribution patterns.

Regarding bicosoecids, only three of the 55 environmental bicosoecid UIRs could be related to a free-living soil flagellate species (*Adriamonas peritocrescens* [AF243501]), albeit with low pairwise identity (~94%; Fig. 6). The rest of the UIRs phylogenetically clustered with uncultured and unresolved taxa. Many of these reference sequences were isolated from sometimes very distant aquatic habitats (Lake Baikal [JN547286], Lake Tanganyika [GU290106, GU290103], Lac Pavin/France [AY642710], Praz Rodet/ Switzerland [GQ330587], Stechlinsee/Germany [HM135093]) and soil samples (Socompa Volcano/Puna de Atacama [FJ592448]). Some were even isolated from a purification plant (Japan [AB721035, AB721079]) as well as under increased CO_2 levels (EF023669,

EF023971). Bicosoecids, known as colorless marine and freshwater heterotrophs, are one of the least studied flagellate groups in soil. Along with apusomonads and choanoflagellates, they are counted to the rare species in soil (Griffiths et al. 2001). A previous study indicated that their presence and abundances (0.2 - 2.6 x 10^6 ind./g) in middle European grasslands are proof that these flagellates are not so rare in soil; in some instances accounting for up to 10% of the total cultivable flagellate abundance (Domonell 2013). We agree with this notion, but for those few cultivation based reference sequences in the database, studies employing HTS can contribute to understand *in-situ* diversity and dominance patterns in their natural environments.

In our study, low average pairwise identity (95.3%, min: 92.3, max: 98.8) existed between our bicosoecid environmental sequences and mostly uncultivated environmental sequences in the PR^2 database (Fig. 6). Bicosoecids are represented by 590 reference sequences in the PR^2 database, of which >434 are of uncertain taxonomic position or from uncultured environmental samples. This low correspondence between environmental sequences and the culture based sequences was already noticed for freshwater bicosoecids (del Campo and Massana 2011). Throughout this HTS study, we attempted to combine the semi-quantitative and biogeographic knowledge associated with HTS reads and the taxonomic knowledge of morphological studies. This situation is however paradoxical, in that one is literally using the unknown rare species diversity to detect and describe the diversity of an ever greater unknown diversity, to formulate an ecological understanding of the unknown.

Conclusions

Descriptions of new organisms should be based on morphology, genotypes (18S rRNA gene) and their ecology (Bass et al. 2009; Carr et al. 2017; Cavalier-Smith and Chao 2010; Howe et al. 2009; Howe et al. 2011). But, few studies evaluate the HTS recovery rate of the barcodes for organisms from the same site of discovery. Here, recovered the reference genes for flagellated organisms previously discovered in a study by Brabender et al. (2012) using a HTS dataset. Among commonly dominant (e.g. sandonids) and rare (e.g. apusozoans, choanoflagellates and bicosoecids) soil flagellates, as well as among the unresolved species sequences, a high unknown diversity was detected. For the dominant soil cercozoan flagellates (Cavalier-Smith and Chao 2010; Bates et al. 2013), we could indicate likelihood of semi-endemic geographic distribution (Foissner 2006), but only when critically dissecting UIRs from OTUs in a phylogenetic tree. This recovery rate by means of phylogeny could be used to describe the geographic dispersal and accuracy of recovery of gene variants for the same morphospecies and how abundant the same species is in the mesoscale (150 sampling sites). Surprising was the discovery of a large diversity of mainly aquatic taxa, although most UIRs formed clusters that associated close to other uncultured environmental sample reference sequences in the curated protist database (PR^2). Many novel environmental sequences had large distances to cultured clones and a huge rare biosphere for flagellates was revealed. Evaluating the UIRs instead of the OTUs that individual reads are usually clustered into, we discovered that much of the hidden diversity may be lost, when phylotypes are clustered at a single pairwise identity cut-off.

Methods

Data collection and soil sampling. We used a high-throughput sequencing (HTS) dataset for samples covering 150 grasslands from three different topo-geographic regions representing large parts (~44 000 km) of Central Europe (Germany, see Venter et al. 2017) in the mesoscale (samples 1 – 1000 km apart). These soil samples were collected in May 2011 as part of the German Biodiversity Exploratories initiative (http://www.biodiversity-exploratories.de/) (Fischer et al. 2010) encompassing three temporally and spatially scaled geo-referenced study plots: the UNESCO biosphere reserve Schorfheide-Chorin (SEG) in north-eastern Germany, Hainich-Dün National Park (HEG) in central Germany and the Schwäbische Alb UNESCO biosphere reserve (AEG) in south-western Germany. Standardized field sampling (Fischer et al. 2010; Brabender et al. 2012) and analyses is described by Venter et al. (2017). In short, 20 x 20 m size grassland plots were selected representing a range of land-use intensities (LUI). At each site, 14 soil sub-samples were cored out (diameter, 8.3 cm) from the upper most 10 cm of the A horizon and the top most 5 cm root-layer was removed, excluding particles >2 cm in diameter. Cores were combined, homogenized and composite samples stored at 4°C while still at field moisture content.

DNA isolation, PCR amplification and HTS. Whole genomic DNA was extracted from 1 g of each composite sample using the PowerSoil® DNA isolation kit (Mo Bio Laboratories, Inc., Carlsbad, CA). DNA concentration was measured using the Nanodrop 100 spectrophotometer (Thermo Fisher Scientific, Germany) and adjusted to 100 ng/µl using ddH_2O. The V4 region of the 18S rRNA gene was directly amplified from the samples using eukaryotic specific primers 590F (5'-3':CGGTAATTCCAGCTCCAATAGC) and 1300R (5'-3':CACCAACTAAGAACGGCCATGC). To separate the sequences, the Titanium

primer design and the recommended multiplex identifier (MID) adapter complex design (Roche, Germany) method was used. The pre-454 sequencing PCR reaction mixture (25 µl) contained: 2µl (100 ng/µl) DNA, 2 µl 10x DNA polymerase buffer with 20 mM $MgSO_4$, 2µl (1 µM) 590F primer + Adaptor A, 2µl (1 µM) 1300R primer, 2µl (2 mM) dNTP each and 0.4 µl (2.5 U/µl) Pfu DNA polymerase (Fermentas, Germany). Cycling conditions were: initial denaturation at 95°C for 3 min; 30 cycles at 95°C for 30 s, 55°C for 30 s, 72°C for 2 min; and a final extension at 72°C for 10 min. Each sample was amplified in triplicate and pooled to a final concentration of 20 ng/µl to eliminate possible PCR biases. HTS using the GS-FLX sequencer and Titanium sequencing kit XLR70 (Roche, Germany) was performed by GATC Biotech AG, Germany. Sequencing was done from the forward primer (adaptor A).

Bioinformatics. DNA sequences were demultiplexed and filtered for (1.) 100% forward primer match to remove false positive PCR amplifications of non-rRNA genes; (2.) minimum sequence length of 200bp and (3.) a maximum sequence length of 710bp - to remove possible sequencing artefacts; and (4.) ambiguities (N's), to exclude sequences containing uncertain base pairs. Sequences were scanned for chimeric sequences against the Protist Ribosomal (PR^2 v203) reference database (Guillou et a. 2013) using the uchime_ref algorithm in the USEARCH v. 7.0.1090 package (Edgar et al. 2011) and trimmed to a maximum length of 530bp to avoid terminally located read errors and focus downstream analyses on the V4 region of interest (Niklas et al. 2013). Trimmed reads were dereplicated into 100% identical unique individual reads (UIRs) using the VSEARCH script (Rochnes et al. 2016) to identify singletons and align sequences to the database.

Taxonomic analysis of sequencing data. UIRs with an initial abundance of one were termed singletons and removed to circumvent the dangers of pyrosequencing related artifacts (Tedersoo et al. 2010). UIRs were annotated by their

closest matching BLAST hit reference sequence (accession number) in the Protist Ribosomal Reference (PR^2) database (GenBank, version 203, downloaded June 2016) for 18S rDNA sequences (Guillou et al. 2013). This applied a closed-reference-binning (CRB), where the accession number was used to identify centroids, similar to that for operational taxonomic units (OTUs). UIRs with matching accession numbers were grouped into single CRBs. Because multiple individual UIRs could be linked to a single accession number, a more stabilized and accurate information on complex communities could be ascertained using CRBs. This compensated for pyrosequencing errors and inferred an upper (UIRs) and lower (CRBs) estimate of the actual species richness of protists in soil (see Venter et al. 2017). Default BLAST parameters (open gap penalty 5, cost gap extension penalty 2, nucleic match 2, nucleic penalty mismatch -3 and word size 11) were applied and single hits were retained if E-value $\leq e^{-100}$. Lineages for UIRs were inspected for ambiguous identification using the metagenome analyzer (MEGAN v. 5) program. Using 50 BLAST hits per query sequence, conserved sequences were correctly identified to the high-order taxa in the database, based on the lowest common ancestor (LCA) algorithm in MEGAN (Huson et al. 2007). UIRs with 100% coverage and annotations to taxa for which the primer region is not a suitable (Pawlowski et al. 2012) as well as non-protistan taxa (Metazoa, Fungi and Embryophyta) were removed. For the scope of this paper, Ciliophora and Amoebozoa annotated sequences were removed too.

OTUs vs. morphospecies richness. Determining species richness in environmental samples using HTS presented the added difficulty of determining species units (Caron et al. 2009). For this purpose UIRs were further binned based on pairwise identity cut-off values to determine species richness at taxonomic levels. Low pairwise identity values might have binned congeneric UIRs (Caron et al. 2009; Nebel et al. 2011) and too high values might have excluded too many UIRs. Due to

evolutionary rate differences within the V4 region of the SSU across taxa (Nebel et al. 2011) as well as sequencing and PCR error rates (Huse et al. 2007; Niklas et al. 2013; Stoeck et al. 2010), UIRs with single base differences (≥99.7% sequence similarity) were binned to the species taxonomic level. The commonly used ≥97% pairwise identity level could be used as a conservative proxy level to separate genera (Caron et al. 2009), but was used in this study to compare results with other studies. The ≥80% pairwise identity and last level for inclusion, stood as proxy for class taxonomic level (Stoeck et al. 2010). UIRs with <80% pairwise identity were excluded.

Phylogenetic analysis of sequencing data. Complementary to local pairwise alignment, we placed UIRs into a phylogenetic context with reference sequences for known diversity. For groups where many query sequences (UIRs) associated with a taxonomic group (Cercozoa), UIRs together with their hit reference sequences (CRBs) from the PR^2 v203 database were compiled into an input FastA file for clustering into classical OTUs using the global alignment USEARCH v9.0.2132 program (Edgar 2010). Using the *-cluster smallmem* and *-sortedby size* options, the USEARCH package algorithm matched all sequences to all seeds until all hits to a centroid were found at ≥97% sequence identity.

In all cases OTUs (centroid sequence) or UIRs were aligned to sequences for standardized reference trees of Cercomonadida (Bass et al. 2009), Choanoflagellida (Carr et al. 2017), Apusomonadida (Cavalier-Smith and Chao 2010) and Bicosoecidae (del Campo and Massana 2011). Multiple sequence alignment was performed using the MAFFT v7.123b (Guindon and Gascuel 2003) and MUSCLE v3.8.31 algorithm (Edgar 2004). Ambiguously aligned columns were corrected by hand. Maximum likelihood (ML) analyses were carried out in RAxML-HPC v7.2.8

using GTR + I + GAMMA model for distribution of evolutionary rates across sites and the following parameters: raxmlHPC -f a -x 12345 -p 12345.

Statistical analysis. Data for UIRs and their sampling sites were assembled into pivot-tables and converted to presence-absence matrices to avoid the effect of inflated abundances. Species richness could be derived from the number of OTUs/UIRs per site and within each region using the collective binary data at taxonomic levels. Statistical analysis was conducted in R software version 3.3.1 (Core Team R, 2014), using the Vegan 2.0-7 (Oksanen 2017) package. Correlations were performed using the “cor.test()” function in R using the Spearman rank test to analyse significance values.

Acknowledgements

We acknowledge bioinformatical and programming help by Dr. Peter Heger. The work was supported by grants from the German Research Foundation (DFG) to H.A. (grant numbers AR 288/16-1,2, SFB 2011, INST 216/862-1), a stipendium of the German Academic Exchange Service (DAAD) to P.V. (grant number 91525927). We thank the managers of the three Exploratories, Swen Renner, Sonja Gockel, Martin Gorke and all former managers for their work in maintaining the plot and project infrastructure; Simone Pfeiffer for giving support through the central office, Jens Nieschulze for managing the central database, and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project. The work has been partly funded by the DFG Priority Program 1374 "Infrastructure-Biodiversity-Exploratories". Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg (according to § 72 BbgNatSchG).

References

Arndt H, Dietrich D, Auer B, Cleven E, Gräfenhan T, Weitere M, Mylnikov AP (2000) Functional diversity of heterotrophic flagellates in aquatic ecosystems. In: Leadbeater BSC & Green JC (Eds) The Flagellates, (pp 240-268) Taylor & Francis Ltd, London

Bass D, Howe AT, Mylnikov AP, Vickerman K, Chao EE, Smallbone JE, Snell J, Cabral C Jr., Cavalier-Smith T (2009) Phylogeny and classification of Cercomonadida (Protozoa, Cercozoa): *Cercomonas*, *Eocercomonas*, *Paracercomonas*, and *Cavernomonas* gen. nov. Protist **160**:483-521

Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R, Fierer N (2013) Global biogeography of highly diverse protistan communities in soil. ISME J **7**:652-659

Blüthgen N, Dormann CF, Prati D, Klaus VH, Kleinebecker T, Hölzel N, Alt F, Boch S, Gockel S, Hemp A, Müller J, Nieschulze J, Renner SC, Schöning I, Schumacher U, Socher SA, Wells K, Birkhofer K, Buscot F, Oelmann Y, Rothenwöhrer C, Scherber C, Tscharntke T, Weiner CN, Fischer M, Kalko EKV, Linsenmair KE, Schulze E-D, Weisser WW (2012) A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. Basic Appl Ecol **13**:207–220

Brabender M, Kiss ÁK, Domonell A, Nitsche F, Arndt H (2012) Phylogenetic and Morphological Diversity of Novel Soil Cercomonad Species with a Description of Two New Genera (Nucleocercomonas and Metabolomonas). Protist **163**:495–528

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD, Dennett MR, Moran DM, Jones AC (2009) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. Appl Environ Microbiol **75**:5797–5808

Carr M, Richter DJ, Fozouni P, Smith TJ, Jeuck A, Leadbeater BSC, Nitsche F (2017) A six-gene phylogeny provides new insights into choanoflagellate evolution. Mol Phyl Evol **107**:166-178

Cavalier-Smith T, Chao EE (2010) Phylogeny and evolution of Apusomonadida (Protozoa: Apusozoa): new genera and species. Protist **161**:549-576

Core Team R (2013) R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna).

del Campo, Massana R (2011) Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. Protist **162**:435-448

del Campo J, Ruiz-Trillo I (2013) Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. Mol Biol Evol **30**:802-805

Domonell AK (2013) High resolution analysis of local and global protistan soil community structures using morphological and modern molecular techniques. Dissertation, Universität zu Köln, 155pp

Domonell A, Brabender M, Nitsche F, Bonkowski M, Arndt H (2013) Community structure of cultavable protists in different grassland and forest soils of Thuringia. Pedobiologia **56**:1-7

Dunthorn M, Otto J, Berger SA, Stamatakis A, Mahé F, Romac S, de Vargas C, Audic S, Stock A, Kauff F, Stoeck T, Edvardsen B, Massana R, Not F, Simon N, Zingone A (2014) Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. Mol Biol Evol **31**:993–1009

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**:1792–1797

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26**:2460–2461

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves

sensitivity and speed of chimera detection. Bioinformatics **27**:2194-2200

Ekelund F, Patterson DJ (1997) Some heterotrophic flagellates from a cultivated garden soil in Ausralia. Arch Protistenkd **148**:461-478

Ekelund F, Rønn R, Griffiths BS (2001) Quantitative estimation of flagellate community structure and diversity in soil samples. Protist **152**:301-314

Fischer M, Bossdorf O, Gockel S, Hänsel F, Hemp A, Hessenmöller D, Korte G, Nieschulze J, Pfeiffer S, Prati D, Renner S, Schöning I, Schumacher U, Wells K, Buscot F, Kalko EKV, Linsenmair KE, Schulze E-D, Weisser WW (2010) Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic Appl Ecol **11**:473–485

Foissner W (1999) Soil protozoa as bioindicators: pros and cons, methods, diversity, representative examples. Agricult Ecosys Environ **74**:95-112.

Foissner W (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta Protozoologica **45**:111–13

Geisen S, Bonkowski M (2017) Methodological advances to study the diversity of soil protists and their functioning in soil food webs. Appl Soil Ecol (*in press*)

Geisen S, Mitchell EAD, Wilkinson DM, Adl S, Bonkowski M, Brown MW, Fiore-Donno AM, Heger TJ, Jassey VEJ, Krashevska V, Lahr DJG, Marcisz K, Mulot M, Payne R, Singer D, Anderson OR, Charman DJ, Ekelund F, Griffiths BS, Rønn R, Smirnov A, Bass D, Belbahri L, Berney C, Blandenier Q, Chatzinotas A, Clarholm M, Dunthorn M, Feest A, Fernández LD, Foissner W, Fournier B, Gentekaki E, Hájek M, Helder J, Jousset A, Koller R, Kumar J, La Terza A, Lamentowicz M, Mazei Y, Santos SS, Seppey CVW, Spiegel FW, Walochnik J, Winding A, Lara E (2017) Soil protistology rebooted: 30 fundamental questions to start with. Soil Biol Biochem **111**:94-103

Geisen S, Tveit AT, Clark IM, Richter A, Svenning MM, Bonkowski M, Urich T (2015) Metatranscriptomic census of active protists in soils. ISME J **9**:1–13

Gill BA, Sandbarg JB, Kondratieff BC (2015) Evaluation oft he morphological species concepts of 16 western nearctic *Isoperla* species (Plecoptera: Perlodidae) and their respective species groups using DNA barcoding. Illiesia **11**:130-146

Glücksman E, Bell T, Griffiths RI, Bass D (2010) Closely related protist strains have different grazing impackts on natural bacterial communities. Environ Microbiol **12**:3105-3113

Gossner MM, Lewinsohn T, Kahl T, Grassein F, Boch S, Prati D, Birkhofer K, Renner SC, Sikorski J, Arndt H, Baumgartner V, Blaser S, Blüthgen N, Börschig C, Buscot F, Diekötter T, Jorge LR, Jung K, Keyel AC, Klein A-M, Klemmer S, Krauss J, Lange M, Müller J, Overmann J, Pašaliæ E, Penone C, Peroviæ D, Purschke O, Schall P, Socher SA, Sonnemann I, Tschapka M, Tscharntke T, Türke M, Venter PC, Weiner CN, Werner M, Wolters V, Wurst S, Westphal C, Wubet T, Fischer M, Weisser WW, Allan E (2016) Land-use intensification causes multitrophic homogenization of grassland communities. Nature **540**:266-269

Griffiths BS, Ritz K, Wheatley R, Kuan HL, Boag B, Christensen S, Ekelund F, Sørensen SJ, Muller S, Bloem J (2001) An examination of the biodiversity-ecosystem function relationship in arable soil microbial communities. Soil Biol Biochem **33**:1713-1722.

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet A-L, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R (2013) The Protist

Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res **41**:597–604

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52**:696–704

Harder CB, Rønn R, Brejnrod A, Bass D, Al-Soud WA, Ekelund F (2016) Local diversity of heathland Cercozoa explored by in-depth sequencing. ISME J **10**:2488-2497

Howe TA, Bass D, Vickerman K, Chao EE, Cavalier-Smith T (2009) Phylogeny, taxonomy, and astounding genetic diversity of Glissomonadida ord. nov., the dominant gliding zooflagellates in soil (Protozoa: Cercozoa). Protist **160**:159-189

Howe TA, Bass D, Chao EE, Cavalier-Smith T (2011) New genera, species, and improved phylogeny of Glissomonadida (Cercozoa). Protist **162**:710-722

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol **8**:R143

Huson D, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenome data. Gennome Res **17**:377–386

Jeuck A, Arndt H (2013) A short guide to common heterotrophic flagellates of freshwater habitats based on the morphology of living organisms. Protist **164**:842-860

Jeuck A, Arndt H, Nitsche F (2014) Extended phylogeny of the Craspedida (Choanomonada). Eur J Protistol **50**:430-443

Jeuck A, Nitsche F, Wylezich C, Wirth O, Bergfeld T, Brutscher F, Hennemann M, Monir S, Scherwaß A, Troll N, Arndt H (2017) A comparison of methods to analyze aquatic heterotrophic flagellates of different taxonomic groups. Protist **168**:375-391

Kadiri M, Spencer KL, Heppell CM (2012) Potential contaminant release from agricultural soil and dredged sediment following managed realignment. J Soils Sediments **12**:1581-1592

Lara E, Mitchell EAD, Moreira D, López Gracía P (2011) Highly diverse and seasonally dynamic protist community in a pristine peat bog. Protist **162**:14-32

Leadbeater BSC, Green JC (2000) The Flagellates: Unity, Diversity and Evolution. Taylor & Francis, London

Leadbeater BSC (2015) The Choanoflagellates: Evolution. Cambridge University Press, Cambridge, doi:10.1017/CBO9781139051125

Mahé F, de Vargas C, Bass D, Czech L, Stamatakis A, Lara E, Singer D, Mayor J, Bunge J, Sernaker S, Siemensmeyer T, Trautmann I, Romac S, Berney C, Kozlov A, Mitchell E, Seppey C, Egge E, Lentendu G, Wirth R, Trueba G, Dunthorn M (2017) Parasites dominate hyperdiverse soilprotist communities in Neotropical rainforests. Nature Ecol Evol **1**:0091

Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ **3**:e1420

Nebel M, Pfabel C, Stock A, Dunthorn M, Stoeck T (2011) Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. Environ Microbiol Rep **3**:154-158

Niklas N, Pröll J, Danzer M, Stabentheiner S, Hofer K, Gabriel C (2013) Routine performance and errors of 454 HLA exon sequencing in diagnostics. BMC Bioinformatics **14**:176

Nitsche F (2014) *Stephanoeca arndti* spec. nov. – First cultivation success including molecular and autecological data from a freshwater acanthoecid choanoflagellate from Samoa. Eur J Protistol **50**:412-421

Nitsche F, Arndt H (2015) Comparison of similar arctic and Antarctic morphotypes of heterotrophic protists regarding their genotypes and ecotypes. Protist **166**:42-57

Nitsche F, Carr M, Arndt H, Leadbeater BSC (2011) Higher level taxonomy and molecular phylogenetics of the Choanoflagellatea. J Euk Microbiol **58**:452-462

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hare RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2017) Package 'Vegan': Community ecology package. R package version 2.4-4. https://github.com/vegandevs/vegan/issues

Paul M (2012) *Acanthocorbis mongolica* nov. spec. – Description of the first freshwater loricate choanoflagellate (Acanthoecida) from a Mongolian lake. Eur J Protistol **48**:1-8

Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M, Fiore-Donno AM, Gile GH, Holzmann M, Jahn R, Jirkù M, Keeling PJ, Kostka M, Kudryavtsev A, Lara E, Lukeš J, Mann DG, Mitchell EAD, Nitsche F, Romeralo M, Saunders GW, Simpson AGB, Smirnov AV, Spouge JL, Stern RF, Stoeck T, Zimmermann J, Schindel D, de Vargas C (2012) CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol **10**:e1001419

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ **4**:e2409v1

Santoferrara LF, Grattepanche J-D, Katz LA, McManus GB (2014) Pyrosequencing for assessing diversity of eukaryotic microbes: analysis of data on marine planktonic ciliates and comparison with traditional methods. Environ Microbiol **16**:2752-2763

Schoenle A, Jeuck A, Nitsche F, Venter P, Prausse D, Arndt H (2016) Methodological studies on estimates of abundance and diversity of heterotrophic flagellates from the deep-sea floor. J Mar Sci Eng **4**:1-11

Soliveres S, Van der Plas F, Manning P, Prati D, Grossner MM, Renner SC, Alt F, Arndt H, Baumgartner V, Binkenstein J, Birkhofer K, Blaser S, Blüthgen N, Boch S, Böhm S, Börschig C, Buscot F, Diekötter T, Heinze J, Hölzel N, Jung K, Klaus VH, Kleinebecker T, Klemmer S, Krauss J, Lange M, Morris EK, Müller J, Oelmann Y, Overmann J, Pašaliæ E, Rillig MC, Schaefer HM, Schloter M, Schmitt B, Schöning I, Schrumpf M, Sikorski J, Socher SA, Solly EF, Sonnemann I, Sorkau E, Steckel J, Steffan-Dewenter I, Stempfhuber B, Tschapka M, Türke M, Venter PC, Weiner CN, Weisser WW, Werner M, Westphal C, Wilcke W, Wolters V, Wubet T, Wurst S, Fischer M, Allan E (2016) Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality.
Nature **536**:456–459

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, Richards TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol **19**:21–31

Stoupin D, Kiss AK, Arndt H, Shatilovich AV, Gilichinski DA, Nitsche F (2012) Cryptic diversity within the choanoflagellate morphospecies complex *Codosiga botrytis* – phylogeny and morphology of ancient and modern isolates. Eur J Protistol **48**:263-273

Tedersoo L, Nilsson RH, Aberenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G, Kõljalq U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. New Phytologist **188**:291–301

Venter PC, Nitsche F, Domonell A, Heger P, Arndt H (2017) The protistan microbiome of grassland soil: diversity in the mesoscale. Protist **168**:546-564

Weisse T (2002) The significance of inter- and intraspecific variation in bacterivorous and herbivorous protists. Antonie van Leeuwenhoek **81**:327-341

Weisse T, Anderson R, Arndt H, Calbet A, Hansen PJ, Montagnes DJ (2016) Functional ecology of aquatic phagotrophic protists - concepts, limitations, and perspectives. Europ J Protistol **55**:50–74

Supplementary Material

The hidden diversity of flagellated protists in soil

by Paul Christiaan Venter, Frank Nitsche and Hartmut Arndt.

Tables

Table S1: The congeneric variance within all *Cercomoas*, *Eocercomonas*, *Metabolomonas*, *Nucleocercomonas* and *Paracercomonas* reference sequences. The variance in p-distance (mean, minimum and maximum) was determined by including all sequences for the species within the genera from the Protist Ribosomal Reference (PR^2) database (GenBank, version 203, downloaded June 2016).

Genus	Number of sequences in PR^2 database	Mean p-distance	Min p-distance	Max p-distance	Sequence similarity inclusion cut-off
Cercomonas	137	0.050	0.000	0.183	95%
Eocercomonas	20	0.060	0.000	0.137	94%
Metabolomonas	1	0.000	0.000	0.000	100%
Nucleocercomonas	3	0.000	0.000	0.000	100%
Paracercomonas	77	0.095	0.000	0.280	91%

Table S2: The flagellate diversity within temperate central Europe. Taxa results after BLAST results for flagellate taxa returned when single hits for unique individual reads (UIRs) were annotated at the ≥97% sequence pairwise identity cut-off. UIRs were binned to the closest accession number (OTUs) and grouped within each Biodiversity Exploratory site (Schwäbische Alb - AEG, Hainich-Dün - HEG, Schorfheide-Chorin - SEG).

			Biodiversity Exploratory site		
Supergroup	**Class**	**OTUs**	**AEG**	**HEG**	**SEG**
Alveolata	Apicomplexa_X	15	126	158	104
	Dinophyceae	4	18	35	26
	Perkinsida	2	7	12	11
Apusozoa	Apusomonadidae_Group-1	3	29	52	52
Archaeplastida	Chlorophyceae	34	55	128	87
	Chlorophyta_X	2	1	17	1
	Pedinophyceae	1	0	0	1
	Trebouxiophyceae	46	168	414	140
	Ulvophyceae	5	13	44	12
Hacrobia	Centroheliozoa_X	3	2	2	4
	Telonemia_X	1	0	0	7
Opisthokonta	Choanoflagellatea	5	11	5	15
	Filasterea	1	6	0	1
	Ichthyosporea	5	67	39	36
Rhizaria	Cercozoa_X	4	41	41	47
	Endomyxa	5	53	59	63
	Endomyxa-Phytomyxea	15	65	106	74
	Filosa	2	1	3	7
	Filosa-Granofilosea	14	51	88	114
	Filosa-Imbricatea	19	173	201	225
	Filosa-Sarcomonadea	175	992	1342	1447
	Filosa-Thecofilosea	11	79	79	75
Stramenopiles	Bacillariophyta	5	8	27	10
	Bicoecea	3	2	6	15
	Chrysophyceae-Synurophyceae	56	339	408	627
	Eustigmatophyceae	1	4	16	7
	Hyphochytriomyceta	1	17	21	23
	MAST	4	17	7	26
	Oomycota	35	315	251	220
	Stramenopiles_X-Group-3	1	3	0	0
	Xanthophyceae	21	54	131	73
	Total OTUs:	499			

Table S3: The flagellate diversity within temperate central Europe. Taxa results after BLAST returned single hits were equated at the 99.7% sequence pairwise identity cut-off, binned to the closest accession number (OTUs) and grouped within each Biodiversity Exploratory site (Schwäbische Alb - AEG, Hainich-Dün - HEG, Schorfheide-Chorin - SEG).

				Biodiversity Exploratory site		
Supergroup	**Clas**	**Species**	**OTUs**	**AEG**	**HEG**	**SEG**
Alveolata	Apicomplexa_X	Gregarines_XX sp.	3	11	6	9
		Monocystis agilis	1	7	18	13
	Dinophyceae	*Gambierdiscus australes*	1	1	5	12
		Lepidodinium chlorophorum	1	1	0	0
Apusozoa	Apusomonadidae_Group-1	*Apusomonas proboscidea*	2	16	28	29
Archaeplastida	Chlorophyceae	*Bracteacoccus minor*	1	2	7	2
		Bracteacoccus sp.	4	5	15	7
		Chlamydomonas meslinii	1	0	2	0
		Chlamydomonas noctigama	1	0	0	2
		Chlamydomonas sp.	1	0	5	1
		Chlorococcum sp.	1	0	1	2
		Chlorococcum tatrense	1	0	1	2
		Chloromonas subdivisa	1	3	3	2
		Chlorosarcinopsis delicata	1	0	1	1
		Coelastrella oocystiformis	1	0	1	2
		CW-Chlamydomonadales_X sp.	4	3	5	10
		Cystomonas sp.	1	3	1	2
		Deasonia granata	1	0	0	2
		Desmodesmus sp.	1	0	0	3
		Dictyococcus varians	1	1	3	0
		Fasciculochloris boldii	1	2	0	2
		Heterochlamydomonas lobata	1	0	6	2
		Oocystis sp.	1	17	13	12
		Pseudomuriella engadinensis	1	2	5	0
		Tetracystis aplanospora	1	0	2	0
		Tetracystis pampae	1	0	2	0
		Tetracystis sp.	1	1	1	2
		Tetracystis vinatzeri	1	6	11	5
	Chlorophyta_X	*Scotinosphaera lemnae*	1	0	4	0
		Scotinosphaera sp.	1	1	10	1
	Pedinophyceae	*Pedinomonas minor*	1	0	0	1
	Trebouxiophyceae	*Apatococcus lobatus*	1	6	10	3
		Auxenochlorella protothecoides	1	11	16	11
		Auxenochlorella sp.	1	1	7	2
		Chlorella mirabilis	1	8	8	0
		Chlorella sp.	1	0	0	3
		Chlorella vulgaris	1	3	3	3
		Chloroidium ellipsoideum	1	21	23	10
		Coccomyxa sp.	1	0	2	0
		Diplosphaera sp.	2	4	13	4
		Elliptochloris sp.	1	2	0	0
		Elliptochloris subsphaerica	2	7	11	5
		Leptosira obovata	1	1	6	3
		Lobosphaera incisa	1	0	2	0
		Marvania sp.	1	8	36	4
		Muriella terrestris	1	11	35	7
		Myrmecia astigmatica	1	0	8	2
		Myrmecia sp.	1	1	5	1
		Nannochloris bacillaris	1	0	1	2
		Neocystis brevis	1	13	24	12
		Parietochloris alveolaris	1	0	0	5
		Prasiolales_XX sp.	2	1	5	5
		Pseudococcomyxa sp.	1	1	5	0
		Pseudostichococcus monallantoides	1	8	18	12
		Raphidonema nivale	1	2	5	3
		Stichococcus sp.	3	6	11	5
		Trebouxiophyceae_XXX sp.	1	1	4	0
		Trochisciopsis tetraspora	1	2	16	4
		Watanabea-Clade_XX sp.	2	10	18	6

Table S3 Continued

				Biodiversity Exploratory site		
Supergroup	**Clas**	**Species**	**OTUs**	**AEG**	**HEG**	**SEG**
		Watanabea reniformis	1	1	2	0
		Xylochloris sp.	1	1	7	1
	Ulvophyceae	*Planophila* sp.	1	10	13	3
		Trichosarcina polymorpha	1	0	4	0
		Ulothrix zonata	1	2	13	4
Hacrobia	Telonemia_X	Telonemia-Group-1_X sp.	1	0	0	5
Opisthokonta	Ichthyosporea	*Anurofeca* sp.	1	27	15	18
Rhizaria	Cercozoa_X	Cercozoa_XXXX sp.	1	8	3	6
	Endomyxa	Leptophryidae_X sp.	1	0	5	0
		Platyreta germanica	1	0	2	0
	Endomyxa-Phytomyxea	Plasmodiophorida_XX sp.	1	0	2	1
		Polymyxa graminis	1	13	20	14
	Filosa	Filosa_XXX sp.	1	0	1	4
	Filosa-Granofilosea	*Limnofila anglica*	1	2	3	3
		Limnofilidae_X sp.	2	4	11	11
		Novel-Gran-6_X sp.	1	1	10	8
	Filosa-Imbricatea	Spongomonadidae_X sp.	1	2	0	8
		Spongomonas solitaria	1	0	2	0
		Spongomonas sp.	1	2	0	3
	Filosa-Sarcomonadea	*Allantion* sp.	1	3	3	3
		Allapsa sp.	1	2	1	2
		Allapsidae_X sp.	1	6	31	18
		Bodomorpha sp.	2	3	9	6
		Cercomonadidae_X sp.	1	33	25	22
		Cercomonas elliptica	1	0	1	2
		Cercomonas fastiga	1	4	5	9
		Cercomonas media	1	14	9	11
		Cercomonas plasmodialis	1	11	20	22
		Cercomonas sp.	14	93	107	139
		Eocercomonas ramosa	1	2	2	6
		Eocercomonas sp.	2	4	5	5
		Filosa-Sarcomonadea_XXX sp.	2	1	4	6
		Flectomonas ekelundi	1	1	0	1
		Group_Te sp.	1	0	2	0
		Mollimonas vickermani	1	2	12	6
		Nucleocercomonas praelonga	1	3	3	2
		Paracercomonas ambulans	1	0	8	8
		Paracercomonas astra	1	8	2	9
		Paracercomonas compacta	1	18	26	25
		Paracercomonas elongata	1	2	1	1
		Paracercomonas minima	1	4	1	1
		Paracercomonas oxoniensis	1	0	0	2
		Paracercomonas paralaciniaegerens	1	1	0	4
		Paracercomonas sp.	5	33	25	39
		Sandona aporians	1	3	14	6
		Sandona sp.	12	16	28	26
		Sandonidae_X sp.	10	23	53	32
Stramenopiles	Bacillariophyta	Raphid-pennate_X sp.	1	0	2	0
	Chrysophyceae-Synurophyceae	Chrysophyceae-Synurophyceae_XXX sp.	2	12	14	26
		Clade-C_X sp.	18	128	161	207
		Clade-D_X sp.	1	2	0	2
		Clade-H_X sp.	1	0	0	3
		Pedospumella encystans	1	22	23	22
		Spumella sp.	4	46	43	43
		Spumella sp._strain1	1	0	9	13
		Spumella vulgaris	1	38	43	42

Table S3 Continued

				Biodiversity Exploratory site		
Supergroup	**Class**	**Species**	**OTUs**	**AEG**	**HEG**	**SEG**
	Eustigmatophyceae	*Eustigmatos vischeri*	1	4	15	6
	Hyphochytriomyceta	*Hyphochytrium catenoides*	1	14	16	21
	Oomycota	*Haptoglossa* sp.	2	17	12	18
		Lagenidium sp.	1	0	0	1
		Phytophthora drechsleri	1	3	0	4
		Phytophthora iranica	1	2	0	0
		Phytophthora megasperma	1	0	0	2
		Phytophthora sp.	1	19	1	0
		Pythiopsis cymosa	1	5	0	0
		Pythium attrantheridium	1	37	2	23
		Pythium glomeratum	1	25	2	25
		Pythium macrosporum	1	2	0	1
		Pythium monospermum	1	1	2	3
		Pythium rostratifingens	1	4	0	5
		Pythium takayamanum	1	17	0	1
		Pythium viniferum	1	7	0	7
		Pythium volutum	1	5	0	7
	Xanthophyceae	*Botrydiopsis callosa*	1	0	1	3
		Botryochloris sp.	1	0	2	1
		Chlorellidium pyrenoidosum	1	0	1	1
		Chlorellidium tetrabotrys	1	8	16	7
		Heterococcus caespitosus	1	4	12	0
		Heterococcus pleurococcoides	1	8	15	9
		Xanthonema tribonematoides	1	0	1	2
		Xanthophyceae_XXX sp.	1	3	3	0
		Total OTUs:	218			

Supplementary figure

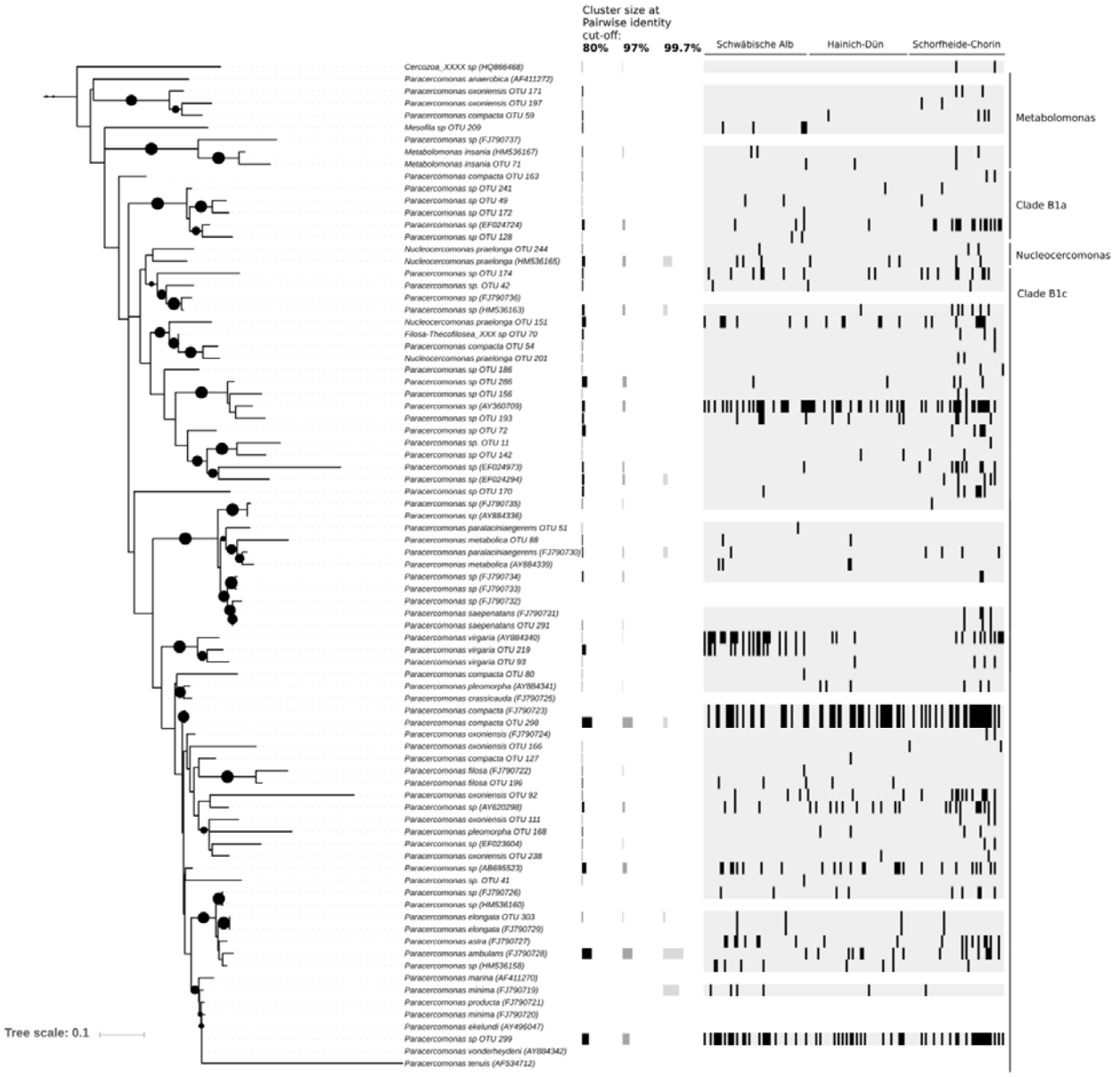


Fig. S1: Maximum likelihood (ML) phylogenetic tree pruned to represent the *Metabolomonas*, *Nucleomonas* and *Paracercomonas* clades indicating the affiliation of 266 full length reference library-derived GenBank (August 2017) and 330 partial environmental unique individual reads (UIRs) sequences for the V4 region of the 18S rRNA gene; 569 sequences; 2939 positions. The backbone tree (101 sequences) is derived from a *Cercomonas* and *Eucercomonas* gene library tree by Bass et al. (2009). ML bootstraps, produced by 1000 replicates in RAxML, are indicated by the size of the black filled circles, all greater than 50%. Horizontal barcharts at the tips of leaves in the ML tree indicate the cluster size at the relative pairwise identity cut-off levels (≥80%, ≥97% and ≥99.7%) included in centroids. Heatmaps associated with leaves in the ML tree indicate the presence/absence of UIRs centroids within the 150 Biodiversity Exploratory grassland sites.

Conclusive Summary

The general aim of this dissertation was to increase the understanding of protist diversity and community structure in temperate grassland soils of Germany within a mesoscale using High-Throughput Sequencing (HTS). Community structure was analyzed along gradients (e.g. geographic separation and land-use intensity [LUI] index) and species role-players identified based on their 18S rRNA gene barcode association with the closest protist reference sequence in the curated Protist Ribosomal Reference (PR^2) database.

We developed a unique pipeline to determine ecological patterns and protist community structure and discovered that the interpretation of HTS results were analysis-biased regarding protist richness and diversity (**Chapter 1**). Results further support the expectation that low database representation led to result biases (e.g. Hacrobia, Excavata, Apusozoa), because even unexpected, mostly aquatic taxa (e.g. choanoflagellates, MAST-12C), could be discovered across all of the mesoscale grasslands. Typical soil protists (~35% rhizarians, ~28% alveolates, ~18% stramenopiles) were uncovered, although most data consisted of rare taxa with low abundance and undersampling could not be excluded using HTS and a large number of samples. By extension, the hypothesis in support of a dominant alveolate rather than rhizarian presence in soils could be confirmed.

Furthermore, results depended on whether pyrotags (also called Unique Individual Reads - UIRs) were directly analyzed or clustered to a database close centroid sequence (closed-reference-clustering) and the level of sequence similarity inclusion value. Due to the method used we often refered to the term pyrotag and UIRs interchangeably to optimally explain difficult concepts. Most sequences were database close in that only 11 OTUs were excluded at the <80% pairwise identity cut-off and only 11% of all UIRs could not be

matched to a reference sequence over the full length of the query sequence (100% query coverage). On the other hand, we could confirm our hypothesis that a large unknown diversity was present, because less than 1% of the UIRs could be matched with 100% to a protist reference sequence. Unknown diversity (low pairwise identity) as well as unresolved and uncultured environmental reference sequences were ever present (20%), in which case affiliations to known taxonomic lineages was only possible by phylogenetic means.

Contrary to what we hypothesized, mesoscale separation indicated more similar communities than local plots, and a taxa-area relationship typical to land plants and animals (slope of the relationship *z*-value: 0.1 – 0.3) and not as that of other microorganism like bacteria and fungi (*z*-value: 0.02 – 0.08) was observed. This result confirmed our hypothesis that geographic separation can lead to semi-endemicity and challenges the ubiquitous-distribution-due-to-small-body-size hypothesis (Foissner 2006). This was confirmed by some organisms being exclusively and highly present in some biodiversity exploratory sites only and others being present across continental scales, while no ubiquitous UIR or OTU could be identified. Geographic separation and not LUI indicated more overlap in similar species. The disrupting effect of land-use (fertilization, mowing, livestock grazing) influenced biodiversity differently to what was hypothesized, in that diversity increased with increasing LUI and became less homogeneous. The increasing non-overlap between biodiversity exploratory sites raised the suspicion that rare taxa may provide a multifunctional response mechanism to the detrimental effects of increased LUI.

Protists (bacterivores) were included alongside eleven other above- and belowground trophic groups, to study the multitrophic diversity and community structural changes to land-use (**Chapter 2**). Because even study sites associated with small increases in LUI indicated homogenization of both above- and belowground communities, the first hypothesis could be confirmed – the point of change being medium LUI. We could also confirm that changes in local community diversity (α-diversity) did not seem to influence β-diversity (community

dissimilarity between sites), and that the effect of land-use was different for below- and aboveground communities. Aboveground communities were more sensitive (better indicators) to changes in land use, where belowground trophic groups responded slower.

Unexpectedly, the α-diversity increased for belowground communities (including protists) as opposed to aboveground communities with increasing LUI. As hypothetically expected, the β-diversity between trophic groups decreased at higher LUI, most probably due to a loss of specialist species, which may have been caused by homogenization, suggesting changes in the trophic structure. The mechanism seems however complicated, because independent of the aboveground communities, species composition belowground was homogenized, reducing β-diversity, without reducing α-diversity. The opposite effects of LUI on belowground compared to aboveground community structure did not conform to the initial hypothesis, and was altogether new to multitrophic community understanding. In short, we confirm our hypothesis that β-diversity was reduced in the multitrophic communities, but belowground α-diversity was less affected by LUI as compared to aboveground communities.

Compared to common species, rare species with low abundance more effectively increased biodiversity and so promoted multifunctionality (**Chapter 3**). Where increased land use led to decreased diversity and therefore to decreased ecosystem functioning, especially common species diversity declined – confirming the importance of the diversity of common species. Common species were however only associated with intermediate (not highly) multifunctional capacity, due to their abundance-related effects. Because increased levels of LUI led to a decline in species abundance, we discovered that the rare species must be responsible for the increased capacity for delivering multiple ecosystem functions (multifunctionality), contradictory to our first hypothesis. We speculate that rare species are less redundant in their functional traits than common species and therefore become functionally more important to sustain multiple ecosystem functions. In contrast, the second hypothesis could be confirmed, which stated that multifunctionality driven by high diversity

will decrease with increased LUI due to functional compositional changes. But this was only true for aboveground trophic groups, since belowground trophic groups (including protists) were less influenced by anthropogenic disturbances. The third hypothesis which states that aboveground organism diversity is a stronger predictor of ecosystem multifunctionality, could therefore be confirmed. More specifically, the aboveground rare species were often related to the highest levels of multifunctionality, but also extremely sensitive to anthropogenic disturbances, which made them important indicator species for high multifunctionality. Conclusively, this last statement also agreed with our final hypothesis, where 10% of the species tested could function as indicator species in natural ecosystems, due to their individual influence on the general ecosystem and providing multiple functions across the 14 investigated ecosystem functions. These indicator species were present across multiple trophic groups and signal LUI changes by their abundance changes.

When regarding all trophic species diversity contributing to multifunctionality, and not just single trophic groups, then filling the gaps in ecosystem services becomes a concerted effect (**Chapter 4**). A study of the relationships between species richness and abundance across nine trophic groups indicated that a multitrophic richness, rather than any single trophic group, had a stronger impact on altogether 14 ecosystem services. This relationship remained the same, even when the diversity was negatively affected by increased land-use. The differences in variance explained by multitrophic richness and abundance, compared to LUI and environmental services indicated that high species richness in multiple trophic groups could best explain high level ecosystem functioning. This was however not the case for abundance, since the abundance of particular trophic groups can negatively affect the ecosystem functioning of others. Variance in ecosystem functioning was more robustly explained by multitrophic richness and abundance, than by environmental abiotic conditions and increased LUI. Therefore, diversity is an extremely strong support mechanism for ecosystem stability and conservation of ecosystem services. Bacterivorous protists were

responsible for positive variance in multitrophic richness and abundance on grassland functioning, especially regarding changes in provisioning services and regulating services. Negative variance with regards to supporting services may be due to the fact that belowground organisms respond slowly to aboveground land-use impacts. In agreement with the previous study, this study emphasizes the importance of aboveground trophic groups to maintain ecosystem functioning.

In order to study the biodiversity of the small protistan life forms, a combination of techniques is required (**Chapter 5**). This becomes extremely evident when the aim is to gain high resolution quantitative and qualitative information. Ultimately, a combination of culture-independent (molecular next generation sequencing - NGS) and culture-dependent methods is required. For surveys covering large samples numbers in a mesoscale size sampling area, such as the present study, DNA-barcoding recovering genotype fingerprints of protists present in complex soil samples, becomes a more viable option. Some unresolved issues with regard to NGS make quantitative analyses less accurate, therefore the focus during analyses was drawn exclusively to presence-absence data. In such instances, a combination of two molecular techniques or multiple markers may become useful in elucidating the shortcomings of the other.

One of the issues, when using HTS for high resolution qualitative information became more visible when specific dominant ciliate protists, known to be well studied in central European soil, were analyzed (**Chapter 6**). The issue of discrepancies between morphological data information and the reference sequences for lineages in the PR^2 database was analyzed for the first time, and gave an overview of the magnitude of the problem. First, only a third of the morphologically described species had a reference sequence deposited in the molecular database, for the most studied ciliate taxon – the Colpodea. Ciliate flagship species were not detected in the magnitude as was expected. Only a third of the discovered morphospecies species in similar European temperate grasslands could be recovered, as was expected by

leading researchers in the field (e.g. Chao et al. 2006; Foissner et al. 2008). Phylogenetic relationships for this monophyletic taxon remaind monophyletic even when pyrotags were included with full length reference sequences in phylogenetic analyses. What was not expected was the very few (~1%) exact matches of pyrotags to morphologically described flagship ciliates from this region, therefore our first hypothesis could not be accepted. Eight of the ten most abundant pyrotags rather associated with other uncertain environmental sequences discovered in very distant sites (e.g. Greenland marine Arctic sediments) and even more unknown lineages (at <97% pairwise identities to reference sequences) were uncovered. Phylogenetic resolution of the pyrotags revealed that well described taxa (*Colpoda steinii* and *Colpoda inflata*) and even their gene variants could be differentiated. Even though the richness of species discovered (roughly 175 taxa) was similar to that in other soil studies (Esteban et al. 2006; Acosta-Mercado & Lynn 2004), a three to eight times higher microdiversity for uncovered taxa could be discovered than described taxa matches in the PR^2 database. These results made detecting ecological gradients for taxonomic groups difficult. Even though strong ecological gradients were present (e.g. soil water content and soil type) less conclusive results were obtained, especially for the many taxa less well represented by the PR^2 database (e.g. Heterotrichea, Prostomatea and Nassophorea).

To prove how accurate our discovery of the heterotrophic flagellate "dark unknown diversity" in soil is, we had to first prove the recovery rate of morphologically described species, isolated and cultured from the same grassland sites (**Chapter 7**, first part). Knowing that gaps exist between molecular database reference sequences and that most HTS retrieved pyrotags are from the rare biosphere, we accepted the challenge of accurate diversity reporting. Here, we proved for the first time the recovery rate for ten cercomonad species hidden between the rare species with low abundance in environmental samples. Putting the pyrotags for these cercomonad species in a phylogenetic tree with other cercozoans, indicated biogeographic variance among the UIRs associated with single species. Biogeographic

variance would have been masked in OTUs by clustering and could have concealed this hidden diversity within OTUs. A further large hidden diversity was uncovered, where UIRs associated with unresolved, unknown environmental sequences and some even forming new clades in phylogenetic trees. We accepted the hypothesis that species isolated at the site of discovery would have a higher recovery rate than those from further away, because the recoverable species are usually among the rare taxa and their recovery is method specific. Only when the resolution or method is set, is it possible to determine their distribution in the mesoscale.

The robust method, however, made it possible to not only discover typical grassland and soil species with high distribution (*Mollimonas vickermani*, *Neoheteromita globosa*), but also the rare taxa (*Apusomonas proboscidea*), all probably freeliving heterotrophic flagellates (**Chapter 7**, second part). Deep phylogeny with high support values were not possible for typical abundant (the cercomonad group *Sandona*) and rare taxa (apusozoans, choanoflagellates) in soil, hence no ecological associations could be deduced based on phylogenetic relationships and their recovery in soils (e.g. occasions of gene transfer or the occurrence of typical marine clades). Unexpectedly, some typical aquatic flagellates, rarely discovered in terrestrial habitats were also uncovered (choanoflagellates), as well as some taxa with little ecological and morphological investigation regarding their occurrence in soil – the bicosoecids. Collectively, the known diversity makes up only a small part of the large unknown diversity.

General references

Acosta-Mercado D, Lynn DH (2004) Soil ciliate species richness and abundance associated with the rhizosphere of different subtropical plant species. J Eukar Microbiol **51**:582–588

Adl SM, Simpson AGB, Lane CE, Lukes J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EAD, Mozley-Stanridge SE, Parfrey LW, Pawlowski J,

Rueckert S, Shadwick L, Schoch CL, Smirnov A, Spiegel FW (2012) The revised classification of eukaryotes. J Eukar Microbiol **59**:429–493

Allan E, Bossdorf O, Dormann CF, Prati D, Gossner MM, Tscharntke T, Blüthgen N, Bellach M, Birkhofer K, Boch S, Böhm S, Börschig C, Chatzinotas A, Christ S, Daniel R, Diekötter T, Fischer C, Friedl T, Glaser K, Hallmann C, Hodac L, Hölzel N, Jung K, Klein AM, Klaus VH, Kleinebecker T, Krauss J, Lange M, Morris EK, Müller J, Nacke H, Pašalić E, Rillig MC, Rothenwöhrer C, Schall P, Scherber C, Schulze W, Socher SA, Steckel J, Steffan-Dewenter I, Türke M, Weiner CN, Werner M, Westphal C, Wolters V, Wubet T, Gockel S, Gorke M, Hemp A, Renner SC, Schöning I, Pfeiffer S, König-Ries B, Buscot F, Linsenmair KE, Schulze E-D, Weisser WW, Fischer M (2014) Interannual variation in land-use intensity enhances grassland multidiversity. Proc Natl Acad Sci USA **111**:308-313

Arndt H, Dietrich D, Auer B, Cleven E, Gräfenhan T, Weitere M, Mylnikov AP (2000) Functional diversity of heterotrophic flagellates in aquatic ecosystems. In: Leadbeater BSC & Green JC (Eds) The Flagellates, (pp 240-268) Taylor & Francis Ltd, London

Azam F, Fenchel T, Field JG, Gray JS, Meyer-Reil LA, Thingstad F (1983) The ecological role of water-column microbes in the sea. Mar. Ecol. Prog. Ser. **10**:257-263

Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R, Fierer N (2013) Global biogeography of highly diverse protistan communities in soil. ISME J**7**:652–659

Barreto DP, Conrad R, Klose M, Claus P, Enrich-Prast A (2014) Distance-decay and taxa-area relationships for bacteria, archaea and methanogenic archaea in a tropical lake sediment. PLoS One **9**:e110128

Berney C, Fahrni J, Pawlowski J (2004) How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. BMC Biol **2**:1-13

Blüthgen N, Dormann CF, Prati D, Klaus VH, Kleinebecker T, Hölzel N, Alt F, Boch S, Gockel S, Hemp A, Müller J, Nieschulze J, Renner SC, Schöning I, Schumacher U, Socher SA, Wells K, Birkhofer K, Buscot F, Oelmann Y, Rothenwöhrer C, Scherber C, Tscharntke T, Weiner CN, Fischer M, Kalko EKV, Linsenmair KE, Schulze E-D, Weisser WW (2012) A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. Basic Appl Ecol **13**:207–220

Blüthgen N, Simons NK, Jung K, Prati D, Renner SC, Boch S, Fischer M, Hölzel N, Klaus VH, Kleinebecker T, Tschapka M, Weisser WW, Gossner MM (2016) Land use imperils plant and animal community stability through changes in asynchrony rather than diversity. Nat Commun **7**:10697

Boenigk J, Arndt H (2002) Bacterivory by heterotrophic flagellates: community structure and feeding strategies. Antonie van Leeuwenhoek **81**:465-480

Bonkowski M (2004) Protozoa and plant growth: The microbial loop in soil revisited. New Phytologist **162**:617–631

Bonkowski M, Clarholm M (2012) Stimulation of plant growth through interactions of bacteria and Protozoa: testing the auxiliary microbial loop hypothesis. ACTA Protozoologica **51**:237-247

Bonkowski M, Roy J (2005) Soil microbial diversity and soil functioning affect competition among grasses in experimental microcosms. Oecologia **143**:232-240

Brabender M, Kiss ÁK, Domonell A, Nitsche F, Arndt H (2012) Phylogenetic and Morphological Diversity of Novel Soil Cercomonad Species with a Description of Two New Genera (Nucleocercomonas and Metabolomonas). Protist **163**:495–528

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD, Dennett MR, Moran DM, Jones AC (2009) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. Appl Environ Microbiol **75**:5797-5808

Chao A (1987) Estimating the population size for capture-recapture data with unequal catchability. Biometrics **43**:783-791

Chao A, Li PC, Agatha S, Foissner W (2006) A statistical approach to estimate soil ciliate diversity and distribution based on data from five continents OIKOS **114**:479-493

Collins MD, Vázquez DP, Sanders NJ (2002) Species-area curves, homogenization and the loss of global diversity. Evol Ecol Res **4**:457-464

De Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J-M, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horák A, Jaillon O, Lima-Mendez G, Lukeš J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans Coordinators: Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E (2015) Eukaryotic plankton diversity in the sunlit ocean. Science **348**:1261605

del Campo, Massana R (2011) Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. Protist **162**:435-448

Domonell A, Brabender M, Nitsche F, Bonkowski M, Arndt H (2013) Community structure of cultavable protists in different grassland and forest soils of Thuringia. Pedobiologia **56**:1-7

Dunthorn M, Otto J, Berger SA, Stamatakis A, Mahé F, Romac S, de Vargas C, Audic S, Consortium B, Stock A, Kauff F, Stoeck T (2014) Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. Mol Biol Evol **31**:993–1009

Egge E, Bittner L, Andersen T, Audic S, de Vargas C, Edvardsen B (2013) 454 Pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. PLOS One **8**:e74371

Ekelund F, Frederiksen HB, Rønn R (2002) Population dynamics of active and total ciliate populations in arable soil amended with wheat. Appl Environ Microbiol **68**:1096–1101

Ekelund F, Patterson DJ (1997) Some heterotrophic flagellates from a cultivated garden soil in Australia. Arch. Protistenkd. **148**:461-478

Esteban GF, Clarke KJ, Olmo JL, Finlay BJ (2006) Soil protozoa - An intensive study of population dynamics and community structure in an upland grassland. Appl Soil Ecol **33**:137–151

Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. Science **296**:1061-1063

Fischer M, Bossdorf O, Gockel S, Hänsel F, Hemp A, Hessenmöller D, Korte G, Nieschulze J, Pfeiffer S, Prati D, Renner S, Schöning I, Schumacher U, Wells K, Buscot F, Kalko EKV, Linsenmair KE, Schulze E-D, Weisser WW (2010) Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic Appl Ecol **11**:473–485

Foissner W (1999) Soil protozoa as bioindicators: Pros and cons, methods, diversity, representative examples. Agriculture, Ecosystems and Environment **74**:95–112

Foissner W (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta Protozoologica **45**:111–136

Foissner W, Bourland WA, Wolf KW, Stoeck T, Dunthorn M (2014) New SSU-rDNA

sequences for eleven colpodeans (Ciliophora, Colpodea) and description of Apocyrtolophosis nov. gen. Eur J Protistol **50**:40–46

Foissner W, Chao A, Katz LA (2008) Diversity and geographic distribution of ciliates (Protista: Ciliophora) Biodivers Conserv **17**:345-363

Foissner W, Moon-ven der Staay SY, van der Staay GWM, Hackstein JHP, Krautgartner W-D; Berger H (2004) Reconciling classical and molecular phylogenies in the stichotrichines (Ciliophora, Spirotrichea), including new sequences from some rare species. Eur J Protistol **40**:265-281

Forster D, Dunthorn M, Mahé F, Dolan JR, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie J-M, Decelle J, Edvardsen B, Egge E, Eikrem W, Gobet A, Kooistra WHCF, Logares R, Massana R, Montresor M, Not F, Ogata H, Pawlowski J, Pernice MC, Romac S, Shalchian-Tabrizi K, Simon N, Richards TA, Santini S, Sarno D, Siano R, Vaulot D, Wincker P, Zingone A, de Vargas C, Stoeck T (2016) Benthic protists: The under-charted majority. FEMS Microbiol Ecol **92**:1–11

Geisen S, Tveit AT, Clark IM, Richter A, Svenning MM, Bonkowski M, Urich T (2015) Metatranscriptomic census of active protists in soils. ISME J **9**:1–13

Gossner MM, Lewinsohn TM, Kahl T, Grassein F, Boch S, Prati D, Birkhofer K, Renner SC, Sikorski J, Wubet T, Arndt H, Baumgartner V,Blaser S, Blüthgen N, Börschig C, Buscot F, Diekötter T, Ré Jorge L, Jung K, Keyel AC, Klein A-M, Klemmer S, Krauss J, Lange M, Müller J, Overmann J, Pašalić E, Penone C, Perović DJ, Purschke O, Schall P, Socher SA, Sonnemann I, Tschapka M, Tscharntke T, Türke M, Venter PC, Weiner CN, Werner M, Wolters V, Wurst S, Westphal C, Fischer M, Weisser WW, Allan E (2016) Land-use intensification causes multitrophic homogenization of grassland communities. Nature **540**:266–269

Green J, Bohannan BJM (2006) Spatial scaling of microbial biodiversity. Trends Ecol Evol **21**:501-507

Grossmann L, Jensen M, Heider D, Jost S, Glücksman E, Hartikainen H, Mahamdallie SS, Gardner M, Hoffmann D, Bass D, Boenigk J (2016) Protistan community analysis: key findings of a large-scale molecular sampling. ISME J **10**:2269-2279

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet A-L, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R (2013) The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res **41**:597–604

Hadziavdic K, Lekanq K, Lanzen A, Jonassen I, Thompson EM, Troedsson C (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. PLoS One **9**:e87624

Hess S, Sausen N, Melkonian M (2012) Shedding light on vampires: the phylogeny of vampyrellid amoebae revisited. PLoS One **7**:e31165

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Gen Biol **8**:R143

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OUT clustering. Environ Microbiol **12**:1889-1898

Jeuck A, Nitsche F, Wylezich C, Wirth O, Bergfeld T, Brutscher F, Hennemann M, Monir S, Scherwaß A, Troll N, Arndt H (2017) A comparison of methods to analyze aquatic heterotrophic flagellates of different taxonomic groups. Protist **168**:375-391

Jousset A, Rochat L, Péchy-Tarr M, Keel C, Scheu S, Bonkowski M (2009) Predators promote defence of rhizosphere bacterial populations by selective feeding on non-toxic cheaters. ISME J **3**:666-674

Lara E, Mitchell EAD, Moreira D, López Gracía P (2011) Highly diverse and seasonally

dynamic protist community in a pristine peat bog. Protist **162**:14-32

Lentendu G, Wubet T, Chatzinotas A, Wilhelm C, Buscot F, Schlegel M (2014) Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. Mol Ecol **23**:3341-3355

Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F, Bittner L, Darzi Y, Wang J, Audic S, Berline L, Bontempi G, Cabello AM, Coppola L, Cornejo-Castillo FM, d'Ovidio F, De Meester L, Ferrera I, Garet-Delmas M-J, Guidi L, Lara E, Pesant S, Royo-Llonch M, Salazar G, Sánchez P, Sebastian M, Souffreau C, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans coordinators: Gorsky G, Not F, Ogata H, Speich S, Stemmann L, Weissenbach J, Wincker P, Acinas SG, Sunagawa S, Bork P, Sullivan MB, Karsenti E, Bowler C, de Vargas C, Raes J (2015) Ocean plankton. Determinants of community structure in the global plankton interactome. Science **348**:e1262073

Mahé, F, de Vargas C, Bass D, Czech L, Stamatakis A, Lara E, Singer D, Mayor J, Bunge J, Sernaker S, Siemensmeyer T, Trautmann I, Romac S, Berney C, Kozlov A, Mitchell EAD, Seppey CVW, Egge E, Lentendu G, Wirth R, Trueba G, Dunthorn M (2017) Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. Nat Ecol Evol **1**:0091

Mahé, F, de Vargas C, Bass D, Czech L, Stamatakis A, Lara E, Singer D, Mayor J, Bunge J, Sernaker S, Siemensmeyer T, Trautmann I, Romac S, Berney C, Kozlov A, Mitchell EAD, Seppey CVW, Egge E, Lentendu G, Wirth R, Trueba G, Dunthorn M (2017) Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. Nature Ecology and Evolution **1**:0091

Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ **3**:e1420

Medinger R, Nolte V, Pandey RV, Jost S, Ottenwälder B, Schlötterer C, Boenigk J (2010)

Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol Ecol **19**:32-40

Nebel M, Pfabel C, Stock A, Dunthorn M, Stoeck T (2011) Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. Environ Microbiol Rep **3**:154-158

Niklas N, Pröll J, Danzer M, Stabentheiner S, Hofer K, Gabriel C (2013) Routine performance and errors of 454 HLA exon sequencing in diagnostics. BMC Bioinformatics **14**:176

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hare RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2017) Package 'Vegan': Community ecology package. R package version 2.4-4. https://github.com/vegandevs/vegan/issues

Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M, Fiore-Donno AM, Gile GH, Holzmann M, Jahn R, Jirků M, Keeling PJ, Kostka M, Kudryavtsev A, Lara E, Lukeš J, Mann DG, Mitchell EAD, Nitsche F, Romeralo M, Saunders GW, Simpson AGB, Smirnov AV, Spouge JL, Stern RF, Stoeck T, Zimmermann J, Schindel D, de Vargas C (2012) CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. PLoS Biol **10**:e1001419

Scherber C, Eisenhauer N, Weisser WW, Schmid B, Voigt W, Fischer M, Schulze E-D, Roscher C, Weigelt A, Allan E, Beßler H, Bonkowski M, Buchmann N, Buscot F, Clement LW, Ebeling A, Engels C, Halle S, Kertscher I, Klein A-M, Koller R, König S, Kowalski E, Kummer V, Kuu A, Lange M, Lauterbach D, Middelhoff C, Migunova VD, Milcu A, Müller R, Partsch S, Petermann JS, Renker C, Rottstock T, Sabais A, Scheu S, Schumacher J, Temperton VM, Tscharntke T (2010) Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. Nature **468**:553-556

Schlebusch S, Illing N (2012) Next generation shotgun sequencing and the challenges of de novo genome assembly. S Afr J Sci **108**:1-8

Schoenle A, Jeuck A, Nitsche F, Venter P, Prausse D, Arndt H (2016) Methodological studies on estimates of abundance and diversity of heterotrophic flagellates from the deep-sea floor. J Mar Sci Eng **4**:1-11

Šlapeta J, Moreira D, López-García P (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. Proc Biol Sci **272**:2073–2081

Soliveres S, Manning P, Prati D, Gossner MM, Alt F, Arndt H, Baumgartner V, Binkenstein J, Birkhofer K, Blaser S, Blüthgen N, Boch S, Böhm S, Börschig C, Buscot F, Diekötter T, Heinze J, Hölzel N, Jung K, Klaus VH, Klein A-M, Kleinebecker T, Klemmer S, Krauss J, Lange M, Morris EK, Müller J, Oelmann Y, Overmann J, Pašalić E, Renner SC, Rillig MC, Schaefer HM, Schloter M, Schmitt B, Schöning I, Schrumpf M, Sikorski J,Socher SA, Solly EF, Sonnemann I, Sorkau E, Steckel J, Steffan-Dewenter I, Stempfhuber B, Tschapka M, Türke M, Venter P, Weiner CN, Weisser WW, Werner M, Westphal C, Wilcke W, Wolters V, Wubet T, Wurst S, Fischer M, Allan E (2016a) Locally rare species influence grassland ecosystem multifunctionality. Philos TransR Soc LondB Biol Sci **371**:20150269

Soliveres S, van der Plas F, Manning P, Prati D, Gossner MM, Renner SC, Alt F, Arndt H, Baumgartner V, Binkenstein J, Birkhofer K, Blaser S, Blüthgen N, Boch S, Böhm S, Börschig C, Buscot F, Diekötter T, Heinze J, Hölzel N, Jung K, Klaus VH, Kleinebecker T, Klemmer S, Krauss J, Lange M, Morris EK, Müller J, Oelmann Y, Overmann J, Pašalić E, Rillig MC, Schaefer HM, Schloter M, Schmitt B, Schöning I, Schrumpf M, Sikorski J, Socher SA, Solly EF, Sonnemann I, Sorkau E, Steckel J, Steffan-Dewenter I, Stempfhuber B, Tschapka M, Türke M, Venter PC, Weiner CN, Weisser WW, Werner M, Westphal C, Wilcke W, Wolters V, Wubet T, Wurst S, Fischer M, Allan E (2016b) Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. Nature

536:456-459

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, Richards TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. MolEcol **19**:21–31

Venter PC, Nitsche F, Domonell A, Heger P, Arndt H (2017) The protistan microbiome of grassland soil: diversity in the mesoscale. Protist **168**:546-564

Zinger L, Boetius A, Ramette A (2014) Bacterial taxa-area and distance-decay relationships in marine environments. Mol Ecol **23**:954-964

Erfolgsbilanz und Teilpublikationen

[1] **Venter PC, Nitsche F, Domonell A, Heger P, Arndt H** (2017) The protistan microbiome of grassland soil: diversity in the mesoscale. Protist **168**:546-564

Die Konzeption, molekularbiologische Arbeit, die bioinformatische Auswertung, die statistischen Analysen, und der wesentliche Teil des Textes wurden vom Autor durchgeführt bzw. erarbeitet (insgesamt etwa 80%).

[2] **Gossner MM, Lewinsohn TM, Kahl T, Grassein F, Boch S, Prati D, Birkhofer K, Renner SC, Sikorski J, Wubet T, Arndt H, Baumgartner V,Blaser S, Blüthgen N, Börschig C, Buscot F, Diekötter T, Ré Jorge L, Jung K, Keyel AC, Klein A-M, Klemmer S, Krauss J, Lange M, Müller J, Overmann J, Pašalić E, Penone C, Perović DJ, Purschke O, Schall P, Socher SA, Sonnemann I, Tschapka M, Tscharntke T, Türke M, Venter PC, Weiner CN, Werner M, Wolters V, Wurst S, Westphal C, Fischer M, Weisser WW, Allan E** (2016) Land-use intensification causes multitrophic homogenization of grassland communities. Nature **540**:266–269

Die bioinformatische Auswertung der Protisten, die Literaturarbeit und die dazugehörige Diskussion, sowie die Überarbeitung des Manuskriptes in allen Phasen des Publikationsprozesses wurden von Herrn Venter durchgeführt bzw. erarbeitet (insgesamt etwa 10%).

[3] **Soliveres S, Manning P, Prati D, Gossner MM, Alt F, Arndt H, Baumgartner V, Binkenstein J, Birkhofer K, Blaser S, Blüthgen N, Boch S, Böhm S, Börschig C, Buscot**

F, Diekötter T, Heinze J, Hölzel N, Jung K, Klaus VH, Klein A-M, Kleinebecker T, Klemmer S, Krauss J, Lange M, Morris EK, Müller J, Oelmann Y, Overmann J, Pašalić E, Renner SC, Rillig MC, Schaefer HM, Schloter M, Schmitt B, Schöning I, Schrumpf M, Sikorski J,Socher SA, Solly EF, Sonnemann I, Sorkau E, Steckel J, Steffan-Dewenter I, Stempfhuber B, Tschapka M, Türke M, Venter P, Weiner CN, Weisser WW, Werner M, Westphal C, Wilcke W, Wolters V, Wubet T, Wurst S, Fischer M, Allan E (2016) Locally rare species influence grassland ecosystem multifunctionality. Philos TransR Soc LondB Biol Sci **371**:20150269

Die bioinformatische Auswertung der Protisten, die Literaturarbeit und die dazugehörige Diskussion, sowie die Überarbeitung des Manuskriptes in allen Phasen des Publikationsprozesses wurden von Herrn Venter durchgeführt bzw. erarbeitet (insgesamt etwa 10%).

[4] **Soliveres S, van der Plas F, Manning P, Prati D, Gossner MM, Renner SC, Alt F, Arndt H, Baumgartner V, Binkenstein J, Birkhofer K, Blaser S, Blüthgen N, Boch S, Böhm S, Börschig C, Buscot F, Diekötter T, Heinze J, Hölzel N, Jung K, Klaus VH, Kleinebecker T, Klemmer S, Krauss J, Lange M, Morris EK, Müller J, Oelmann Y, Overmann J, Pašalić E, Rillig MC, Schaefer HM, Schloter M, Schmitt B, Schöning I, Schrumpf M, Sikorski J, Socher SA, Solly EF, Sonnemann I, Sorkau E, Steckel J, Steffan-Dewenter I, Stempfhuber B, Tschapka M, Türke M, Venter PC, Weiner CN, Weisser WW, Werner M, Westphal C, Wilcke W, Wolters V, Wubet T, Wurst S, Fischer M, Allan E** (2016) Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. Nature 536:456-459

Die bioinformatische Auswertung der Protisten, die Literaturarbeit und die dazugehörige Diskussion, sowie die Überarbeitung des Manuskriptes in allen Phasen des Publikationsprozesses wurden von Herrn Venter durchgeführt bzw. erarbeitet (insgesamt etwa

10%).

[5] **Schoenle A, Jeuck A, Nitsche F, Venter P, Prausse D, Arndt H** (2016) Methodological studies on estimates of abundance and diversity of heterotrophic flagellates from the deep-sea floor. J Mar Sci Eng **4**:1-11

Die bioinformatische Auswertung und die Beratung hinsichtlich der Methodenpräsentation und die dazugehörige Diskussion sowie die Überarbeitung des Manuskriptes in allen Phasen des Publikationsprozesses wurden von Herrn Venter durchgeführt bzw. erarbeitet (insgesamt etwa 10%).

[6] **Venter PC, Nitsche F, Scherwass A, Arndt H** (*under review*) Descrepancies between molecular and morphological databases of soil ciliates studied for temperate grasslands of central Europe. Protist

Die Konzeption, molekularbiologische Arbeit, die bioinformatische Auswertung, die statistischen Analysen, ein wichtiger Teil der Literaturarbeit und der wesentliche Teil des Textes wurden von Herrn Venter durchgeführt bzw. erarbeitet (insgesamt etwa 80%).

[7] **Venter PC, Nitsche F, Arndt H** (*subm.*) The hidden diversity of flagellated protists in soil. Protist

Die Konzeption, molekularbiologische Arbeit, die bioinformatische Auswertung, die statistischen Analysen, ein wichtiger Teil der Literaturarbeit und der wesentliche Teil des Textes wurden von Herrn Venter durchgeführt bzw. erarbeitet (insgesamt etwa 80%).

Erklärung (gemäß § 4 Abs. (1) Nr. 9)

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit − einschließlich Tabellen, Karten und Abbildungen −, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie − abgesehen von unten angegebenen Teilpublikationen − noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Hartmut Arndt betreut worden.

_____

Paul Christiaan Venter