Evolutionary genetic analysis of an invasive population of sculpins in the Lower Rhine

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Arne W. Nolte

aus Oldenburg

Köln, 2005

Berichterstatter: Prof. Dr. D. Tautz

Prof. Dr. H. Arndt

Prof. Dr. A. Meyer

Tag der letzten mündlichen Prüfung: Freitag der 9. Juli 2005

Acknowledgements

I thank Prof. Dr. D. Tautz for his continuous interest in sculpin biology. This has not only provided me with a fascinating project but also contributed significantly to the fact that sculpins are now among the genetically best-explored freshwater fishes in Europe. His inventive spirit and support have made it possible to put ambitious ideas into practice.

Prof. Dr. H. Arndt and Prof. Dr. A. Meyer have kindly agreed to evaluate this thesis.

I thank Dr. J. Freyhof, who made the initial discovery that invasive sculpins are “different” and introduced me to this problem during undergraduate times. Since then, he has been extremely helpful in all questions surrounding fishes and, with respect to this work, in providing insights into ecological aspects.

The sculpin project was possible only with benevolent support and permissions from Dr. A. Mellin, T. Heilbronner, W. Fettweis, C. Bode, L. Jörgensen, L. Kroll, Dr. C. Köhler, W. Fricke, Dr. H. Arzbach and M. Kämmereit and with material or logistic support from Dr. F. Volckaert, Dr. B. Hänfling, Dr. D. Neely, Dr. L. Bervoets, Dr. G. Knaepkens, Dr. E. Winter, I. Steinmann, Dr. E. Korte, U. Weibel, C. Dümpelmann, Dr. W. Dönni, C. von Landwüst and Dr. A. Waterstraat as well as numerous local fishermen.

As a newcomer to molecular biology I had to learn first how to produce data in the lab and at later stages how to make use of those data. I thank all current and former members of the Tautz Group for a pleasant, productive and helpful working environment, most of all Kathryn Stemshorn who joined the *Cottus* project in several ways.... As for the computer work, I acknowledge the persistent support of our specialists Chris Voolstra and Till Bayer. Dr. Röbbe Wünschiers has introduced me to the world of “Computational Biology” and Till Bayer and Dr. Tina Harr have helped me with extremely useful scripts to handle raw data. Many thanks to Susanne Kipp for help in careful phenotypic analyses that have been the second pillar of my work besides molecular data.

Above all, I thank my parents, Dr. Helga Nolte and Karl Heinz Nolte, for allowing me to pursue and develop my interest in biology.

Contents

Preface: Of sculpins and model organisms

A straightforward approach to study biological phenomena is to start with theory and test predictions. Proceeding this way allows one to choose beforehand, which study organism best suits the needs. Most likely, the choice would be among the so-called model organisms for which an overwhelming wealth of background information is available. Admittedly, the increase of knowledge in all disciplines of biology reflects the tremendous success and utility of this approach.

My motivation for this work stems less from initial theoretical considerations but was borne out of an interest in natural history of fishes. I have been intrigued by simple and yet open questions regarding the appearance of sculpins where nobody expected them to be. At a glance, finding answers on where sculpins may come from and what has permitted them to appear in the first place seems to be a minor detail. Yet, careful analyses of "experiments" conducted by nature will reveal novel aspects and approach limits of our knowledge, most likely from a unique perspective. Most importantly, to unravel a complex situation as found in natural ecosystems must not be seen as imposing limitations to peel out a single factor but forces one to integrate knowledge into a complex picture.

I hope that the reader may share a fascination with nature's ugly ducklings (Baugh 1980, Katula 1998) and wish that studies inspired by and devoted to natural history can be carried on to merge with what we know from models, be it theoretical or live ones.

Baugh, T. (1980) A Netful of Natives. RCM Publications. Sierra Madre, California.
Katula, R. (1988) The Good, the Bad, and the Very Ugly Sculpins, American Currents, March-June.

Introduction

An invasion of the Lower Rhine

Less than 20 years ago populations of the European sculpin *Cottus gobio* L. 1758 (Cottidae, Scorpaeniformes; Teleostei) were discovered in the main channel of the German part of the Lower River Rhine for the first time (Schleuter 1991). Besides new records from the German stretch of the Lower Rhine a recent study on the fish fauna in the Netherlands has demonstrated a tremendous range expansion as well. While sculpins were found in few places before 1980 more recent surveys have revealed widespread occurrence during the following decades (De Nie 1997). Sculpins were particularly abundant in large rivers, artificial canals and the IJsselmeer. These waters represent the summer-warm *potamal* harbouring a typical fish community characteristic for the lower reaches of large rivers (Lelek and Buhse 1992). European sculpins have not been described from habitats including muddy and stagnant backwaters before and the fact that they suddenly occurred together with species like flounder (*Pleuronectes flesus*), ide (*Leuciscus idus*) and ruffe (*Gymnocephalus cernuus*) lead Volz and Cazemier (1991) to the conclusion that common views about the autecology of sculpins would have to be revised. *Cottus gobio* is traditionally thought to be confined to headwater regions of streams (Vogt and Hofer 1909) to which they are tied by their need of high amounts of dissolved oxygen (Nikolsky 1978).

It needs to be emphasized here, that this novel pattern of distribution of sculpins must bee seen in the context of a wake of ecological perturbations that have affected the River Rhine during the last centuries. Heavy pollution from industrial, communal and agricultural wastewaters has dramatically affected water quality resulting in a decrease of species diversity in the past (Lelek and Buhse 1992). Moreover, the main channel of the Rhine has been dredged and fortified with rocks to become an important inland waterway for ships. Increased sensitivity to environmental issues resulted in concerted conservation measures by neighbouring countries now organised in the IKSR (International Commission for the Protection of the Rhine) and from the mid-seventies on, the water quality improved steadily. This has resulted in an increase of rheophilic species of fish that were absent or rare during times of heavy pollution (Staas 1997). Ecological change of the Rhine is also documented by marked fluctuations in the invertebrate community including the reappearance of previously vanished taxa and waves of invasion by alien species (Neumann 2002). All of these changes are thought to reflect improved water quality and this is likely to have affected sculpins as well.

On the other hand the Rhine drainage remains subdivided into distinct ecological subregions as described by Illies (1961). Of these, the basic division into *potamal* and *rhithral* are relevant with respect to sculpins because according to current knowledge they have not recolonized the former but in fact invaded a new ecological region (Fig. 1). In this study population genetic approaches are used to reconstruct the history of colonization within the Rhine and some of its tributaries. Presently this does not include experimental approaches to identify relevant adaptations but relies on the idea that

population genetic structure in current populations should reflect both population history and persistent patterns of selection and adaptation. This bears an advantage in that one can approach the phenomenon of invasion without the need of precise hypothesis about the nature of adaptations as these are difficult to develop given the recurrent ecological change and dynamics of the Rhine.

Ecological genetics and phylogeographic inference

Two key questions can be phrased with respect to the appearance of sculpins within the Rhine. The first relates to the source populations and the second would ask for the reasons that have allowed for this invasion. Both can be approached with molecular markers using the methodologies of phylogeography (Avise 2000) and ecological genetics (Lowe *et al.* 2004). These two fields largely overlap and complement one another as for the methods that are currently applied. However, they differ in that phylogeography reconstructs past processes from current patterns and thus integrates a historical component while ecological genetics is more focused on mechanisms of natural selection and adaptation at the level of the gene. A key concept of ecological genetics is that associations of genotypes with environments can arise as a result of habitat specific fitness of a given genotype. If a genotype was adaptive it would rise in frequency while a maladaptive genotype would be removed by natural selection. However, associations of genotypes with environments can also originate from population subdivision, i.e. all genotypes of a population living on a mountain would be associated with that mountain. This results from the historical fact that the population occurs on that mountain and would not be informative of mechanisms acting on separate genetic factors. Thus the concerted action of processes of selection and of population history shapes the distribution of genotypes and populations in nature. As a result the key challenge in this field is to disentangle these alternative explanations. Of course, this is only possible in study systems where both factors have an influence and can be traced.

Figure 1: Sculpins have recently invaded large river habitats (upper picture - the Lower Rhine at Duisburg), downstream of headwaters where they typically occur (middle picture – Stream Broel at Winterscheid). So-called *invasive* sculpins (upper specimen) differ from those in streams of the Lower Rhine area (lower specimen) in having a deeper body and more protruding eyes among other characters (Pictures by I. Steinmann and A. Hartl).

The study system: continuous bodies of water and leaky genepools

Sculpins within the Lower Rhine area can be exploited in ecological genetics studies for two main reasons. Above all, the study sites within the River Rhine are part of a continuous body of water, which is divided ecologically but not by physical barriers to dispersal. In such a setting, the distribution of genotypes can be (cautiously) interpreted as a result from a choice situation. In other words, if the distribution of genotypes within an open system is not random, then mechanisms have to be assumed that determine this distribution. Secondly, previous studies have already documented that the involved populations are part of a wide-ranging area of postglacial secondary contact (Englbrecht *et al.* 2000, Volckaert *et al.* 2002). This finding has important implication for evolutionary processes since old phylogeographic lineages of sculpins were shown to represent “porous” diverging genepools, which can exchange genetic material despite their persisting isolation. For this study this justifies a consideration of genetic factors independently of old evolutionary lineages. Recombination of genetic material originating from divergent genepools is thought to be an important evolutionary factor (Arnold 1997, Barton 2001) and it is widely accepted that hybrids can be inviable or less fit than their parents. However, the view that hybridization may enhance fitness and promote evolutionary progress is currently more widespread among botanists than zoologists. Arnold *et al.* (1999) noted a common reluctance to openly interpret the occurrence of natural hybrids by stating that *“Individual organisms with different genotypes demonstrating varying fitnesses depending on environment is a fundamental concept of evolutionary theory. However, this truism is not usually extended to cases where natural hybridization has occurred…”.* It is this simple relationship of fitness, genotype and environment that has served as a primary backbone in the interpretation of the population genetic patterns in this study, regardless of whether hybrids or pure lineages were concerned. The chapters of this thesis illustrate how this principle together with phylogeographic and ecological genetic approaches can explain aspects of the distribution dynamics of sculpins in the Rhine drainage.

Aspects treated in this study

Source populations and colonization history

The range expansion of invasive sculpins is analysed and documented in Chapters 1 and 4. In Chapter 1 the genenepool of the invasive sculpins is traced back to its phylogeographic origins. Extensive sampling of surrounding areas shows that invasive sculpins are recent hybrids that have resulted from a mixture of two ancestral lineages. Streams surrounding the distribution range of invasive sculpins still harbour the ancestral lineages. Note that Invasive sculpins represent a *hybrid lineage* in contrast to *recent hybrids* that are studied in later chapters (Chapter 4, 5, 6).

Development of genetic markers and genomic recources

Further analysis of the genetics of hybridization relies on access to genetic markers and map information. In Chapter 2 the isolation of microsatellites from *Cottus* is described. These markers were applied in the construction of the first genetic linkage map for *Cottus* (Chapter 3). The obtained linkage information served to optimize the analysis of genetic structure at hybrid zones (Chapter 4) and to gain first insights into the genetic architecture of quantitative trait divergence (Chapter 6).

Evolutionary ecology of hybrids

Because representatives of the ancestral lineages were not found to colonize large river habitats there is a suspicious correlation between hybridism and ecological success in a novel habitat. It is hypothesized that (Chapter 1) mechanisms of hybrid speciation have contributed to the invasive sculpins ecological success. The opposite seems to be true for recent hybrids, which are not able to compete with parental lineages (Chapter 4). This observation is most likely attributable to selection inflicted by the environment (Chapter 4).

Genetic architecture of divergence

Hybrid speciation theory assumes that transgressive segregation is a key process that may create evolutionary novelty. In Chapter 5 evidence for transgressive phenotypes in body shape is found in recent natural hybrids. Finally in Chapter 6 first insights into the genetic architecture across sculpin hybrid zones are obtained. Genetic factors affecting morphological differentiation and habiat specificity are apparently physically linked. Both Chapter 5 and 6 provide evidence that the differentiation that is observed across hybrid zones is heritable. Moreover the apparent genomic linkage of different divergent traits has important implications for the evolutionary ecology of hybrids.

Literature:

Avise, J. C. (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge, viii+447 pp.

Arnold, M. L. (1997) Natural Hybridization and Evolution. Oxford University Press, New York.

Arnold, M. L. Bulger, M. R. Burke, J. M. Hempel, A. L. Williams, J. H. (1999) Natural Hybridization: how low can you go and still be important? Ecology, 80, 2, 371–381.

Barton, N. H. (2001) The role of hybridization in evolution. Mol. Ecol. 10, 551-568.

De Nie, H. W. (1997) Atlas van de Nederlandse Zoetwatervissen. Media Publishing, Doetinchen, 151 pp.

Englbrecht, C. C. Freyhof, J. Nolte, A. Rassmann, K. Schliewen, U. Tautz, D. (2000) Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-pleistocene origin of the major central European populations Mol. Ecol. 9, 709-722.

Illies, J. (1961) Versuch einer allgemeinen biozönotischen Gliederung der Fließgewässer. Int. Rev. Ges. Hydrobiol. 46, 205-213.

Lelek, A. Buhse, G. (1992) Fische des Rheins – früher und heute. Springer Verlag, Berlin Heidelberg. 214 pp.

Lowe, A. Harris, S. Ashton, P. (2004) Ecological genetics: design, analysis, and application. Blackwell Publishing, Oxford, UK. vii-x+320pp.

Neumann, D. (2002) Ecological Rehabilitation of a Degraded Large River System – Considerations Based on Case Studies of Makrozoobethos and Fish in the Lower Rhine and Its Catchment Area. Internat. Rev. Hydrobiol. 87, 139-150.

Nikolsky, G. V. (1978) The Ecology of Fishes. T.F.H Publications. Neptune City, New York. v-ix+329 pp.

Schleuter, M. Nachweis der Groppe (*Cottus gobio*) im Niederrhein. Fischökologie 4, 1-6 (1991).

Staas, S. (1997) Das Jungfischaufkommen im Niederrhein und in angrenzenden Nebengewässern unter Berücksichtigung der Uferstruckturen am Strom, hrsg.: Landesanstalt für Ökologie, Bodenordnung und Forsten / Landesamt für Agrarordung NRW, LÖBF-Schriftenreihe, Band 12. 111pp.

Volz, J. Cazemier, G. (1991) Die Fischfauna im niederländischen Rhein – eine aktuelle Bestandsaufnahme. Fischökologie 5, 3-18.

Volckaert, F. A. M. Hänfling, B. Hellemans, B. & Carvalho, G. R. (2002). Timing of the population dynamics of bullhead *Cottus gobio* (Teleostei: Cottidae) during the Pleistocene. J. Evol. Biol. 15, 930-944.

Vogt, C. Hofer, B. (1909) Die Süsswasserfische von Mittel-Europa. Engelmann, Leipzig, xxiv+558 pp.

Declaration of collaborators contributions

I have designed the research together with Diethard Tautz and conducted the largest part of the practical work. This includes sampling, molecular work and data analyses. However, significant input of others has improved the quality of this study. I had the pleasure to cooperate with several colleagues whose input and contribution I acknowledge below.

Chapter 1: Hybrid invasion

Jörg Freyhof helped to develop the sampling scheme and took part in the fieldwork. Further, the inferred patterns of distribution and life history data are results of J. Freyhofs' long-term studies on fishes in the River Rhine system. Kathryn Stemshorn conducted the SNP screen and contributed to the microsatellite genotyping of Scheldt sculpins.

Chapter 2: Microsatellites

Kathryn Stemshorn has contributed data on the mendelian inheritance of all markers from screens in parents and offspring.

Chapter 3: Genetic Map

Kathryn Stemshorn genotyped mapping families and performed data analyses with my assistance. I have developed markers and protocols, collected sculpins to produce mapping families, crossed them and raised the offspring.

Chapter 4: Nascent Hybrid Zones

Jörg Freyhof provided background knowledge on the fish communities and ecological settings in the River Sieg system and helped to identify hybrid zones. We have also collaborated to conduct the underlying fieldwork.

Chapter 5: Transgressive Phenotypes

H. David Sheets has put our joint idea of an assignment procedure with statistical tests into practise. He has contributed the geometric morphometric methods to this work and wrote the software that I have used.

Chapter 1: An invasive lineage of sculpins, *Cottus* sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups

Arne W. Nolte, Jörg Freyhof, Kathryn C. Stemshorn and Diethard Tautz

Abstract

Fish abundance surveys in the Rhine system have shown in the past two decades that there is a rapid upriver invasion of a freshwater sculpin of the genus *Cottus*. These fish are found in habitats that are untypical for the known species *Cottus gobio*, which is confined to small cold streams within the Rhine drainage. Phylogeographic analysis based on mitochondrial haplotypes and diagnostic single nucleotide polymorphisms indicates that the invasive sculpins are hybrids between two old lineages from the River Scheldt drainage and the River Rhine drainage, although it is morphologically more similar to the Scheldt sculpins. Most importantly, however, the invasive population possesses an unique ecological potential that does not occur in either of the source populations from the Rhine or the Scheldt, which allows the colonisation of new habitats that have previously been free of sculpins. Microsatellite analysis shows that the new lineage is genetically intermediate between the old lineages and that it forms a distinct genetic group across its whole expansion range. We conclude that hybridization between long separated groups has lead to the fast emergence of a new, adaptationally distinct sculpin lineage.

Introduction

In Central Europe, sculpins commonly known as *Cottus gobio* L. are stenoecious inhabitants of well-oxygenated cold streams and lakes (Vogt & Hofer 1909). They are generally absent from downstream habitats such as large rivers or artificial canals. Unexpectedly, less than 20 years ago, Cazemir (1988) and Brink *et al.* (1990) reported sculpins to be common in the Lower Rhine of the Netherlands, which is a typical summer warm potamal habitat. In the ecologically similar German stretch of the Lower Rhine, parallel new records of sculpins were made by Schleuter (1991) and by Lelek & Köhler (1993). A comparison between the distribution in 1960 -1979 and the more recent situation in 1991-1995 in the Netherlands was published by De Nie (1997). While sculpins were identified only in a few lowland streams in the earlier surveys, the later ones showed them to be very abundant in the Rhine Delta, the River Maas and the IJsselmeer. Intriguingly, sculpins were now preferentially found in large bodies of water, which were not used as a habitat

before (De Nie 1997). Again, this was paralleled by findings from the German Lower Rhine where the invading sculpins were also found in large stagnant water bodies connected to the Rhine, for instance backwaters and harbours (own data). All of these observations indicate that the sculpins recently invading parts of the Lower Rhine drainage display a previously unknown tolerance to summerwarm and turbid waters within the lower reaches of large rivers. The sudden appearance of the invasive sculpins raises the question of their origins. Köhler *et al.* (1993) discussed two possible ways of colonisation. Sculpins could have persisted in the Lower Rhine and recolonized the riverine habitats after the improvements in water quality, starting in the 1980th. Alternatively, they could have colonised the main river downstream from source populations in its tributaries. The invasive sculpins could also represent introduced, non-native invaders since they share the conspicuous skin prickling with populations from elsewhere in Europe (Koli 1969). This includes the Scheldt system, which became connected to the Rhine system via canals in the past centuries and would therefore be an obvious source for a non-native invasion. However, apart of the identification of the source populations for the invasion, the most intriguing question relates to the factors that have allowed this population expansion in the first place, since neither the surveys in the Scheldt nor in the Rhine basin (Vandelannoote *et al.* 1998, Köhler et al 1993) have previously documented sculpins to be eudominant members of the fish communities of large lowland rivers.

Phylogeographic analyses of mitochondrial DNA of European sculpin populations, the so-called *Cottus gobio* complex, have revealed several clearly distinct groups across Central Europe (Englbrecht *et al.* 2000; Volckaert *et al.* 2002). This makes them one of the most deeply substructured European fish taxa studied so far. Sculpins have persisted during glacial cycles within separate refugia across much of their Central European range. The oldest lineages are separated since up to 3 Myr, and the ones from the Rhine and the Scheldt are separated since up to 1 Myr. Intriguingly, the Upper Rhine tributaries harbour a different lineage than the Lower Rhine tributaries. Both have not formed a homogenized genepool although a river capture united them in a single basin about 1 Myr ago. Thus, riverine habitats that would connect subpopulations apparently acted as a prohibitive barrier. The only previously noted exception to the otherwise clear phylogeographic structure was related to sculpins from the Lower Rhine. In some populations from the area where the above described surveys have indicated a recent invasion, mtDNA haplotypes originating from adjacent phylogeographic lineages were discovered (Englbrecht *et al.* 2000; Volckaert *et al.* 2002).

Here, we show that the invasive sculpin from the Lower Rhine harbours indeed a hybrid gene pool, derived from its ancestral lineages from the Scheldt and the Lower Rhine system. Only the hybrid population has been successful in invading riverine habitats of the Lower Rhine. Representatives of the ancestral lineages have not expanded their ranges despite the absence of geographical barriers. We discuss these results in the context of models for hybrid speciation.

Methods:

Sampling

Fish were collected using portable elektroshockers. Fin clips were preserved in 99% ethanol. Specimens for morphological analysis were fixed in 4% formaldehyde and later transferred to 70 % ethanol. The fish community of the River Sieg was studied by J.F. in a monitoring project since 1992. Sampling was done twice annually allowing to reconstruct population dynamics of sculpins in detail. The fish fauna of the River Mosel and the Middle and Northern Upper Rhine was studied from 1997 – 2001. Further details for all sampling sites are provided in Chapter 1 - suppl. Tab. 1.

Morphology

Sculpins vary in the degree to which spinelike scales (prickling) cover the body (Koli 1969). Five classes were distinguished for this study: 0 = prickles absent; 1 = less than ten prickles present beneath pectoral fin; 2 = more than ten prickles but all covered by the pectoral fin; 3 = prickling extends beyond pectoral fin but ends anterior to the middle of the second dorsal fin; 4 = prickling extends back beyond the middle of the second dorsal fin. A subset of individuals was examined for differentiation in body shape. Well-preserved Rhine sculpins from small streams (populations # 23, 24, 25, 26, 28, 29, 31, 34, 35), Invasive sculpins (populations # 8, 10, 12, 13) and Scheldt sculpins (populations # 65, 66) were analysed using landmark based methods (Rohlf and Marcus 1993). A set of 14 anatomical landmarks was chosen to capture the shape from a lateral view. Data were analysed using the software packages TPS (Rohlf 2003) and IMP (Sheets 2002). All specimens were superimposed by procrustes methods. Partial warp scores were used for the morphometric analysis. Overall differentiation in shape was measured as bootstrapped full procrustes distance between group means. CVA was used in order to identify those shape vectors serving to discriminate best between groups. The shape change implied by the discriminant axes was visualized as vectors on a deformation grid. Partial warp scores were regressed on centroid size to evaluate the confounding effects of allometry. This did not notably affect the outcome of the analysis suggesting that growth contributes little to the observed differentiation.

Life history data

Fully mature females were collected at the beginning of the spawning season (March 2003). 43 females from two Invasive sculpin populations (# 13, 64) and 93 females from four Rhine sculpin populations (# 24, 25, 28, 63) were included. Premature females of Scheldt sculpin were collected in January 2004 from two streams (# 65, 66) in the River Scheldt drainage. Own laboratory observations showed that all sculpins studied are single clutch spawners. All specimen were measured (SL) before the gonads were removed. Age was determined by otholith analysis.

Molecular analysis

The mitochondrial control region of sculpins was sequenced as previously described (Englbrecht *et al.* 2000). All mt - haplotypes could be unambiguously assigned to known groups (following Vockaert *et al.* 2002 and Knapen *et al.* 2003). SNP markers were developed from sequences of 12 random genomic clones. PCR primers were chosen to amplify fragments ranging from 500 to 700 bp, which were directly sequenced. The sequences were screened for polymorphisms in 20 specimen representing different phylogeographic lineages. Using *Cottus* sibiricus and *Cottus* ricei (River Olkha, southwest of Irkutusk, Siberia / Smoky River, Alberta, Canada; by courtesy of D. Neely, St Louis) as outgroups we could identify high frequency alleles that were derived states for a given lineage. Out of the 12 loci analysed in this way, the following loci yielded diagnostic SNP differences (primer details in Chapter 1 - suppl. Tab. 2; genotype details in Chapter 1 - suppl. Tab. 3): CgoSNP1, CgoSNP2, CgoSNP3 (two seperate SNPs, A and B), CgoSNP4 and CgoSNP5 (GenBank Accn. CL242132; CL242133; CL242134; CL242135; CL242136). Additional specimens were typed for these loci by pyrosequencing on a PSQ 96 MA (Pyrosequencing AB). The microsatellite loci Cgo18, Cgo33, Cgo56, Cgo42, Cgo1114 and Cgo1033 (Englbecht *et al.* 1999) were typed on a MegaBace 1000 sequencer (Amersham Biosciences) to assess the population substructure of 23 populations (populations # 6 – 9, 11 –16, 22 – 26, 28 – 29, 31, 33 – 35, 65 - 66; total n: 950 individuals) within the Lower and Middle Rhine basin. An analysis of overall genetic distance of the Invasive sculpin genepool to representatives of the presumed ancestral lineages was based on fewer populations (Invasive sculpin #10, Rhine sculpin, # 24, 26; Scheldt sculpin, # 65, 66) but involved typing of 120 microsatellite loci from Nolte *et al.* (2005) (Chapter 1 – suppl. Tab. 4).

Population genetic analyses

MSA 3.15 (Dieringer and Schlötterer. 2003) was used to calculate pairwise genetic distances. We used Nei's standard genetic distance to make our data comparable to Knapen *et al.* (2003). Furthermore, deviations from a stepwise mutation model were detected for some loci, thus distances had to be based on an infinite allele model. The choice of alternative genetic distance measures did not significantly change the outcome of the analyses presented here (not shown). Exact tests for population differentiation in diagnostic SNP frequencies and for differential contribution of haplotype groups were conducted using ARLEQUIN version 2001 (Schneider *et al.* 2000).

Results:

Invasion of a new sculpin into river habitats of the Rhine

Fish abundance surveys detected in 1992 sculpins with intense skin prickling (see below) in the lowest reaches of the Sieg. In the following ten years, this sculpin expanded its range upriver with an average rate of approximately 4-8 km per year (Fig. 1). During this whole period, sculpins without skin prickling remained confined to their native streams and did not colonize riverine habitats despite the fact that many sculpin larvae are known

to drift downstream after they hatch (Bless 1990). A similar pattern of colonisation was observed for the German part of the River Mosel, another major tributary of the Rhine (Fig. 1). Previous surveys have documented numerous populations of sculpins in small tributaries to the Mosel but never in riverine habitats (Pelz 1985; Pelz and Brenner 2000). In 2000 - 2001, we found unprickled sculpins to be restricted to small tributaries, while prickled sculpins were abundant in the main channel of the Mosel (Fig. 1). Finally, the explosive spread of sculpins in the Netherlands described by de Nie (1997) was apparently due to the spread of prickled sculpins as well, since the specimens we examined from Lake IJsselmeer were prickled sculpins.

In order to resolve the origins of the invasive fish and to uncover its population structure we have obtained samples from the Lower Rhine drainage and from the western Scheldt drainage. In the following, we will distinguish three groups of animals: the "Invasive sculpins" that have recently colonized the main stream and the large tributaries of the Rhine, the "Scheldt sculpin" from small upstream tributaries of the Scheldt drainage, and the "Rhine sculpin" from the small upstream tributaries of the Lower Rhine drainage (note that the sculpins in the small upstream tributaries of the Upper Rhine drainage belong to the Danubian phylogeographic group - Englbrecht *et al.* 2000).

Morphological and ecological analysis of all samples shows a clear differentiation between the Invasive sculpins and the Rhine sculpins and a less pronounced differentiation between the Invasive sculpins and the Scheldt sculpins (Fig. 2). A particularly prominent diagnostic difference between Invasive sculpins and Rhine sculpins is the occurrence of skin prickling (Fig. 2A) which is strong in the Invasive sculpins and virtually absent in the Rhine sculpins (Fig. 2B). However, Scheldt sculpins show also skin prickling comparable to that of the Invasive sculpins (Fig. 2B). A significant differentiation among the latter two groups can be observed in a geometric morphometric analysis (Fig. 2C). Invasive sculpins differ from adjacent Rhine sculpins by having a deeper trunk and caudal peduncle, a shorter body and a relatively larger head. Invasive sculpins differ from Scheldt sculpins in having a less deep anterior trunk (Fig. 2D). The anterior body depth of Invasive sculpins is intermediate between Scheldt sculpins and Rhine sculpins. Still, the average shape of Invasive sculpins is more similar to Scheldt sculpins as compared to Rhine sculpins (procrustes distance: 0.0217 vs. 0.0321 respectively). The morphological characteristics of the Invasive sculpins are retained in animals raised in aquaria and are thus not simply a plastic response to riverine habitats (Nolte *et al.* in prep.).

There are also life history differences between the groups. Female Rhine sculpins grow for at least two years before first reproduction. In contrast, Scheldt sculpins and Invasive sculpins start to reproduce in their first year (Fig. 2E). Both, Scheldt sculpins and Invasive sculpins compensate for the smaller size at reproduction by producing smaller eggs relative to their size than Rhine sculpins (Fig. 2F).

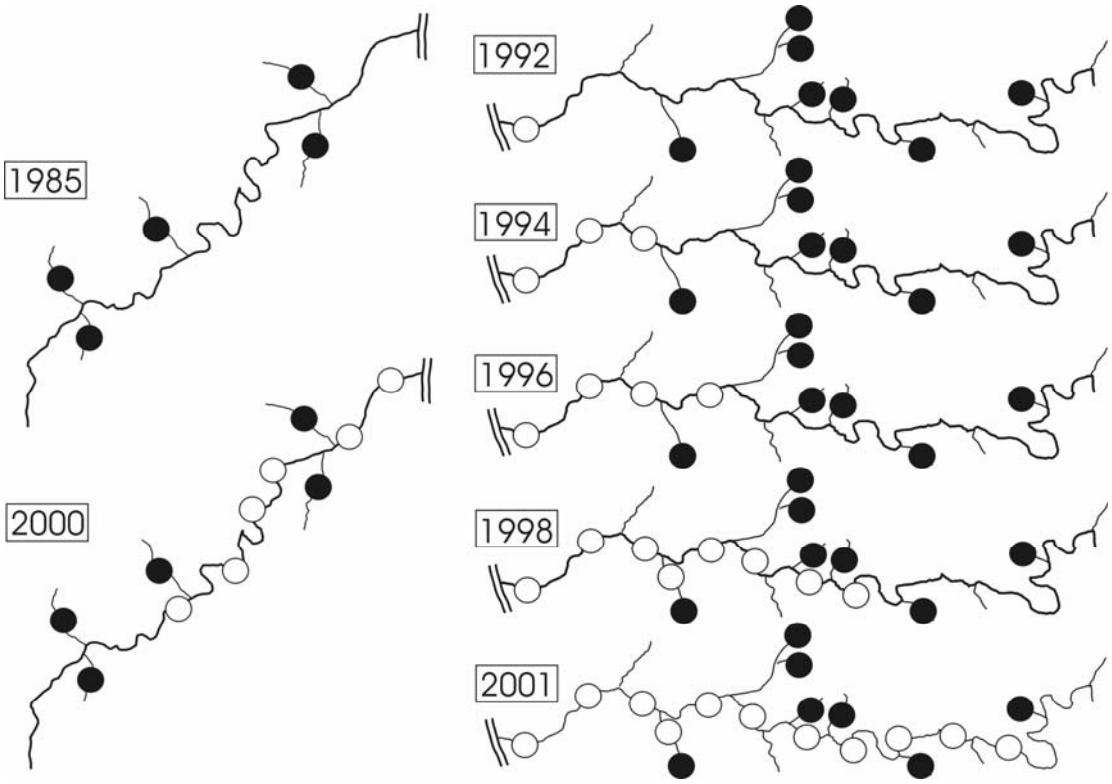
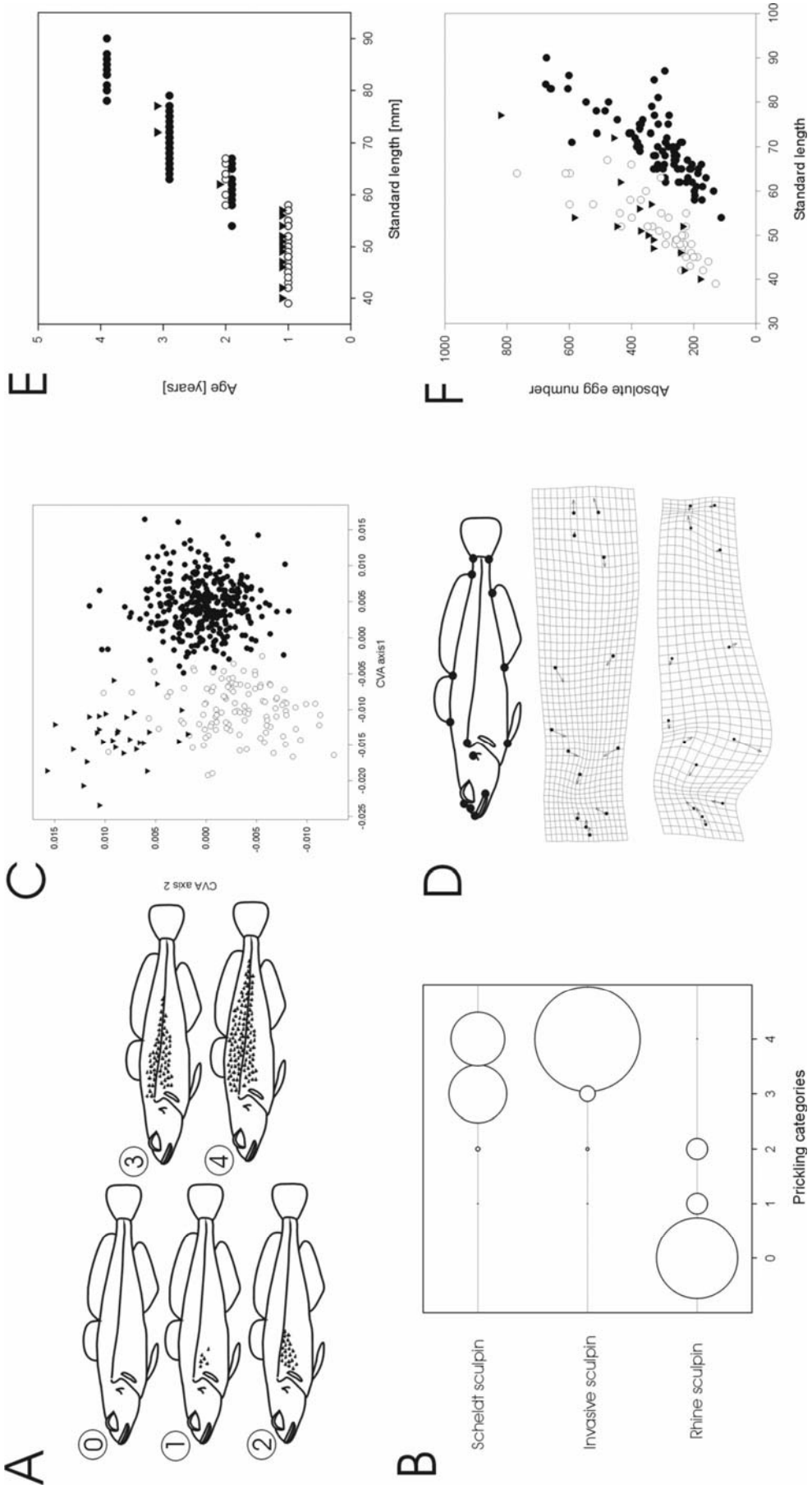


Figure 1: Colonization of rivers through Invasive sculpins in the recent past. Data from regular fish abundance surveys in the River Mosel (left) and the River Sieg (right) are depicted for selected years. Open circles indicate the increasing range of Invasive sculpins (as inferred from skin prickling) within main rivers, black circles represent the more or less static populations of non-prickled Rhine sculpins. The depicted length of the Mosel is approximately 200 km, the depicted length of the Sieg approx. 65 km, the double line depicts the Rhine.

Figure 2 (overleaf): Morphological and ecological analysis of sculpin samples. (A) Depiction of the five categories of spinelike scales covering the body (see Methods) and (B) frequencies for all groups (size of circle represents frequencies found). (C) Differentiation in body shape among the three lineages. Each comprises a distinct cluster that separates along two CVA axes (axis 1: Lambda = 0.13, chisq = 941.6, df = 48, $p < 0.01$; axis 2: Lambda = 0.62, chisq = 215.3, df = 23, $p < 0.01$). (D) The shape change captured by the CVA axes plotted as vectors at 14 anatomical landmarks (depicted in upper panel) on deformation grids (middle panel: Invasive sculpins vs Rhine sculpins; lower panel: Invasive sculpins vs. Scheldt sculpins). (E, F) Life history characters (age, fecundity, size) from reproducing females of Rhine sculpins (black circles), Invasive sculpins (open circles) and Scheldt sculpins (black triangles).


A
0
1
2
3
4
B
Scheldt sculpin
Invasive sculpin
Rhine sculpin
0 1 2 3 4
Prickling categories
C
CVA axis 2
CVA axis1
D
E
Age [years]
Standard length [mm]
F
Absolute egg number
Standard length

Invasive sculpins are hybrids

Given the hints from phylogeographic analysis of mitochondrial haplotypes (Englbrecht *et al.* 2000; Volckaert *et al.* 2002) of recent admixture of sculpins within the Lower Rhine, we developed nuclear diagnostic markers to trace a possible hybridization within the nuclear genome. Primers were constructed for randomly cloned genomic fragments and sequences were determined from 280 animals from an expanded dataset covering the major phylogeographic lineages that occur adjacent to the Rhine drainage (see methods for details). This allowed us to identify five single nucleotide polymorphisms (SNPs) that represent derived states and that were diagnostic in at least one of a known phylogeographic lineages (Fig. 3). These SNPs, as well as a further mitochondrial haplotype sequence analysis was then used to characterize the groups of sculpins described above (Tab. 1).

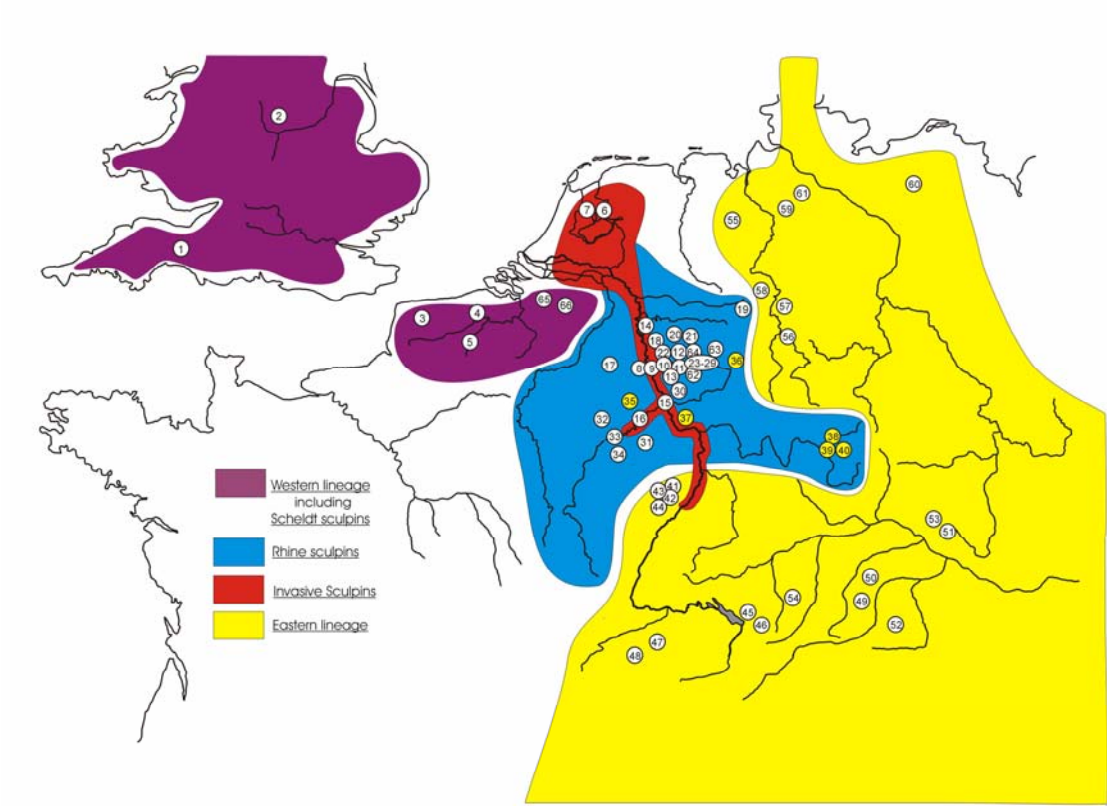


Figure 3: Distribution of evolutionary lineages of sculpins around the River Rhine basin as inferred from a combined analysis of mt-haplotypes, nuclear SNPs and literature data. The numbers refer to the sampling sites of fish used in this study. Note that there are further populations of sculpins in the West, which represent different lineages but which are not relevant for this study and are therefore omitted for clarity. Several populations (marked yellow) in tributaries to the middle Rhine or River Main carry intogressed haplotypes of danubian origin but belong to the Rhine sculpin lineage according to nuclear data (see text).

Rhine sculpins from upstream tributaries in the Lower Rhine drainage carry specific mitochondrial haplotypes (group I and III; this and the following haplotype groups sensu Englbrecht *et al.* 2000) and derived alleles at SNP

loci CgoSNP 1, 2 and 3a (Tab. 1). Scheldt sculpins carry group IV haplotypes and a diagnostic fixed allele at locus CgoSNP4 (Tab. 1). Sculpins from the upper Rhine drainage and from neighbouring drainages in the east of the Rhine show group I haplotypes and diagnostic fixed alleles at loci CgoSNP 3b and 5 (Tab. 1). In contrast to all stream populations, Invasive sculpins reveal a mixture of group I, III and IV haplotypes as well as a mixture of SNP alleles otherwise diagnostic for Rhine or Scheldt sculpins (Tab. 1). Thus, Invasive sculpins harbour a hybrid genome that is derived from Scheldt sculpins and Rhine sculpins.

Table 1: Diagnostic SNP alleles and mt-haplotype groups for different sculpin lineages. SNP alleles were classified as ancestral (a) or derived (d), based on outgroup comparisons (see Methods). The table provides frequencies for the respective populations, dashes indicate absence of character states. Invasive sculpins (bottom) show a combination of diagnostic alleles from Rhine sculpins (Lower Rhine streams) and Western sculpins (including Scheldt sculpins and samples from Great Britain)) (loci SNP1, SNP2, SNP3a and SNP4), but none of the diagnostic alleles of the Eastern animals (SNP 3b and SNP5). Note that the mt-haplotype group I sequences of Rhine sculpins are restricted to a few tributaries only (Fig. 3; see discussion).

population locations	N pop / N indiv typed	Phylogeographic lineage	locus Cgo SNP1	locus Cgo SNP2	locus Cgo SNP3a	locus Cgo SNP3b	locus Cgo SNP4	locus Cgo SNP5	haplo type group
41 - 61	21/72	Eastern Danubian sculpin	a: 1.0	a: 1.0	a: 1.0	a: -	a: 1.0	a: -	I: 1.0
			d: -	d: -	d: -	d: 1.0	d: -	d: 1.0	III: -
									IV: -
1-55; 65; 66	7/45	Western Scheldt sculpin	a: 1.0	a: 1.0	a: -	a: 1.0	a: -	a: 1.0	I: -
			d: -	d: -	d: 1.0	d: -	d: 1.0	d: -	III: -
									IV: 1.0
17-21; 24-32; 34-40	21/87	Lower Rhine Rhine sculpin	a: 0.14	a: 0.07	a: -	a: 1.0	a: 1.0	a: 1.0	I: 0.29
			d: 0.86	d: 0.93	d: 1.0	d: -	d: -	d: -	III: 0.71
									IV: -
6; 10; 15	3/76	Invasive sculpin	a: 0.67	a: 0.81	a: -	a: 1.0	a: 0.84	a: 1.0	I: 0.13
			d: 0.33	d: 0.19	d: 1.0	d: -	d: 0.16	d: -	III: 0.53
									IV:0.34

Invasive sculpins form a genetically distinct group

To test for a possible population substructure within the lineages, we typed six microsatellite loci for 950 individuals from 23 sampling sites. We find that the three groups of sculpins form three corresponding genetic clusters (Fig. 4). Most notably, the Invasive sculpins from the IJsselmeer, Mosel and Sieg cluster together, suggesting that they form a genetically homogeneous group. The genetic differentiation (Fig. 4a) among different

populations within the Rhine sculpin cluster is greater than within the Invasive sculpin cluster (average D: 0.47, range: 0.22-0.77 vs. 0.12, range: 0.02-0.24), which indicates a shallower population substructure of the Invasive sculpin. This is in line with the fact that Rhine sculpins occur in isolated subpopulations in separate streams whereas Invasive sculpins are interconnected in their distribution (Fig. 3).

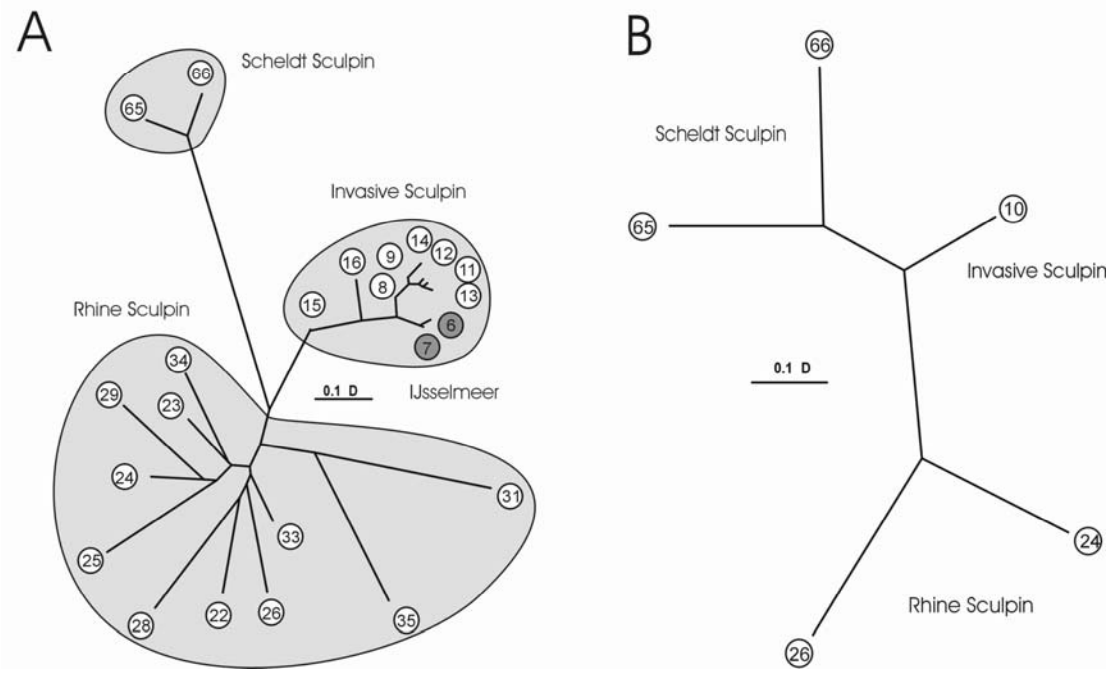


Figure 4: Neighbour-joining trees of genetic distances (Nei's standard distance) among sculpin populations based on analysis of six microsatellite loci. (A) Tree based on six loci from 23 sampling sites and 14-48 individuals per site. The clustering is congruent with the combined SNP/haplotype inference as depicted on Figure 3. Populations of stream sculpins are genetically more subdivided than invasive sculpins. (B) Tree based on 120 loci from five sampling sites and 12-48 individuals per site, representing the ancestral lineages and invasive sculpins. This shows that invasive sculpins are not particularly similar to either one of the presumed ancestral lineages on the genome level.

To assess the relative genomic contribution of the different lineages within the hybrid gene pool, we typed a smaller number of populations for a larger number of microsatellites. For this we used a population of Invasive sculpins from the Sieg, two populations of Rhine sculpin and two populations of Scheldt sculpin (see methods) and typed them for 120 microsatellite loci. Genetic distance (D, as above) analyses revealed the same grouping as in Fig. 3 and more importantly, places the sample of the Invasive sculpin similarly distant to both the Scheldt and the Rhine sculpins (Fig. 4b - average D to Rhine sculpins: 0.62; average D to Scheldt sculpins: 0.46). This result

suggests an almost equal mixture of the source gene pools in the Invasive sculpins on average. However, this statement needs to be cautioned, since the primary source populations that have been involved in the initial hybridization event are not necessarily those that we have sampled here.

Discussion

Hybridization between distinct populations or between closely related species is well known to occur both in animals and plants (Arnold 1997). Usually the evidence stems from non-concordant phylogenies of nuclear and mitochondrial markers (Adams *et al.* 2003; Avise 2000; Rognon and Guyomard 2003), However, it has so far only been shown for plants that rapid adaptations to new habitats and corresponding new colonisations can be caused by hybridization (Rieseberg *et al.* 2003). Still, it is a priori likely that the same should also occur in animals. As a possible example, Schliewen and Klee (2004) provide evidence for a hybrid species with very distinct ecology among the sympatrically evolved cichlid species flock in the crater lake Barombi Mbo in Cameroon. Seehausen (2004) has recently presented a model where he considers invasion of new areas and habitats as a trigger for hybridization between lineages and species, which may result in new lineages with a new adaptive potential. This process is thought to contribute greatly to speciation in those systems. Our observations on sculpins suggest also that a new lineage has emerged from hybridization of long separated groups and most importantly, that this new lineage indeed has a novel adaptive potential that is absent in its ancestors. This is apparently a very recent and still dynamic process that we observe directly, which makes it very different from cases where the event of hybridization lies back in time. It is also somewhat different from situations where artificial introductions have lead to fast adaptations in the respective new environment (Hendry *et al.* 2000; Streelmann *et al.* 2004). In our case, hybridization and new adaptation not only occurred in the same lineage but also appear to have emerged jointly. In the following we evaluate the molecular phylogeographic context, the evidence for specific adaptations and propose a scenario for the course of the hybridization.

Molecular phylogeographic context

The new phylogeographic data presented here including those from mitochondrial haplotypes as well as SNPs are largely in line with previous inferences on the pan-european population structure of sculpins (Englbrecht *et al.* 2000; Volckaert *et al.* 2002). However, there is an interesting deviation. Animals of the upper River Main as well as some tributaries to the middle Rhine (Pops. 35, 36, 37) harbour eastern (Danubian; group I) mitochondrial haplotypes, but belong to the Rhine sculpins based on SNP analysis. Such a grouping with populations from the Lower Rhine drainage was already reported by Hänfling *et al.* (2002). It seems likely that a mitochondrial introgression has occurred in previous times in these populations, which has by now lead to the fixation of typical eastern group I haplotypes in a nuclear genome background of the Rhine sculpins. For the context of the present

study this is only important with respect to understanding the rare occurrence of the eastern group I haplotypes in the Invasive sculpin (Tab. 1), while no corresponding eastern SNP alleles were found in these animals, i.e. the occurrence of eastern mt-haplotypes in the invasive sculpin genepool does not require a direct involvement of danubian ancestors.

Apart of this slight complication with the group I mitochondrial haplotypes, the pattern revealed in our study is remarkably clear. A mixture of haplotypes and diagnostic SNP alleles is only seen in animals from riverine habitats of the Rhine drainage, while populations representative of the parental lineages persist in small headwater streams throughout the study area. The hybridization that has formed the hybrid invasive genepool must have occurred very recently, since we have not found private SNP alleles in these hybrids.

The group IV haplotypes in the Invasive sculpin gene pool are identical to haplotypes H41 and H42 that otherwise occur in Scheldt sculpins within the River Nete (Volckaert et al 2002; Knapen *et al.* 2003, our study). Intriguingly, exactly this tributary to the River Scheldt is the closest to the Lower Rhine and was one of the first that has become connected to the Rhine system via manmade canals, starting about 200 years ago. In contrast to the group IV haplotypes, the group III haplotypes provide no hint for the origin of the source population of the hybrids since only broadly distributed common haplotypes are found.

The hybrids form a distinct and rather homogeneous lineage in comparison to representatives of the parental lineage according to the microsatellite data (Fig. 4 and 5). Furthermore, populations of the Invasive sculpin are largely in Hardy Weinberg equilibrium (data not shown) and represent a morphologically homogeneous group (Fig. 2). Thus, the Invasive sculpin represents a homogeneous population and not simply a mixed pool of individuals from different origins.

Morphology and adaptations

The animals invading the Lower Rhine drainage were morphologically identifiable by their strong prickling. Since we know now that this is characteristic for the hybrid lineage, we can conclude in retrospect, that the Rhine sculpins from streams flowing into major rivers remained confined to their native habitats and did not expand their range, which is in perfect agreement with our field data. Thus, despite their geographic proximity and direct waterway connection, only the Invasive sculpin was able to colonize the vacant habitats. This pattern can only be explained by an autecological advantage of Invasive sculpin over Rhine sculpins in large river habitats. While the exact nature of this advantage remains unknown, the most inclusive measure of overall fitness, namely thriving populations vs. apparent inability to survive in large rivers provides the best general evidence for differential adaptation.

Comparable systematic survey data as for the Sieg and the Mosel do not exist for the main river of the Lower Rhine. Nevertheless, samples from the Mosel (# 16), the confluence of the Mosel and the Rhine at Koblenz (# 15), the Sieg and the Rhine at Bonn (# 8, 9 and 11 – 13), the confluence of the Düssel and the Rhine at Düsseldorf (# 14) and the IJsselmeer in the Netherlands (# 6, 7) all correspond to the Invasive sculpin (Fig. 3, 4). Moreover, sculpins from the Dutch lowlands show the mixture of mitochondrial haplotypes (Volckaert *et al.* 2002; this study) characteristic for the Invasive sculpin and it is therefore likely that these are representatives of the invasive lineage, although the SNP data to confirm this are not available.

In contrast to the Rhine sculpin, only circumstantial evidence for an ecological differentiation of the scheldt sculpin and the invasive sculpin is available. First, a population expansion or invasion of the main River Scheldt has not been described so far and past survey data suggest that Scheldt sculpins are usually confined to headwaters of small tributaries (Vandelannoote *et al.* 1998). Knapen *et al.* (2003) have analysed the population genetic structure of Belgian sculpin populations with a partially identical set of microsatellite markers that was used for this study. They found an equally or even more pronounced population substructure among Scheldt sculpins than we observe here for Rhine sculpins. This confirms that the overall population structure of Scheldt sculpins resembles that of Rhine sculpins in that isolated subpopulations are restricted to headwaters of small tributaries with no regular gene flow between separate streams. With respect to invasive sculpins this implies that Scheldt sculpins from small tributaries constitute the ancestral population that has no pronounced ability to colonise riverine habitats.

A hybridization scenario

All major European rivers like the Rhine and Scheldt as well as their large tributaries such as the Mosel have been channelized, dredged and dammed in the past centuries to create waterways suitable for large ships. In addition, the waterways were fortified with rocks which act as microhabitats for sculpins (Knaepkens *et al.* 2002). Furthermore, the Rhine Delta and large bays (e.g. the IJsselmeer) were cut off from the open sea eliminating tide and saltwater as ecological key factors in large areas. In this way a novel interconnected system of new habitats has become available through human activities. The Invasive sculpin has originated in the lowest reaches of the Rhine drainage and has then spread within the Netherlands and upstream the Rhine. The fact that there are no known records of the Invasive sculpin in the Scheldt drainage yet, suggests that there might have been an initial unidirectional migration of Scheldt sculpins via the newly built canals towards the Rhine drainage, where the hybridization began. Because of the major changes to the river systems, one can only speculate of how these lineages would have originally met. An initial phase would require at least short-term success within restricted areas and could have been highly dependent on transient ecological conditions or chance. It is known that larvae of European sculpins can enter the open water (Wanzenböck *et al.* 2000) and can be washed downstream after they hatch (Bless 1990). Such drifting larvae might

survive for some time in some parts of those riverine habitats, which were changed by human activities and could thus have contributed to matings with sculpins from other lineages. This drift effect could contribute to homogenize the hybrid genepool and would explain why the eastern mitochondrial haplotypes from middle rhine tributaries occur in the hybrid gene pool, even in the Ijselmeer, without a direct contact of the respective source populations. Thus, over time a population of hybrid animals and backcrosses would have become established somewhere in the Lower Rhine area which eventually obtained the capacity to spread into new habitats.

From plant studies it is known that such hybridization is a potential source of new genotypes from which selection could pick favourable heterospecific gene combinations to allow adaptation to new environments (Rieseberg *et al.* 2003). Barton (2001) has recently reviewed the role of hybridization in generating new adaptations. The respective models usually assume that hybrid genotypes are less fit than the parental genotypes, although some of the very large number of possible backcross hybrid genotypes may be fitter than either parent. This effect will be particularly pronounced under conditions where habitat changes occur for the parental lineages, as it would have happened in the Rhine delta in the past centuries. Under such conditions, a hybrid population exploiting a combined pool of traits could adapt to new habitats that are not available to either parental lineage. This process of hybridization, backcrossing and adaptation may have gone on for some time, before a distinct new lineage emerged, which constitutes now the invasive population.

An alternative scenario would posit that the Scheldt animals have started to spread along the Rhine before the hybrid gene pool was established, but would have picked up alleles from Rhine animals at contact sites in the tributaries. Such contact sites do indeed exist (Chapter 4), but this scenario is still difficult to reconcile with the data. Given that the spread of the Invasive sculpins has occurred within two decades, there would not have been enough time to allow the degree of mixing of the gene pool along the Rhine that we observe. For example, we would not expect in this scenario to find such a close grouping of genotypes from the Middle Rhine and the IJsselmeer (compare Fig. 4), since the allelic influx would have come from different source populations. Also, this scenario cannot explain why the Scheldt animals should have spread in the first place, given that they have been confined to non-river habitats since a long time. The scenario of an initial formation of a hybrid genepool with new adaptations, before the spread began, remains therefore more likely. Still, the possibility of further introgression of alleles from the stream populations remains. We have recently completed the draft of a genome map for *Cottus* (Stemshorn *et al.* 2005), which will allow to map regions of the genome that have introgressed.

Acknowledgements

This work was possible only with permissions from A. Mellin, T. Heilbronner, W. Fettweis, C. Bode, L. Jörgensen, L. Kroll, C. Köhler, W. Fricke, H. Arzbach and M. Kämmereit. We thank F. Volckaert, B. Hänfling, D.

Neely, L. Bervoets, G. Knaepkens, E. Winter, I. Steinmann, E. Korte, U. Weibel, C. Dümpelmann, W. Dönni, C. von Landwüst, A. Waterstraat as well as numerous local fishermen for support, samples and help with the fieldwork. We thank B. Harr and J. Bohlen for suggestions on the manuscript. This study was financed by the Deutsche Forschungsgemeinschaft and Verband der Chemischen Industrie.

Literature

Adams, J. R.; Leonard, J. A.; Waits, L. P. (2003) Widespread occurrence of a domestic dog mitochondrial DNA haplotype in southeastern US coyotes. Mol. Ecol. 12, 541-546.

Arnold, M. L. (1997) Natural Hybridization and Evolution. Oxford University Press, New York.

Avise, J. C. (2000) Phylogeography: The History and Formation of Species. Harvard Press.

Barton, N. H. (2001) The role of hybridization in evolution. Mol. Ecol. 10, 551-568.

Bless, R (1990). Die Bedeutung von gewässerbaulichen Hindernissen im Raum-Zeit-System der Groppe (*Cottus gobio* L.). Natur und Landschaft 65: 581-585.

Brink van den F. W. B. Velde van der, Cazemir, W. G. (1990) The faunistic composition of the freshwater section of the River Rhine in the Netherlands: present state and changes since 1900. Limnologie aktuell 1: 191-207.

Cazemir, W. G. (1988). Fish and their environment in large European river ecosystems. The Dutch part of the River Rhine. Sciences de l´eau 7: 97-116.

De Nie, H. W. (1997) Atlas van de Nederlandse Zoetwatervissen. Media Publishing, Doetinchen, 151 pp.

Dieringer, D. Schlötterer, C. (2003) microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3 (1), 167-169

Englbrecht, C.C. Largiader, C.R. Hänfling, B. Tautz, D. (1999) Isolation and characterization of polymorphic microsatellite loci in the European bullhead *Cottus gobio* L. (Osteichthyes) and their apllicability to related taxa. Mol. Ecol. 8, 1966 - 1969.

Englbrecht, C. C. Freyhof, J. Nolte, A. Rassmann, K. Schliewen, U. Tautz, D. (2000) Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-pleistocene origin of the major central European populations Mol. Ecol. 9, 709-722.

Hänfling, B. Hellemans, B. Volckaert, F. A. M. Carvalho, G. R. (2002) Late glacial history of the cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. Mol. Ecol. 11, 1717-1729.

Hendry, A.P. Wenburg, J.K. Bentzen, P. Volk, E.C. Quinn, T.P. (2000). Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. Science 290, 516-519.

Knapen, D. Knaepkens, G. Bervoets, L. Taylor, M. I. Eens, M. Verheyen, E. (2003) Conservation units based on mitochondrial and nuclear DNA

variation among European bullhead populations (*Cottus gobio* L. 1758) from Flanders, Belgium. Conservation Genetics 4, 129-140.

Knaepkens, G. Bruyndonc, L. Bervoets, L. Eens, M (2002) The presence of artificial stones predicts the occurrence of the European bullhead (*Cottus gobio*) in a regulated lowland river in Flanders (Belgium). Ecology of Freshwater Fish 11, 203–206.

Koli, L. (1969) Geographical variation of *Cottus gobio* L. (Pisces, Cottidae) in Northern Europe. Annales Zoologici Fennici 6, 353-390.

Köhler, C. Lelek, A. Cazemir, W. G. (1993). Die Groppe (*Cottus gobio*) im Niederrhein - Merkwürdigkeit oder etablierter Bestandteil der Fischartengemeinschaft? Natur und Museum 123 (12), 373-386.

Lelek, A. Köhler, C. (1993) Erfassung der Fischfauna des deutschen Niederrheinabschnittes mit besonderer Berücksichtigung der Rheinsohle im Hauptstrom. Studie für die Rheinfischereigenossenschaft, Bonn, 27 pp.

Nolte, A. W. Stemshorn, K. C. Tautz, D. (2005). Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Mol. Ecol. Notes (*in press*)

Pelz, G. R. (1985): Fischbewegungen über verschiedenartige Fischpässe am Beispiel der Mosel. - Cour. Forsch.-Inst. Senckenberg 76:1-190.

Pelz, R. Brenner, T. (2000). Fische und Fischerei in Rheinland-Pfalz: Bestandsaufnahme, fischereiliche Nutzung, Fischartenschutz. Ministerium für Umwelt und Forsten Rheinland-Pfalz. 258 pp.

Rieseberg, L. H. Raymond, O. Rosenthal, D. M. Lai, Z. Livingstone, K. Nakazato, T. Durphy, J. L. Schwarzbach, A. E. Donovan, L. A. Lexer, C. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301, 1211-1216.

Rognon, X. Guyomard, R. (2003) Large extent of mitochondrial DNA transfer from Oreochromis aureus to O. niloticus in West Africa; Mol. Ecol. 12, 435-445.

Rohlf, F. J. Marcus, L. F. (1993) A revolution in morphometrics. Trends Ecol. Evol. 8, 129-132.

Rohlf, F. J. (2003) TpsDig, digitize landmarks and outlines, version 1.39. Department of Ecology and Evolution, State University of New York at Stony Brook.

Schleuter, M. (1991) Nachweis der Groppe (*Cottus gobio*) im Niederrhein. Fischökologie 4, 1-6.

Schliewen, U.K. Klee, B. (2004). Reticulate sympatric speciation in Cameroonian crater lake cichlids. Frontiers in Zoology, 1:5

Schneider, S. Roessli, D. Excoffier, L. (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Seehausen, O. (2004) Hybridization and adaptive radiation Trends Ecol. Evol. 19 (4): 198-207.

Sheets, H. D. (2002) IMP-Integrated Morphometrics Package, Deptartment of Physics, Canisius College, Buffalo, Ney York.

Stemshorn, K. C. Nolte, A. W. Tautz, D. (2005). A Genetic Map of *Cottus gobio* (Pisces, Teleostei) based on microsatellites can be linked to the Physical Map of *Tetraodon nigroviridis*. J. Evol. Biol (*in press*)

Streelman, J.T. Gmyrek, S.L. Kidd, M.R. Kidd, C. Robinson, R.L. Hert, E. Ambali, A.J. Kocher T.D. (2004). Hybridization and contemporary evolution in an introduced cichlid fish from Lake Malawi National Park. Mol Ecol 13, 2471-2479

Vandelannoote A, Yseboodt R, Bruylants B, Verheyen R, Coeck J, Belpaire C, Van Thuyne G, Denayer B, Beyens J, De Charleroy D, Maes J, Vandenabeele P (1998) Atlas van de Vlaamse beeken riviervissen. WEL v.z.w.Wijnegem.

Volckaert, F. A. M. Hänfling, B. Hellemans, B. & Carvalho, G. R. (2002). Timing of the population dynamics of bullhead *Cottus gobio* (Teleostei: Cottidae) during the Pleistocene. J. Evol. Biol. 15, 930-944.

Vogt, C. Hofer, B. (1909) Die Süsswasserfische von Mittel-Europa. Engelmann, Leipzig, xxiv+558 pp.

Wanzenböck J. B. Lahnsteiner & K. Maier (2000). Pelagic early life phase of the bullhead (*Cottus gobio* L.) in a freshwater lake. Journal of Fish Biology 56: 1553-1557.

Appendix – Chapter 1

Chapter 1 - Supplementary Table 1: Sampled Populations, localities with coordinate data, river basins and references to other studies.

Chapter 1 - Supplementary Table 2: SNP Loci, primers and variable sites with ancestral and derived states in phylogeographic lineages of the *Cottus gobio* complex.

Chapter 1 - Supplementary Table 3: Individual mitochondrial haplotype group affinity and SNP genotypes of all specimens analysed (missing values indicated by "?").

Chapter 1 - Supplementary Table 4: Individual microsatellite Genotypes with Lineage affinity of specimens analysed in the multilocus distance tree (missing values indicated by "0"). Data format: Table; saved row by row with fields separated by semicolons. Ends of rows are marked by the insertion of “XXX”.

Chapter 2: Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure

Arne W. Nolte, Kathryn C. Stemshorn and Diethard Tautz

Abstract:

Applying a simplified enrichment procedure we have isolated and characterized 177 microsatellite markers for *Cottus gobio* L. 1758 (Cottidae; Scorpaeniformes; Teleostei). In contrast to using specific probes for the enrichment we use genomic DNAs of unrelated organisms for cross-hybridization. This takes advantage of the fact that simple sequences are the only repetitive elements that are abundantly found in all eukaryotic genomes and that any genome usually contains all permutations of microsatellite motifs. This cross-hybridization principle was employed to enrich genomic libraries of *Cottus* DNA to obtain a large number of non-redundant microsatellite markers without further screening procedures.

Microsatellites from Sculpins

Sculpins (family Cottidae, genus *Cottus*), are small benthic fish that are widespread throughout the northern hemisphere. They have received attention for their pronounced population substructure and the presence of ancient phylogeographic lineages (e.g. Englbrecht *et al.* 2000; Volckaert *et al.* 2002). We are currently developing them into a model species for evolutionary ecological studies with the aim to understand the genetic architecture of divergence. However, the number of microsatellites available (Englbrecht *et al.* 1999) was limited.

To obtain more loci, we have employed an enrichment strategy that relies on cross hybridization of simple sequences between distantly related organisms. It has long been known that simple sequences are universally present in diverse eukaryotic genomes (Tautz and Renz 1984). Thus, instead of using a panel of synthetic probes of all permutations of simple sequences for enrichment procedures, we use directly the genomic DNA of an unrelated organism as probe.

To prepare hybridization filters, genomic DNA from a mouse (*Mus musculus domesticus*) or a crayfish (*Procambarus clarkii*) was fixed to small nylon membrane fragments (10 x 5mm, Hybond N+, Amersham Bioscience). A 100µg/ml solution of genomic DNA was mixed 1:1 with 1M NaOH and dripped onto the pieces of the membrane. After 2-5 minutes these were transferred for 2 min into neutralization buffer (1M Tris-HCl pH 7.0) followed by two 2 min washes in 2xSSC buffer. The membrane chips were then dipped

on tissue to remove excess liquid, dried at 80°C over night and stored dry until use.

Prior to enrichment total genomic DNA of *Cottus gobio* was digested with Tsp509 and size selected for fragments between 500 and 1000bp by double separation on a 1.5% agarose gel. The fragments were cloned into EcoRI digested LambdaZAP (Stratagene). Using ExAssist helper phages, clones were mass excised and single stranded DNA was prepared according to the protocol of the supplier (Stratagene).

For the enrichment, the hybridization chip was preincubated for 15 min in hybridization buffer (5x SSC, 0.5% SDS, 50µg/ml Heparin, 1x Denhardt's solution) at 50°C. Approximately 500ng of single stranded plasmid was heated for 5 min at 95°C in 1 ml of hybridization buffer with the prehybridised chip added to the tube. Temperature was slowly dropped to 50°C over a period of 30 min under gentle agitation. For washing, the chip was transferred to 40 ml hybridization buffer and incubated at 50°C under gentle agitation for 15 min. These washes were repeated another two times with hybridization buffer and three times for 10 min with 0.1x SSC, 0.1% SDS buffer at 50°C. The chip was removed with forceps, gently dipped on a tissue to remove excess liquid and transferred to 500µl TE buffer. This was heated to 80°C for ten minutes to liberate the hybridized DNA. The chip was then removed and the DNA was precipitated. A primer extension was performed using plasmid specific oligonucleotides and Klenow DNA polymerase prior to cloning into electrocompetent SOLR™ cells (Stratagene). Positive clones were sequenced without further screening.

An alternative enrichment approach was derived from the FIASCO protocol (Zane *et al.* 2002). Genomic fragments from an MseI digest / adapter ligation reaction were subjected to a 10 min incubation at 72°C in a complete PCR buffer to close gaps remaining from the adapter ligation. Then fragments were size-selected and 200 – 500ng were enriched as above. The pool of enriched fragments was PCR amplified, ligated into pZeroII™ vector (Invitrogen) and cloned into electrocompetent Top10™ cells (Invitrogen). Positive clones were sequenced without further screening.

For fragments carrying microsatellites, primers were designed using FastPCR (http://www.biocenter.helsinki.fi/bi/bare-1_html/oligos.htm). The optimal primer annealing temperature was tested on a temperature gradient thermal cycler. Loci that were found to be polymorphic in a prescreening were then scored for a set of 480 *Cottus* specimens from three populations from the River Sieg (Germany) and for 165 offspring from laboratory crosses. These data allow to determine allele numbers and confirm mendelian inheritance for all markers (Tab. 1). Deviations from Hardy-Weinberg equilibrium as well as observed and expected heterozygosities (Tab. 1) were determined for 35 specimens from a single site (Lower Sieg; GIS 50°47'N 7°10'E) using Arlequin (Schneider *et al.* 2000). A significant deficiency of heterozygotes was detected for 17 loci (Tab. 1), which suggests the presence of null alleles. However, we retain these loci as we have demonstrated mendelian inheritance for them in laboratory crosses.

The single stranded plasmid enrichment (approach I) yielded about 50% clones with microsatellite repeats while the PCR based protocol (approach II) was slightly more efficient (60-85% repeat containing clones). Enrichment efficiency was similar when mouse DNA or crayfish DNA was

used as probe. Most repeats obtained were dinucleotide repeats, but we also found several clones with higher order repeat units (Tab. 1). Very few clones were found in duplicate, i.e. a large diversity of loci was sampled.

We have tested the enrichment protocol for additional species. In other fish (Telmatherinidae; Nemacheilinidae; Cichlidae) we have obtained a comparably high percentage of microsatellite containing clones. For an insect (*Panorpa communis*) and a bird (*Luscinia megarhynchos*) the efficiency was lower (approximately 10% positive clones). Very low efficiency was obtained for Salamanders (*Salamandra salamandra*), which have apparently an unusual genome structure, which requires much more stringent enrichment procedures (Steinfartz *et al.* 2004). Thus, we recommend the use of the simplified procedure only for organisms in which microsatellites are not particularly rare, which is, however, the case for the majority of organisms.

Table 1: Characteristics of 177 microsatellite loci in *Cottus gobio* with primer sequence, repeat motif, cloned size and optimal annealing temperature (T_A). Total number of alleles (N_A) is given for 480 individuals belonging to 3 different populations except for those marked with * (sample size here is 167 individuals belonging to one full-sib and two half-sib families); observed heterozygosity (H_O) with **statistically significant deviation from Hardy-Weinberg equilibrium ($p < 0.001$), and expected heterozygosity (H_E), n.e. = not evaluated.

Locus	Repeat motif	Primersequences (5'-3')	Size bp	T_A (°C)	N_A	H_o	H_e	GenBank Accession no.
Cott100	$(TG)_{15}$	CTCATCGTGGTTTGATCGGTG	176	60	10	0.48571	0.54244	AY861165
		CCGAGCGTGAGTCAGGCGTG						
Cott105	$(CA)_{15}$	TCCTACAGGGTGCGATCGTG	341	60	12	0.85714	0.81573	AY861166
		TGCAGGAGTCAGGACTCTGC						
Cott112	$(TG)_{14}$	TCATGTGGGATGGAGCTCTG	149	60	28	0.91429	0.94327	AY861167
		TCACCGCATTACTGTTGCAC						
Cott113	$(TG)_{15}$	AGCGCCAGAATGCAGCATCC	167	60	18	0.65714	0.81035	AY861168
		AGTGTGGCGAGCCCAAGATC						
Cott118	$(AC)_{14}$	ACTGGTCTCCAGGCGGTGTC	410	60	12	0.54286	0.56398	AY861169
		GACGCCGTCATGCTCAGGTC						
Cott119	$(GT)_{9}$	ACCGCAATGTGGACTTGACC	289	60	7	0.74286	0.67453	AY861170
		AGGTGACCAGCTCAGCGGTC						
Cott127	$(AC)_{14}$	TTCAGCTGCCTGTGTAAGGC	194	60	11	0.85714	0.76439	AY861171
		TGTATGGAGCACTTGGGCTG						
Cott128	$(TG)_{14}$	TCTGTGGGTGTTTGGTCGTG	318	60	10	0.65714	0.80911	AY861172
		TGAACTCTGCACATGACTGC						
Cott130	$(GT)_{13}$	TCTGGATCCCTCGGACCAGG	146	60	7	0.91429	0.80083	AY861173
		TGAGCTCCATCGTGGGTTCG						
Cott132	$(AC)_{13}$	AGTTTGTCACACACATAGGCAC	216	60	6	0.48571	0.72671	AY861174
		GTTTGTGTGCATGCGTTGTG						
Cott138	$(CA)_{13}$	ACGTGTAGGGCCGCAGGACC	267	60	12	0.85714	0.7764	AY861175
		ACGGCTGCCAGGGAAACCTG						
Cott144	$(CA)_{12}$	TTGGATCGGTGCCTTCACAG	156	60	9	0.74286	0.76439	AY861176

		TAGGGACGACAGCCAGCTCC						
Cott146	$(TG)_{12}$	GCTTCAGGGTAGCGTGTGTG	224	60	11	0.91429	0.87867	AY861177
		TCAGGCGGCAAGCTCTGGTG						
Cott149	$(CA)_{12}$	CATCACTGGACAATGACGGC	332	60	13	0.76471	0.80773	AY861178
		AAGACCTCTGAGGTGTTTCCA						
Cott150	$(GT)_{12}$	ATCTCTCAGCCATGAGCGAC	168	60	8	0.71429	0.73913	AY861179
		ACAGCTGTGCTTTGTGCAGC						
Cott152	$(GT)_{10}N_{49}(GT)_8$	AGAGCGGTAGACACTCGCCA	173	60	6	0.87879	0.82424	AY861180
		TGGCCTCCTCAGGCTTGTTC						
Cott153	$(GT)_{12}$	GGGTTCAGGTGCGTTTGGTC	247	60	4	0.36364	0.42471	AY861181
		ACCTTCAGCTCCGACATCAG						
Cott154	$(AC)_{11}N_{10}(CA)_7$	TCGTCGGGAGTGAGGCGTCC	274	60	15	0.79412	0.83582	AY861182
		TGCTTTCGATCAACATGGCTG						
Cott158	$(AC)_{12}$	TGACGTTAGCGCGTGTTGAG	209	60	9	0.51429	0.58385	AY861183
		AGTCTTGCTTTGTGCGATGTC						
Cott163	$(AC)_{11}$	TGTTCAGGGCTGACTTGCTG	157	60	13	0.60606	0.72075	AY861184
		CATAGGGCTTGTGTCTGAGC						
Cott164	$(TG)_{12}$	AGGAGAATGGCCAGACAGAC	126	60	16	0.82353	0.8266	AY861185
		ACTGCCCTCGTACCGGAGCA						
Cott170	$(CA)_{10}$	ACATGGTGCATAATGTTGCCC	125	60	14	0.8	0.72878	AY861186
		TCAGGGCCTACGCACAGAGC						
Cott173	$(AC)_{16}$	TGAAGTCAGCATGCCTACCA	162	60	5	0.66667	0.5366	AY861187
		GAACCCTGTGTGGGTGGGTC						
Cott175	$(TG)_{11}$	TGTGTGAAGGGTGGTTGCTC	341	60	15	0.48485	0.5627	AY861188
		AGGCAGCTACGTTGAGGTCC						
Cott176	$(TG)_{10}CG(TG)_6$	AGCGAGGTGACCTGGGTGTC	199	60	22	0.77143	0.87495	AY861189
	$CGTGCG(TG)_6$	ATTCAGTGGCCTGCCATGAA						
Cott179	$(GT)_{10}$	AGCGTGGCTCAGCCTGGTGG	135	60	9	0.68571	0.73043	AY861190
		TCCAGGGCTAAGTGCTGCTC						
Cott183	$(GT)_{10}$	AGATGACCAGGGCTCTATGC	310	60	4*	n.e.	n.e.	AY861191
		GTGGGAGCCGCCTCGTACAG						
Cott184	$(TC)_{12}N_{20}$	GGATTAGGCTGGTTGGTGTC	193	60	5	0.54286	0.49027	AY861192
	$(TC)_7$	TAACGCAGCACCTGCAGAGC						
Cott188	$(CA)_{10}$	TCGCCTGCAGGGAGATGCTT	181	60	5	0.57143	0.6472	AY861193
		TCATCCGTGTTTCAGGAGCA						
Cott194	$(TCA)_8$	TCTTCGTCCTGTCTAGCTGC	144	60	7	0.54286	0.47702	AY861194
		AGGTTGAACGACTGCTCGTC						
Cott205	$(CT)_7N_{12}$	TTGACGGTCAAATGTCACTGTG	324	60	7	0.68571	0.72919	AY861195
	$(CT)_{20}$	TCCCGAACATGACTGCGGTG						
Cott207	$(GT)_9$	AGTCCTTGTCGGGAGCCTCG	291	60	46	0.8	0.90393	AY861196
		ATTGGGCGTTGCTCACCAGC						
Cott210	$(GT)_{10}$	TGAGATGCTCTAGATTGCTCAGTG	208	60	7	0.55882	0.64794	AY861197
		CATGTGCTCAAAGACAGTCACG						
Cott214	$(AC)_9$	AGTAGTGCCGCTGCACCCAC	252	60	6	0.75758	0.68904	AY861198
		TCTCCCTCGTGGTGCAAACC						
Cott221	$(CT)_{15}N_{30}(CA)_7$	GCAGGAACTTCACACCGCCA	269	60	3*	n.e.	n.e.	AY861199
		GCTCATTGACTGTGAACTGGGT						
Cott224	$(CT)_{19}$	CCACTGTACTGACTCAGGCAG	165	60	23	0.65714	0.79503	AY861200
		CCATGGATCACTTGGAATGCAC						

Cott228	$(GA)_{11}$	AGGGCATCCACGTGTCATGC GCTGTCTGCACAACCTGCAG	167	60	10	0.62857	0.71636	AY861201
Cott250	$(CT)_{11}$	TCTGGTGGATGTGCGTGTCC TGAATGTGCTTCCGTGTCAG	192	60	5	0.57143	0.71139	AY861202
Cott255	$(CT)_{10}$	TCACTACAGCCAGGTGTCTG GCATGTGCATGCCTCCACAG	172	60	9	0.62857	0.61242	AY861203
Cott269	$(CT)_6N_{55}(CA)_8$	GTGCTGGACACCTCACTGCT AGACATGAGTCATTTGTGGGGT	280	60	2	0.57143	0.47371	AY861204
Cott272	$(GT)_7N_{30}(GT)_6$	ACCGAGGCTTTGTAGCCCTG TCTGAGTCCACGGCACACAC	229	60	6	0.31429	0.38219	AY861205
Cott29	$(TG)_{44}$	ATTTGATTCAGGGCCCAGTG GCCCTAGAAATAACACACACACGT	326	60	5*	n.e.	n.e.	AY861206
Cott290	$(AAACCC)_{10}$	AGCCGTAGCCGACCAGCAGC CTAACACCAGCACTAAGCAGC	279	60	25	0.68571	0.79586	AY861207
Cott293	$(GA)_7N_{16}(AG)_8$	GACAGGCTGCGCTCACCTGG ACGGCTATCCCTGGCCGGAC	169	60	3	0.48571	0.49234	AY861208
Cott296	$(CT)_7$	TGCAAACGTCTCACACGCTG GGAGTCTGAATTCCCTCGAGTG	258	60	2*	n.e.	n.e.	AY861209
Cott300	$(TG)_7$	TCAAGACAGACAGGTCAGCTC CCTTGACCTGATAATGCCCAC	155	60	3	0.02857	0.05673	AY861210
Cott313	$(GT)_6$	CGTACATTGCACCTGTGACAG TGTCCTGCTCTTGCTCAGTG	170	60	2	0.02941	0.05838	AY861211
Cott315	$(AG)_7(CAGA)_{13}$	ACAAGCCTGTGAGGTTGTGT TCTCTATCCATGCTGCCTGG	308	60	50	0.85714	0.95031	AY861212
Cott328	$(GA)_6$	AGATGGACTCCCTGCTGGAC GCAGTCTTTCACCGCTGGAG	162	60	6	0.64706	0.64399	AY861214
Cott348	$(TGC)_7$	TGGACCACTGAGTTGCTGTG TGGTGTGTCATACCTGCAGC	321	60	17	0.77143	0.85549	AY861215
Cott50	$(TG)_{24}$	TCCTGTGCTAATGCTGTGGA AAACCTCACCTCGGTGAGTC	177	60	5*	n.e.	n.e.	AY861216
Cott564	$(GT)_{49}(GCGTGC)$ $(GT)_6(GC)(GT)_{14}$	TGGGTCATACTGGGACGCTG ACAGTGCAGTCGTGTCACTG	350	60	7*	n.e.	n.e.	AY861217
Cott570	$(TG)_{27}(TAT)(GA)_6$	TGTAAGTCACAGTTGGCATTGC ACCACCCATCACGCACAGTG	179	60	24	0.67647**	0.85031	AY861218
Cott580	$(GT)_{37}$	ATTGTCTGGTGGATTTGCCG AGAGACGTCACAACTGCAGC	280	60	30	0.6	0.63851	AY861219
Cott582	$(GT)_{39}$	AGCTCTGGATGAGGACTGTC TCTGTGAGGTTCAGTGCGAC	233	60	27	0.77143	0.80497	AY861220
Cott588	$(TG)_{32}$	CCAGCAGCATCGAACACACC CTTGCGTCGTCAGACCTTCG	272	60	19	0.94118	0.89245	AY861221
Cott619	$(CA)_{15}(TA)(CA)_6$	TGTTTGAGGGAATCCCTGCA GAGATCCACCTGCTTCAAGC	141	60	21	0.62857	0.70973	AY861222
Cott635	$(TC)_8$	AGAAGGGCAACATAAGGGCC GAGGCGAGTGCGATGCCACC	169	60	9	0.68571	0.74369	AY861223
Cott67	$(GT)_{14}$	TGTGAACGAGTCATTGCTGC ACAACATGGACGGCGTGTTC	154	60	5*	n.e.	n.e.	AY861224
Cott675	$(AG)_{10}$	ACTGTGCTGCATGCCAACTG AGAGTTCGCCTGCAGTATGG	280	60	9	0.68571	0.65135	AY861225
Cott68	$(GT)_{22}$	GGACCTGTAATTTACAGTCCACACTG	154	60	7*	n.e.	n.e.	AY861226

		TGTGTGTTGGAGGTGCTCTC						
Cott684	$(TGTA)_{15}$	TGTCATGCTGGGATTCGCAC	300	60	41	0.88571	0.93292	AY861227
		AGGCGAACACAGTCTGGGCT						
Cott685	$(GT)_6$	TCCGGCTCACTCCGGAGGCA	165	60	4	0.14286	0.16273	AY861228
		GGGTCAAGTTTAGGCCTTTGG						
Cott686	$(GCA)_8$	TGGGAGTGCGATCGGGTTGG	244	60	8	0.68571	0.71304	AY861229
		GCGGGCCTTGACACTGTTGG						
Cott687	$(CT)_{17}$	TGTACCTAGTGAGCCTGCTG	144	60	18	0.74286	0.76936	AY861230
		CGTCAGCATCTACTGGGCAG						
Cott688	$(TC)_{12}$	GATCCCAACAATATCGGTCACC	175	60	4*	n.e.	n.e.	AY861231
		AGGCCGTTATCATCTGCAGC						
Cott697	$(AC)_6(CT)(AC)_8$	TCATCACGAGAAGGCTCAGTC	153	60	4	0.23333**	0.60226	AY861232
		GAGCACCAACGCCAATTTGG						
Cott700	$(AC)_{17}$	TGAGACGCCACCGCTCGGAG	250	60	10	0.60606	0.74639	AY861233
		GCTTACCAGTGGACTGCAGC						
Cott708	$(AC)_{12}$	TGGCCTAAATAAGCACCACG	144	60	10	0.73529	0.7568	AY861234
		AGCTAACGTCCTCCCACAAG						
Cott78	$(TG)_{20}$	TCTGCTCCGCAGGGTCGCTG	325	60	26	0.88571	0.88364	AY861235
		TGACACATGGTGAACCCGTG						
Cott91	$(TG)_{15}$	ACGACAGTGCCGGGTTTCAG	161	60	16	0.82857	0.80455	AY861236
		TGGGAGTTCTGGTCTGTTGG						
Cott98	$(TG)_{15}$	CATTGGCCTTGCTGGAAAGC	142	60	5*	n.e.	n.e.	AY861237
		CCCAAGACAACTTTCTAACGCAG						
CottE1	$(GT)_{49}$	ACTTGGCTCCTCTTTGCAGCAG	230	59	6	0.42857**	0.66418	AY861238
		ACAGGTCGGAAGACTGGCTGTC						
CottE10	$(AC)_{22}$	CCGGGTGGCGATAGTACCCTGC	277	59	25	0.82857	0.83602	AY861239
		AGGCACTGGTAATGAACTGCTC						
CottE11	$(TG)_{41}$	TCAAGAGGTGAGAGGCCAAGAG	245	59	31	0.65714**	0.93043	AY861240
		CATCCTGGTGTTACCATTGCTG						
CottE12	$(GT)_{46}$	GAGTCAGCCACGAGTGCAACAC	277	59	28	0.61765	0.66286	AY861241
		TCCAGTCTGATGGGCACAAGAG						
CottE13	$(GT)_{32}$	AGTTACGCAATGCAACACGCCT	200	59	15	0.875	0.86111	AY861242
		TCGAACACGAAGCTCCCGTCAC						
CottE2	$(TG)_{15}$	CGTGATGGGAACCATGTGTC	222	59	23	0.74286	0.85052	AY861243
		CTCTGCCATGAAGCGGCTCG						
CottE20	$(GT)_{20}$	CGTCTGCACAGTTGAGCCTG	170	60	13	0.85294	0.87138	AY861244
		AGCACGCCGTCGCTAACAGC						
CottE21	$(TG)_{17}$	ACTGCACCGGTCCTTGCAGG	233	59	15	0.35**	0.83718	AY861245
		GAAGTGTGTCCACTCACAACC						
CottE23	$(TG)_{27}$	TGAGCAGCTTTACAGTGTTGG	317	59	29	0.91429	0.90352	AY861246
		TCGTGCTCTTCTGGCATCTG						
CottE30	$(GT)_{13}(GC)(GT)_8$	AAGCGGACAGACGCAAAGAG	173	59	22	0.77143	0.88778	AY861247
		ACTCGCCCAAGAAGACCAGC						
CottE31	$(TG)_{19}$	TGACGTAACCACCCGACCAC	267	59	13	0.68571	0.78095	AY861248
		CAGTCAGGACAGCATCATGG						
CottE32	$(TG)_{15}(AT)(TG)_{11}$	GGTCCAGCAGCAACGCACAC	189	59	32	0.79412	0.86831	AY861249
		GCAAAGGCTATCTGTCACAGC						
CottE399	$(CA)_{23}$	GTTGGTCCACAGTTGCTGAC	136	59	19	0.59375	0.72222	AY861250
		AGTAAGTCGGTGACAGGCTC						

CottE6	$(GT)_{29}$	ATGTGAAGGCAGTGCACCTC ATAGTGCCTGGCAGCAACAC	158	59	18	0.77143	0.78385	AY861251
CottE7	$(GT)_{22}$	AGACTCGAGAGGAATGAGTC GCTCCGCCTCCGTCTCTGAG	177	59	18	0.71429	0.86667	AY861252
CottE8	$(TG)_{26}$	TGACTGCAGCATGTGCTACC GCAGTTTCCGTACCGCCCAG	268	61	20	0.78788	0.85781	AY861253
CottE9	$(TG)_{17}$	TCCTGATGGGAGCCACTTGC TTCCTGATGGGCGGGATCTC	297	59	14	0.71429	0.8058	AY861254
CottES1	$(AG)_{15}$	CGAGAGACAACACAGCCCGGTAG AAGAAGTTGCACGCTCCATCG	143	63	7	0.55882	0.73486	AY861255
CottES10	$(GT)_{12}$	CAGGCGGCGACACGGTG TTATGAGGAGTCTGCCAATGCAG	186	63	42	0.88571	0.9089	AY861256
CottES19	$(CT)_{11}$	CCCCTTCCGTGAGCGACG ATATCACAGCACTTTTGGGGATG	144	60	10	0.32353	0.38762	AY861257
CottES2	$(GT)_{13}$	TTATGGATACATTGGCAGCACG TGTACGACCTGGACGGAAACG	191	64	10	0.71429	0.7354	AY861258
CottES21	$(GT)_{12}$	ATGGAAAGGAATTGTGGCGAC GTGGAACAGTATGGCTGCAGCA	153	64	14	0.68571	0.66874	AY861259
CottES3	$(GT)_5(TGCT)(GT)_5$	CTGGTGTGTGTGTTTCAGGGAGG CAGGTGTGAGAGACCCGACTCC	236	63	9	0.65714	0.67122	AY861260
CottES6	$(AG)_{40}$	GCCAGGAATACTTTTGGGAGG CGCTGTGTAGTTAGGACCCAGG	233	63	22	0.91176	0.9065	AY861261
CottES8	$(TG)_{10}N_{37}(TG)_9$	TGTTCAGTGTTTGCATTTTGCG TCCAAAACCTGCGAAACTGC	237	63	7*	n.e.	n.e.	AY861262
LCB12	$(TG)_{21}$	AGGATGTTGCAGGAGGGTGG TAAGAAGAGATATCACAGGCTTTG	188	60	10	0.70588	0.63257	AY861263
LCB16	$(AC)_{23}$	GAGATGGGAATAAAAGAGGAGC CGTCCTCTCATTGACGGGATG	165	59	20	0.85714	0.86625	AY861264
LCB18	$(TG)_{15}N_7(GA)_6$ $(AA)(GA)_6$	ATCACCGAGTATGCCCACTGG CACTTTCGCAGCAATCCCTG	280	58	11	0.5**	0.85909	AY861265
LCB67	$(AC)_6$	GGAGTGCTGACTCCTCTGCTGG TCACCGCTCACTTTGGGGTC	208	60	4	0.02857	0.05673	AY861266
LCE100	$(CA)_{10}$	CGCTCCTCATTACAGCCCACG GGTTTTATGAAGCAGCAAGGTGC	178	60	5*	n.e.	n.e.	AY861267
LCE103	$(GT)_{10}(CTA)(TG)_6$	GCACCTATCTACCTACCGGCTATC GTTCCATTCACCCGAAACTGG	193	60	9	0.65714	0.6472	AY861268
LCE105	$(TG)_9$	GCAATATGCAATCGCTGGCTG CGGATTTACGTGGCCGTCAAC	146	59	12	0.88571	0.81449	AY861269
LCE11	$(TG)_{30}$	GGAGAAGGAGAAGCAGCTTGC GGGTGGAAGGATTCGGAGG	215	57	30	0.61765**	0.77788	AY861270
LCE111	$(CT)_7N_6(CA)_8$	GCTGCAGACAGAGCTGCCAATC GGGTGATTCTGTTTAGGCCAGC	228	60	8	0.8	0.74783	AY861271
LCE12	$(TG)_{46}$	AAGAGGAGGTCTCGTTGCCAG AGCGGGGACGCCTGTCC	217	61	3*	n.e.	n.e.	AY861272
LCE122	$(CA)_8$	CACAGCTAAGTGCACCGGCAC CATGTGTTATTTATGAGCGCTGC	284	59	4	0.48571	0.5205	AY861273
LCE126	$(AC)_8$	CGGTGTTAGACGTTTGGCTTCC CCACACAAGTCCGAGGCTTGG	271	60	8	0.25714	0.28903	AY861274
LCE13	$(GT)_{16}N_{12}(GT)_{12}N_{12}$	TATGGGAAACTACCCTCAGGGT	196	61	20	0.8	0.82733	AY861275

	$(GT)_7N_{12}(GT)_8$	CTCAAACATGTGCAGGCACTGC						
LCE162	$(CA)_6N_{35}(AC)_7$	GAAGTAAACTTCATCAGGTCGACG	269	65	7	0.68571	0.66957	AY861276
		CTTCACATGGATGCTCTCCAGC						
LCE179	$(AC)_7$	AATCCACTTTCACTGGCGGTT	160	63	4	0.05882**	0.24715	AY861277
		CCATGCCTGTAACCTTTGGCAC						
LCE181	$(AC)_6$	GCTCTTGCTTAAGACTCGCCAT	129	65	2	0.51429	0.47371	AY861278
		ATCACAGTCCCTCTCCGCTCTG						
LCE2	$(TG)_{28}$	AACCTAAAGGCCGCTGAATG	155	63	29	0.91176	0.91879	AY861279
		CTTCCTCACGGCACTGCAGG						
LCE20	$(CTCCT)_{18}$	GGATCAGATCAGTTTGAGGAGCG	266	60	6	0.41176	0.54785	AY861280
		TCTCTGTTCACTGTCCCGTCCTC						
LCE21	$(GT)_{24}$	TGAGAGGAGGAGAGGCTTGCAC	177	60	23	0.8	0.85135	AY861281
		TGACGGCACCGGAGCTGAG						
LCE219	$(CCT)_{13}$	TATTTGAGGTGTGTGCCAAGTGG	212	63	14	0.77143	0.72133	AY861282
		CGGAGCAGAAAGAAAGCCAGC						
LCE22	$(GT)_{24}$	CCAGAAGTGCTCGCCTCTCC	368	60	16	0.91429	0.81573	AY861283
		GTGTAGATGAGTGCCCGCTCC						
LCE24	$(TG)_{18}$	AAATAACCTCAGACGTGCACTGC	177	65	24	0.28125**	0.81349	AY861284
		CCAGACGTGAACAACACGGCTG						
LCE25	$(TG)_{26}$	TCATTGTGTGTTCACCCTGCCAC	287	63	13	0.37143	0.37474	AY861285
		CGTGTTCTTGTTTTGCAGCTCG						
LCE26	$(TG)_{20}$	TTTATTAGTTTAGCTCCCCTTGC	303	63	41	0.91429**	0.94948	AY861286
		AAGCATAATAGTCTCCGCCTCTC						
LCE27	$(GT)_{29}$	TCATAGCAGCCCGATTGTGC	243	61	32	0.8	0.81159	AY861287
		TGATTAAGCCAAGCCGTCCTC						
LCE275	$(AGA)_{10}$	ACGCCGCGTCGGCCTAG	306	63	8	0.62857	0.66335	AY861288
		TGGGGCTTTTGTGCCTGC						
LCE279	$(ATG)_8$	ATCAAAACCTCAGGAGGCCAC	209	63	6	0.58065	0.73295	AY861289
		CAGTTTTGCAAGAAAACCCCAC						
LCE29	$(TG)_9(T)(TG)_{22}$	GTGGGGAGAATGACGGATGG	221	63	34	0.82857	0.89482	AY861290
	$(A)(GT)_6$	TACACATGCATTTGGATTGACC						
LCE30	$(GT)_9(AT)(TG)_{10}$	AGTGATGAATCTCTCTGGGACCAG	164	57	19	0.71429	0.88323	AY861291
		GGTACAGCCCACAGCCAAGG						
LCE31	$(TG)_{17}$	GGATTATGACTGTTCAAGGTCCG	281	63	28	0.875	0.8874	AY861292
		CGGCGCAGAGCAACCGTG						
LCE32	$(TG)_{23}$	GTCCACTTTCCAGGGGACTCC	259	65	13	0.68571	0.85797	AY861293
		GGAGGTACGAAGGGAACGAAG						
LCE35	$(TG)_{19}$	ATTTTACACGAGCGGCAGAAG	173	63	11	0.62857	0.68033	AY861294
		CCTGCAGTGTTTAAACCGGAGC						
LCE37	$(TG)_{14}(T)(TG)_{13}$	GCATGGTGCCGGGAGTGG	210	66	19	0.58824	0.60053	AY861295
		CTGATTCAGCTCGCTGGGATG						
LCE38	$(GT)_9(A)(GT)_8$	TTGGGGCCCAACGGAGC	149	63	21	0.65714	0.84017	AY861296
		GGGCCACACACACGGAGACG						
LCE39	$(TG)_{19}$	GACTCTCCTCGTCTTCCCCATG	203	63	15	0.97059	0.89113	AY861297
		TCCACCACATATCCACCCTGG						
LCE4	$(TG)_{25}$	AATAATAACAAATGTCATGGCTTC	159	57	36	0.88571	0.94037	AY861298
		GGAAAACACGTCTGGGGTAGG						
LCE40	$(CA)_{11}$	GTGGCTATGCATGGTTCCTGC	304	61	10	0.8	0.76273	AY861299
		ATTGCCTTCTTTGGAGCGTG						

LCE42	$(GT)_{18}$	CGTTCAGACACGACGAGCAAACC CAATGGCTTCAGAGGCCGAG	132	63	11	0.65625	0.77679	AY861300
LCE43	$(TG)_{19}$	AACGAATCATGTCCCGGTTTG GCTGCGTCTCTCCGGACACC	172	61	20	0.48571	0.48323	AY861301
LCE44	$(GT)_{12}$	CAGCCGGCCGTGACGTG CCCGCCACTCGCCCTTG	252	63	7	0.8	0.70559	AY861302
LCE46	$(TG)_{15}$	TTTAGAGGTGAATCAAGAGGCAG TTTCTCTCGAGTCCTCTCGTCC	191	63	20	0.88571	0.83685	AY861303
LCE48	$(GT)_{15}$	CCTCTTCATCACCCGCTCTTTC AGTCTTTTCACCTGTTGCCAGG	192	61	22	0.91429	0.90228	AY861304
LCE5	$(GT)_{67}$	AACTGTCACTCAAGCCCATTCG GCCGCGTGAGCCAACAGC	217	61	9*	n.e.	n.e.	AY861305
LCE51	$(GT)_{17}$	GCAGCGGCCTCTCGCATGAG CCGGATTTCTTAATGACCAAACCAG	249	58	14	0.62857	0.65424	AY861306
LCE52	$(GT)_{15}$	AGTGTGCCCTTCATGCTGTG GCTAAGGGATGTGTCTGTCG	262	63	7	0.77143	0.80414	AY861307
LCE54	$(TG)_{14}$	ATGTGTGAGGGCCTCAATGC CTGTGTGGTGCGTCCATTGG	233	63	7	0.54286	0.54783	AY861308
LCE59	$(TG)_{15}$	GCCAGTCGACAAACAGCTGC ACAACCGTGTCACCAAGTGC	170	60	18	0.82857	0.83106	AY861309
LCE64	$(TG)_{10}$	TGTCAGCCTGTCCTGACGAG TCTACAAGTGCAACAGCGAC	127	60	18	0.45714	0.59752	AY861310
LCE66	$(GT)_{12}$	AGCATATCGGCAGGAGGCTC ACAAAGTGCTGCTCACTCAC	140	60	21	0.37143	0.44306	AY861311
LCE68	$(TG)_{11}$	ACGTTTGCGTGTGTGCATCG TCATTGGCTTGTGCATGTCC	213	59	7	0.54286	0.58634	AY861312
LCE69	$(AC)_{13}$	GGAATCACGATAGCGAAGACG AACAACAAGCACAGCAGTCG	204	59	7	0.41176	0.59658	AY861313
LCE74	$(TG)_{11}$	GGTAGGCTGTTTCCAAGCAG AACAGCCCACATTCTGCAGC	186	60	9	0.68571	0.71967	AY861314
LCE75	$(CA)_6(GA)(CA)_9$	AGGAACGCATAACGTCTTGC CAAATACCGCTCGACCAGAG	225	58	9	0.68571	0.7234	AY861315
LCE78	$(TG)_{11}$	TCAGTGACTCCTGTGGTGAC ACGGCGTATGTCAAACCGTG	187	60	5	0.67647	0.68569	AY861317
LCE79	$(TG)_{11}$	AAGCCAGGTTTGACATGTGC ACGATGCACCCTGAGCAGAC	334	58	6	0.51429	0.61615	AY861318
LCE80	$(TG)_{10}$	ACTTGCAGGGAACTTGGCAC TCTACCCCTGGTTTGTGTGC	223	60	3*	n.e.	n.e.	AY861319
LCE81	$(TG)_{11}N_{15}(GT)_6$	GCTTGTGAACATTTCGGCCTC AATGTACAAAGCTCCCACGTCC	154	59	17	0.53125	0.625	AY861320
LCE82	$(TG)_{11}$	TGTCAATCAAGAGACGAAGGTCG GCAGGTGGCAGCAGCACTGC	157	60	8*	n.e.	n.e.	AY861321
LCE83	$(GT)_{11}$	AGGCGGAGTGCAGGGCTC CCTTCCACTCATGTCCCCTTTG	175	60	7*	n.e.	n.e.	AY861322
LCE87	$(CA)_{11}$	TGCCAGCTCTACCGTCAGCTG CGGCCCACGCCATCCTC	158	59	8	0.46875	0.48165	AY861323
LCE88	$(AG)_{18}$	TGGGCGTCGGCCTCCTC CGATCTTTATCCTCGGCCAC	235	59	12	0.57143	0.53996	AY861324
LCE89	$(TATC)_5$	AGAGCACACACCCTTCCGGTC	275	60	34	0.82353	0.85514	AY861325

		GAACCTGCACAGGGCTACAGC						
Cott108	$(AC)_7N_{22}(AC)_{16}$	TGTCTGCAATGGCCTTCAGG	301	60	27	0.65714	0.80952	AY861326
		TGCATAACCCACCAACTGTC						
Cott197	$(GA)_{16}$	TGAGCGCCAGCTGCACAAGC	294	60	35	0.28571**	0.50518	AY861327
		ACCTCTTTGGATTGCCAACAG						
Cott213	$(CA)_9$	TTGCCATGGATTTGAGGCAG	338	60	17	0.82857	0.85921	AY861328
		AGCATTGCTATTATCAGGCTGC						
Cott222	$(AG)_{15}$	ACGGATGAACAGCATTGAGG	288	60	13	0.5**	0.74188	AY861329
		TGAACCCAAACTCTTGCCAG						
Cott273	$(ACAT)_{41}$	GTGGCAGCTTTGAGGCAACC	342	60	43	0.57143**	0.95983	AY861330
		AGGATCGGTGAGCTAACTGC						
Cott43	$(TG)_{39}$	CAGAGCTGTTGACTGGTGAC	304	60	24	0.36667**	0.82203	AY861331
		TGGTGTGAACACTCATGCAG						
Cott54	$(AC)_{34}$	CACGTACATGTGAAACGAACCC	156	60	18	0.42857**	0.8911	AY861332
		AACAGAGGAAACGCCATGAC						
Cott584	$(GT)_{21}$	TGCATGGGTGCTCATGCCTG	174	60	33	0.64706	0.64794	AY861333
		TCATCCGGTGTCAAGCTGTG						
Cott722	$(GT)_{54}$	AGGGCCACAGGCTTGCCCAG	328	60	66	0.58065	0.98255	AY861334
		GCTCAGCGCAGTCCCTGCAG						
CottE16	$(GT)_{34}$	CAGCACTTTGTCAGTCAGACC	247	62	38	0.87879	0.92168	AY861335
		TTGGGCACGCGCCTGGTCAG						
CottE19	$(CA)_{27}$	GACCTTGAAGTCTAACAGATGCAC	177	60	24	0.55882**	0.83582	AY861336
		TCCACCTCTACCCACTGTCC						
LCB13	$(GT)_{15}$	ATGATGTCAGTGTGTGCTCTTGCA	174	61	32	0.91176	0.91747	AY861337
		CGGGAGTTTCTAGCTGCCATG						
LCB4	$(CAGTA)_{30}$	CAGACAGACAGGCTGCCCACTG	317	57	3	0.20588	0.2331	AY861338
		GTCGACCTGTTACCATTACTGG						
LCE55	$(GT)_{15}$	CCAGCTGACGTCAGCTGGAG	164	63	30	0.45714	0.5619	AY861339
		GTGGTTCCTTTCCAGAACGAC						
LCE57	$(GT)_7$	AAGACCGGCCTCCAATGAGC	226	63	17	0.6	0.74948	AY861340
		GACGGCGGTGTTGACCCTCG						
LCE58	$(GT)_{13}$	GGCTCCAACGCGACGCACTC	264	63	12	0.47059	0.65847	AY861341
		TGGCAACCCCGGTTGGCTCG						
LCE62	$(TG)_9$	CGATTATGCTCCAGGTCACAC	293	58	36	0.77143	0.92712	AY861342
		GCGATGAATGGCAGCATGTC						
LCE73	$(GT)_{12}$	GCTCATCTGCTCAGTGGGAC	212	58	13	0.51429**	0.84099	AY861343
		TCCACAGCCATCCCCCTAAC						

Acknowledgements:

This work was funded by the DFG and the Verband der Chemischen Industrie. We thank A. Pozhitkov for helpful discussions on hybridization protocols and T. Bayer for his help in streamlining our data analysis.

Literature

Englbrecht, C. C. Largiader, C. R. Hänfling, B. Tautz, D. (1999) Isolation and characterization of polymorphic microsatellite loci in the European bullhead *Cottus gobio* L. (Osteichthyes) and their apllicability to related taxa. Molecular Ecology, 8, 1966 - 1969.

Englbrecht, C. C. Freyhof, J. Nolte, A. Rassmann, K. Schliewen, U. Tautz, D. (2000) Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-pleistocene origin of the major central European populations Molecular Ecology, 9, 709-722.

Schneider, S. Roessli, D. Excoffier, L. (2000) ARLEQUIN: a Software for Population Genetics Data Analysis, Version 2.000. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Steinfartz, S. Küsters, D. Tautz, D. (2004) Isolation and characterization of polymorphic tetranucleotide microsatellite loci in the Fire salamander Salamandra salamandra (Amphibia: Caudata) Molecular Ecology Notes, 4, 626–628.

Tautz, D. Renz, M. (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl. Acids Res. 12, 4127-4138.

Volckaert, F. A. M. Hänfling, B. Hellemans, B. Carvalho, G. R. (2002) Timing of the population dynamics of bullhead *Cottus gobio* (Teleostei: Cottidae) during the Pleistocene. J. Evol. Biol. 15, 930-944.

Zane, L. Pargelloni, L. Patarnello, T. (2002) Strategies for microsatellite isolation: a review. Molecular Ecology 11, 1-16.

Chapter 3: A Genetic Map of *Cottus gobio* (Pisces, Teleostei) based on microsatellites can be linked to the Physical Map of *Tetraodon nigroviridis*

Kathryn C. Stemshorn, Arne W. Nolte und Diethard Tautz

Abstract:

To initiate QTL studies in the non-model fish *Cottus gobio* we constructed a genetic map based on 171 microsatellite markers. The mapping panel consisted of F_1 intercrosses between two divergent *Cottus* lineages from the River Rhine System. BLAST searches with the flanking sequences of the microsatellite markers yielded a significant (e < 10-5) hit with the *Tetraodon nigroviridis* genomic sequence for 45% of the *Cottus* loci. Remarkably, most of these hits were due to short highly conserved non-coding stretches. These have an average length of 40 bp and are on average 92% conserved. Comparison of the map locations between the two genomes revealed extensive conserved synteny, suggesting that the *Tetraodon* genomic sequence will serve as an excellent genomic reference for at least the Acanthopterygii, which include evolutionary interesting fish groups such as guppies (*Poecilia*), cichlids (*Tilapia*) or platy (*Xiphophorus*). The apparent high density of short conserved non-coding stretches in these fish genomes will highly facilitate the identification of genes that have been identified in QTL mapping strategies of evolutionary relevant traits.

Introduction

Syntenic relationships offer the possibility to transfer genomic information available for model organisms to non-model organisms, which are genetically less well characterized (Schmid 2000, Gebhardt *et al.* 2003, Erickson *et al.* 2004). With a number of complete genome sequences becoming publicly available, the possibilities for comparative approaches are increasing. Studies range from basic comparisons of chromosome structure (Chowdhary *et al.* 1998) to the identification of synteny-defined candidate genes (Giampietro *et al.* 1999). Whole genome comparisons of different species reveal information about homologies, conserved regions, syntenic relationships, genome duplications or duplications of genomic fragments, and genome evolution in general. Comparisons like this are only possible for fully sequenced model organisms. However comparisons of the genetic map of one organism with the physical map of another organism can also be very informative. Among plants this possibility has been employed to gain

information about conserved synteny between the plant model Arabidopsis thaliana and different crop species (Dominguez *et al.* 2003, Gebhardt *et al.* 2003). One of the hopes is, that through comparative analysis knowledge about the genetic make-up of non-model organisms can be gained without having to construct a physical map. Depending on the goal of the study these approaches require high degrees of genome colinearity at the genetic level and at the gene level (= microsynteny) (Schmid 2000), as well as sufficient similarity between the sequences to identify homologous regions. Consequently, the question arises of how closely related organisms should be for comparative analysis to be fruitful. Here we show, that there is an unexpectedly large number of highly conserved non-coding regions in fish genomes, which allow to place at least half of randomly cloned non-coding fragments on the map of fully sequenced fish genomes and to use synteny relationships to infer the gene content of the respective regions.

Our study object, *Cottus gobio* (Scorpaeniformes, Cottidae, *Cottus* - common name "bull head"), is a small, benthic freshwater fish, which occurs in streams throughout Europe, with closely related species distributed throughout the northern hemisphere. Our focus area is the River Rhine System, where phylogeographic lineages of *Cottus gobio* occur in parapatry (Englbrecht *et al.* 2000). We have recently found that ancient lineages (stream type) occupy the headwaters of small streams within the Rhine area. Intriguingly, a new invasive lineage has recently appeared in areas of the main rivers that where previously free of *Cottus* (Chapter 1). This indicates a divergent ecological potential of this lineage (invasive type). The invasive type forms contact zones with the stream type, which are coupled with habitat transitions, in particular in regions where small tributaries disembogue into the mainstream. These ecological transition zones are areas of intense admixture but remain extremely narrow suggesting a strong role of differential adaptation (Chapter 4). This situation is predestined to study the genetic basis of adaptive traits. Possible approaches in such hybrid zones would include studies of geneflow (Barton and Gale, 1993), direct inference of selection at marker loci (Kauer *et al.* 2003), trait mapping exploiting linkage disequilibrium (McKeigue 2005) or microarray-based identification of differential gene expression. All of these approaches would ultimately yield candidate loci that point more or less directly to loci that play a key role in understanding genetic processes of evolutionary divergence (Erickson *et al.* 2004). However, as in many organisms of evolutionary interest, the genome of *Cottus gobio* is not yet genetically characterized. We have opted for microsatellites as a basis for a first genetic map, since they have the advantage over AFLPs or other anonymous markers in providing a direct genomic tag via their flanking regions. Here we show, that this is indeed a key towards integrating known genome data from other fish species into identifying genes in the species of interest.

Methods

Construction of mapping families

We established crosses of two divergent lineages (stream type and invasive type) of *Cottus gobio* found in the River Rhine system. All animals

used were taken from the River Sieg drainage, from locations outside of known hybrid zones.

To obtain crosses, mature pre-spawning adults were collected in the field in February 2002 and transferred to laboratory tanks. Fish were fed ad libitum with insect larvae. Spawning occurred readily in artificial shelters partially buried in sand at temperatures between 8-10 °C. After spawning, only the guarding male was left with the egg clusters. After hatching of the larvae the male was removed. Larvae were raised initially using live Artemia nauplii, and later with frozen chironomid larvae and mysiid shrimps until at least 3 cm in length. All animals were preserved in 70% ethanol for future studies.

One cross involves a male from the population “Giertshagener Bach” (stream type; Stream Giertshagener Bach at Giertshagen, North Rhine-Westphalia, Germany; 50°45'N 7°36'E) and 2 females from the population “Wahnbach” (invasive type; Stream Wahnbach, Outlet into River Sieg at Seligenthal, North Rhine-Westphalia, Germany 50°48'N 7°16'E) resulting in two half-sib families (n = 24 and 63 progeny). A full-sib family was obtained from a female from “Ottersbach” (stream type; Stream Ottersbach at Eitorf, North Rhine-Westphalia, Germany; 50°47'N 7°26'E) and a male from "Wahnbach" (see above) and contains 78 progeny. Attempts to create an F2 generation intercross failed for unknown reasons. Note, however, that this is not due to general hybrid sterility as numerous F2 or backcross hybrids were found in natural hybrid zones (Chapter 4).

Genotyping of the microsatellite markers

Loci were taken from Englbrecht *et al.* (1999) and Nolte *et al.* (2005). All individuals were genotyped for 171 microsatellite markers on a Megabace 1000 (Amersham Biosciences). PCR reactions were performed as multiplex; up to 6 fluorescently labelled (Fam, Hex, Tet) primer pairs were combined and amplified using the Multiplex-PCR Kit (Quiagen) as described in Nolte *et al.* (2005). The loci were combined in a way such that all fragments could be separated in a single lane without overlap and scored unambiguously.

Linkage analysis

Linkage distances and marker orderings were determined with the Locusmap software (Garbe and Da, 2003). The sex-averaged LOD-threshold was set to 3. Since many linkages only rely on either male- or female-specific recombination rates, due to the parents not being equally informative for every locus, the resulting sex-specific maps are not reliable. We could thus not evaluate sex-specific differences for the linkages. The Haldane map function was used to convert recombination frequencies to centimorgans (cM). Non-inheritance errors were checked again in the genotyping files and then classified as probable allele-drop-out errors, when the progeny was homozygous for a parental allele only found in one parent, or allele-mutation errors, when the progeny possessed an allele not present in one of the parents, which could be explained by a single step mutation of a parental allele. Graphics of the linkage groups were produced with the MapChart software (version 2.1; Voorrips 2002).

The sex-averaged LOD-Scores ranged from 3.26-94.81 with an average of 21.21. The informative meiosis among the linked loci ranged from

62-330 with an average of 199.9. Identical inheritance was detected for 57 marker-pairs. 20 non-inheritance errors were detected, of which 16 concern a single locus and can be explained by allele drop out in the progeny. The remaining non-inheritance errors are spread over five different loci and can also mainly be explained by allele drop out except for one locus, were a mutation in one of the progeny alleles is the most probable explanation.

Test for Mendelian segregation

Tests for Mendelian segregation were performed using Pearson's chi-square test with an expected segregation ratio of 1:1 for all alleles (significance level $P < 0.05$). Every family was tested separately for every marker, which resulted in 513 pairwise comparisons of observed vs. expected allele numbers. Markers not following Mendelian segregation were checked for genotyping errors (see above).

BLAST searches

BLAST searches (Altschul *et al.* 1990) were done against the *Tetraodon*, Fugu and Danio genomic sequences via the Ensembl Genome Browser (http://www.ensembl.org/). Similarity searches against the Medaka sequences were conducted via the Medaka Genome Project homepage (http://dolphin.lab.nig.ac.jp/medaka/index.php). The *Cottus* sequences had an average length of about 500 bp (range from 119–1109 bp). Hits with $e < 10\text{-}5$ were considered as significant. The corresponding *Tetraodon* sequences were retrieved for sequence comparisons. Local alignments were produced with DIALIGN 2 (Morgenstern 1999) using the default settings.

Results:

Cottus map

Three mapping families consisting together of 170 individuals were genotyped for 171 microsatellite loci. 3.3% of the tests for Mendelian segregation were significant at $P < 0.05$, indicating that the level of segregation distortion was within the limits that are expected by chance. 366 significant pairwise linkages (LOD > 3.0) were detected for 154 of these markers. The loci could be assembled into 20 linkage groups (Fig. 1). The lengths of the linkage groups range from 0-1681.7 cM with 2-49 markers per group. The longest linkage group is linkage group 3 with 1618.7 cM, the cumulative map length is 2738.1 cM (see Methods for further details on the

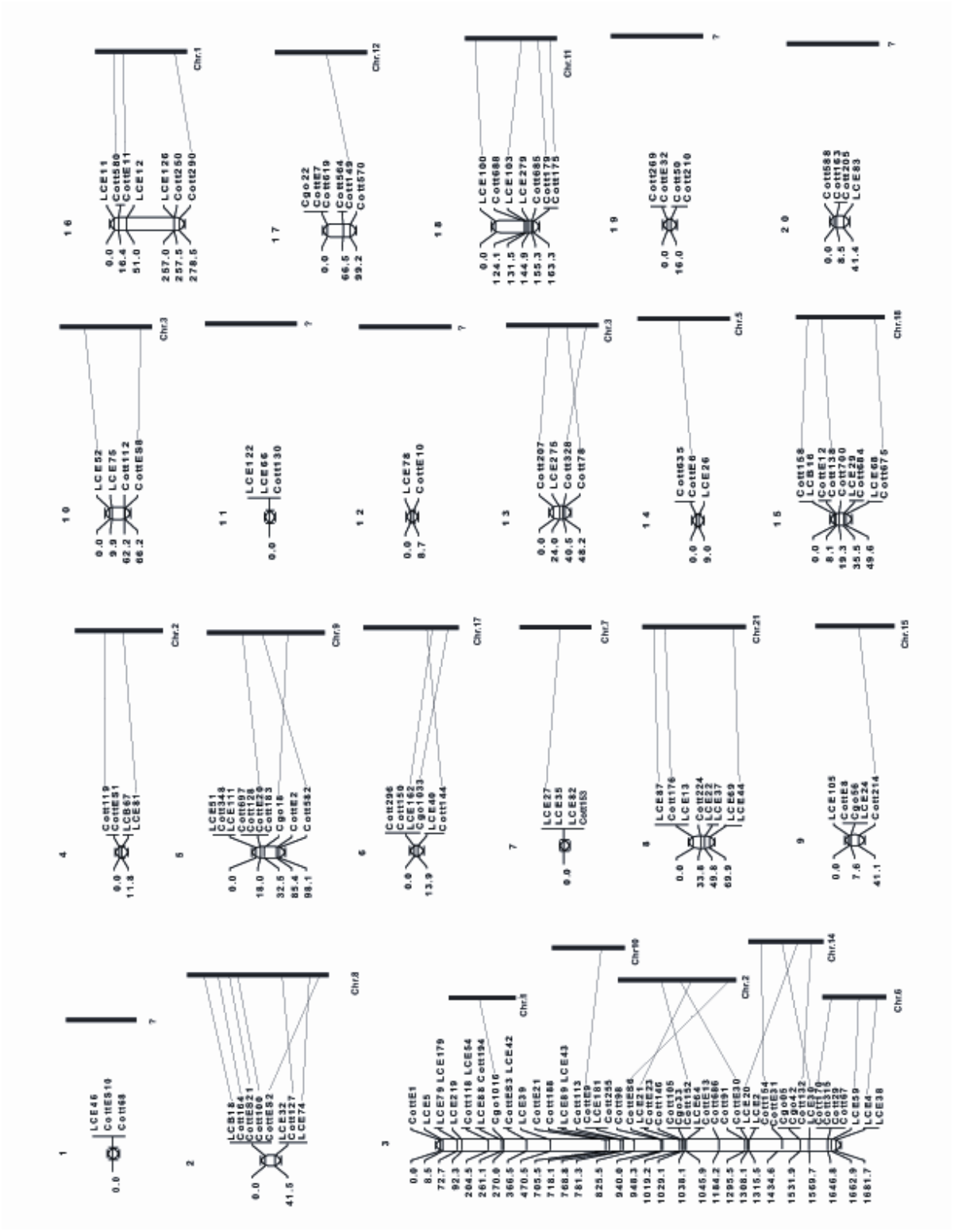


Figure 1: Comparison of *Cottus gobio* linkage groups with the chromosomes of *Tetraodon nigroviridis*. Significant BLAST hits and their relative positions on the *Tetraodon* chromosomes are indicated by connecting lines between the *Cottus* locus and the *Tetraodon* chromosome.

map). Given that the chromosome number in *Cottus* is 24 with no conspicuously large single chromosome (Vitturi & Rasotto 1990), it seems likely that linkage group 3 is artificial and will become fragmented when more mapping groups are included.

The published genome size of close relatives of *Cottus gobio* is slightly below 1 pg per cell (Hardie & Hebert 2003) and this value was also found for C. gobio from both lineages involved in this study in a first estimate (T. R. G pers. com., compare http://www.genomesize.com/). According to Dolezel *et al.* (2003) this can be converted into a genome size of about 1000 Mbp. One centimorgan would thus correspond to 0.36 Mbp.

Locus matching

The flanking sequences of all typed microsatellite loci were used for similarity searches against the Danio, Medaka, Fugu and *Tetraodon* genomes. Using a significance threshold of e < 10 –5 we found between 21 to 87 hits in the different genomes, most of which are even retained at a significance threshold of e < 10-10 (Tab. 1).

The matches were usually due to blocks of very highly conserved sequences. For *Tetraodon* comparisons, these had a length of 19-120 bp (average 40 bp) with sequence similarities between 62-100% (average 92%).

Only about a third of the loci with matching flanking sequences showed a conservation of the microsatellite itself (i.e. at least 5 repeats of the respective sequence motif) in *Tetraodon*, confirming the expected high turnover of such sequences (Schlötterer 2000).

The total length of *Cottus* sequences analysed in these BLAST searches was 86,530 bp. Given that 77 fragments yielded a significant hit with the *Tetraodon* genome sequence, we can estimate that at least one conserved block occurs about every 1100 bp. Thus, it should be possible to analyse even microsyntenic relationships throughout the genomes of these species.

Table 1: Number of BLAST matches of *Cottus* microsatellite flanking sequences in other fish genomes

matches with	$e < 10^{-5}$ N out of 171	$e < 10^{-10}$ N out of 171
Danio	21	11
Medaka	18	11
Tetraodon	77	64
Fugu	87	67

Link with the *Tetraodon* map

An ordered map is available for the *Tetraodon* genome, which covers about 64% of the genome sequence (Jaillon *et al.* 2004). Comparisons of map positions of the *Cottus* markers with a hit in the *Tetraodon* sequence thus allow assessing large-scale synteny patterns. We find that most markers from a single linkage group in *Cottus* yielded also hits on a single chromosome in *Tetraodon* (Fig. 1). The major exception is *Cottus* linkage group 3, which yields hits with five *Tetraodon* chromosomes, but in separate blocks, confirming our inference that *Cottus* linkage group 3 will split up when more data is available. The observed syntenic relationships together with the sequence similarities between the *Cottus* and *Tetraodon* sequences suggest true homology of the associated regions.

Five *Cottus* linkage groups could not be associated with a *Tetraodon* chromosome so far. In some cases this was due to lack of significant hits with the respective markers (groups 12 and 19) and in other cases hits were only found on genomic fragments that are not yet anchored to a *Tetraodon* chromosome (groups 1, 11 and 20).

Given that *Tetraodon* has only 21 chromosomes (Grützner *et al.* 1999), we cannot expect a one to one syntenic relationship between all linkage groups. This is also reflected in our finding that *Cottus* linkage groups 10 and 13 map to a single *Tetraodon* chromosome (Fig. 1). However, the general patterns are clearly comparable and suggest that large parts of the genomes will be alignable.

Discussion:

The tackling of specific evolutionary questions often requires working with non-model organisms. However, when it comes to understanding the genetic basis of an evolutionarily interesting trait, the limited genetic options in non-model organisms may prohibit even standard approaches that are commonly used in model organisms. Erickson *et al.* (2004) noted, that "...synteny among species or genera may lead to the opportunity to complement initial QTL experiments with candidate gene approaches as QTL within an interval may be matched to genes of known function in homologous chromosomal locations identified in related model organisms...". The key question of how closely two species have to be related for such approaches to become feasible remains to be answered empirically.

To identify the genetic basis of a given trait, one has to construct linkage maps that allow the correlation of a molecular marker with the trait. In non-model organisms, it will often only be possible to obtain an F_1 cross for mapping, which limits the map resolution. It is therefore of special interest to assess in how far completed genome projects can help to identify genes in non-model organisms. Studies in plants have already been conducted to evaluate whether microsyntenic relationships exist between model and non-model plant species. Colinearity can generally be observed at the level of genes within flowering plant families and could aid fine-mapping and map-based cloning experiments (Schmid 2000). Our results suggest that this may also hold for teleost fishes.

Microsatellite markers provide both a system for polymorphism analysis and a system for anchoring the locus via the sequences that flank the microsatellite repeat. However, since microsatellites normally reside within non-coding regions, it is often thought that they can only be matched with relatively closely related species. Interestingly, Rico *et al.* (1996) had already found that a given microsatellite locus can be amplified across a large range of fish taxa. We found that almost half of the flanking sequences from *Cottus* yield a significant match with Fugu and *Tetraodon*. Intriguingly, the matches occur with highly conserved short stretches of unknown function. Given the large number of hits that we have detected, it would seem that the density of such conserved non-coding regions is very high in these fish genomes. While it is generally interesting to speculate about the functional role of these sequences (Gaffney and Keightley 2004), they also turn out as potentially highly useful tools for linking genome information between diverse fish species.

Given the known partially conserved synthenies even between mammal and fish genomes (Grützner *et al.* 2002; Jaillon *et al.* 2004), it is not surprising that we find evidence for highly conserved synteny between the fish genomes themselves. However, it is reassuring that even such a simple map construction strategy as the one that we have employed, in conjunction with an only partially annotated genome such as *Tetraodon*, already yields clearly comparable chromosomal parts. Nevertheless we have to consider the occurrence of intrachromosomal rearrangements, which do not allow direct transfer of all positional information from the *Tetraodon* to the *Cottus* genome. Still, because of the apparent high density of conserved sequence elements, it will be possible to trace microsynthenic relationships, even if the whole chromosome segment is rearranged, or fused to another chromosome. We have already started QTL mapping in *Cottus* to define regions possibly associated with differential adaptations. The microsatellite markers that are linked to QTL can then be used to directly identify candidate genes from the *Tetraodon* genome, which can be employed in further fine mapping strategies.

Figure 2 shows a sketch of the phylogenetic relationships between the major fish lineages. Fugu, *Tetraodon* and *Cottus* belong to the Acantopterygii (Nelson 1994), which include also medaka (Oryzias latipes) as a further genome for which full sequence information will soon be available. The interrelationships within the Acantopterygii are still under debate, but both Nelson (1994) and Miya (2003) agree that *Cottus* (Scorpaeniformes) is more closely related to the *Tetraodon*tiforms (Takifugu rubripes, *Tetraodon nigroviridis*) than to the Atheriniforms (Oryzias latipes). The other major model fish, Danio rerio, belongs to the Ostariophysi. Given that we find about a quarter of the *Cottus*/*Tetraodon* matches even in Danio, it would seem that it will be straight foward to link genetic markers that are found in any of these teleost fish species to known genome information of one of the model organisms.

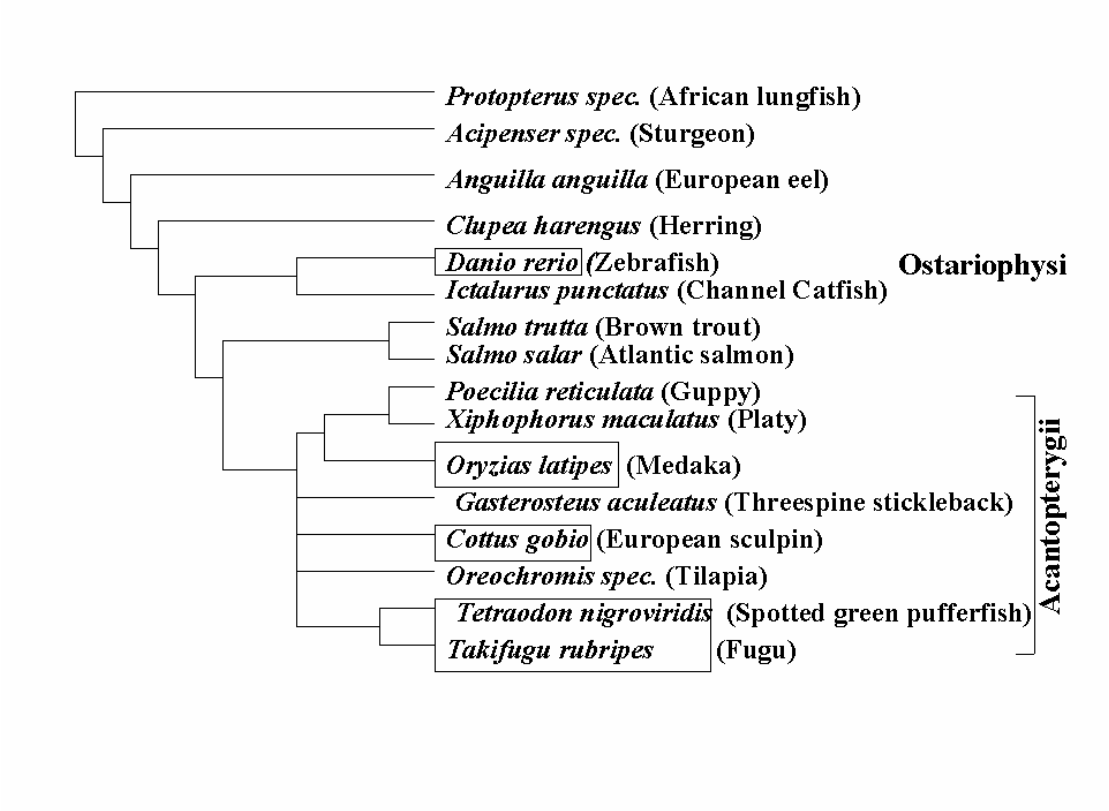


Figure 2: Schematic cladogram illustrating the relativ phylogenetic positions of model fish species such as Danio, Oryzias, *Cottus*, *Tetraodon* and Fugu among other teleost fishes of special interest. Based on Nelson *et al.* (1994) and Miya *et al.* (2003)

Acknowledgements

We thank A. Mellin (Obere Fischereibehörde Köln), W. Fettweis and T. Heilbronner (Siegfischereigenossenschaft) for permission to sample *Cottus* specimen, as well as Susanne Kip and Birgit Schmitz for technical support. This work was funded by a DFG grant to D.T. and the Fond der chemischen Industrie.

Literature

Altschul, S. F. Gish, W. Miller, W. Myers, E. W. Lipmann, D. J. (1990) BLAST: ‚Basic local alignment search tool'. J. Mol. Biol. 215: 403-410.

Barton, N. Gale, K. S. (1993) Genetic analysis of hybrid zones. In: Hybrid zones and the evolutionary process (R. G. Harrison, ed.), pp. 13-45. Oxford University Press, New York, NY.

Chowdhary, B. P. Raudsepp, T. Frönicke, L. Scherthan, H. (1998) Emerging patterns of comparative genome organization in some mammalian species as revealed by zoo-FISH. Genome Res. 8: 577-589.

Dolezel, J. Batos, J. Vogelmayr, H. Greilhuber, J. (2003) Nuclear DNA content and genome size of trout and human. Cytometry 51A: 127-128.

Dominguez, I. Graziano, E. Gebhard, C. Barakat, A. Berry, S. Arus, P. Delseny, M. Barnes, S. (2003) Plant genome archeology: evidence for conserved ancestral chromosome segments in dicotyledonous plant species. Plant Biotechnol. J. 1: 91-99.

Englbrecht, C. C. Largiader, C. R. Hänfling, B. Tautz, D. (1999) Isolation and characterization of polymorphic microsatellite loci in the European bullhead *Cottus gobio* L. (Osteichtyes). Mol. Ecol. 8: 1966-1968.

Englbrecht, C. C. Freyhof, J. Nolte, A. Rassmann, K. Schliewen, U. Tautz, D. (2000) Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggest a pre-Pleistocene origin of the major central European populations. Mol. Ecol. 9: 709-722.

Erickson, D. L. Fenster, C. B. Stenoien, H. K. Prices, D. (2004) Quantitative trait locus analysis and the study of evolutionary process. Mol. Ecol. 13: 2505-2522.

Gaffney, D. J. Keightley, P. D. (2004) Unexpected conserved non-coding DNA blocks in mammals. Trends in Genetics 20: 332-337.

Garbe, J. Da, Y. (2003) Locusmap user manual, Version 1.1 (Department of Animal Science, University of Minnesota).

Gebhardt, C. Walkemeier, B. Henselewski, H. Barakat, A. Delseny, M. Stüber, K. (2003) Comparative mapping between potato (*Solanum tuberosum*) and *Arabisopsis thaliana* reveals structurally conserved domains and ancient dupliations in the potato genome. The Plant J. 34: 529-541.

Giampietro, P. F. Raggio, C. L. Blank, R. D. (1999) Synteny-defined candidate genes for congenital and idiopathic scoliosis. Am. J. Med. Gen. 83: 164-177.

Grützner, F. Lütjens, G. Rovira, C. Barnes, D. W. Ropers, H. Haaf, T. (1999) Classical and molecular cytogenetics of the pufferfish *Tetraodon nigroviridis*. Chromosome Res. 7: 655-662.

Grützner, F. Crollius, H. R. Lutjens, G. Jaillon, O. Weissenbach, J. Ropers, H. H. & Haaf, T. (2002) Four-hundred million years of conserved synteny of human Xp and Xq genes on three *Tetraodon* chromosomes. Genome Res. 12: 316-322.

Hardie, D. C. & Hebert, P. D. N. (2003) The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. Genome, 46: 683-706.

Jaillon, O. Aury, J. Brunet, F. Petit, J. Stange-Thomann, N. Mauceli, E. Boneau, L. Fischer, C. Ozouf-Costaz, C. Bernot, A. Nicaud, S. Jaffe, D. Fisher, S. Lutfalla, G. Dossat, C. Segurens, B. Dasilva, C. Salanoubat, M. Levy, M. Boudet, N. Castellano, S. Anthouard, V. Jubin, C. Castelli, V. Katinka, M. Vacherie, B. Biemont, C. Skalli, Z. Cattolico, L. Poulain, J. de Beradinis, V. Cruaud, C. Duprat, S. Brottier, P. Coutanceau, J. Gouzy, J. Parra, G. Guigo, R. Zody, M.C. Mesirov, J. Lindblad-Toh, K. Birren, B. Nusbaum, C. Kahn, D. Robinson-Rechavi, M. Laudet, V. Schlachter, V. Quetier, F. Saurin, W. Scarpelli, C. Wincker, P. Lander, E.S. Weissenbach, J. & Crollius, H.R. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431: 946-957.

Kauer, M.O. Dieringer, D. and Schlötterer, C. (2003) A Microsatellite variability screen for positive selection associated with the "out of Africa" habitat expansion of Drosophila melanogaster. Genetics. 165: 1137-1148.

McKeigue, P. (2005) Prospects for Admixture Mapping of Complex Traits. Am. J. Hum. Genet. 76:1–7.

Miya, M. Takeshima, H. Endo, H. Ishiguro, N. B. Inoue, J.G. Mukai, T. Satoh, T. P. Yamaguchi, M. Kawaguchi, A. Mabuchi, K. Shirai, S. M. Nishida, M. (2003) Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol. Phyl. Evol. 26: 121-138.

Morgenstern, B. (1999) DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics 15: 211-218.

Nelson, J. S. (1994) Fishes of the world, third ed. Wiley, New York.

Nolte, A. W. Stemshorn, K. C. and Tautz, D. (2005) Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Mol. Ecol. Notes (*in press*)

Rico, C. Rico, I. Hewitt, G. (1996) 470 million years of conservation of microsatellite loci among fish species. Pro. R. Soc. Lond. B 263: 549-557.

Schlötterer, C. (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365-371.

Schmid, R. (2000) Synteny: recent advances and future prospects. Cur. Op. Plant Biol. 3: 97-102.

Vitturi, R. Rasotto, B. (1990) Karyotype analysis of *Cottus gobio* L. (Pisces, Cottidae). Cytobios 62: 81-86.

Voorrips, R. E. (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. 93: 77-78.

Chapter 4: Rapid moulding of nascent hybrid zones results from differential adaptation of two lineages of sculpins

Arne W. Nolte, Jörg Freyhof and Diethard Tautz

Abstract:

The initial formation of hybrid zones is an exceptional event. It is a priori unclear if and how fast a nascent hybrid zone would mould as a response to endogenous and exogenous forces. We witnessed the formation of secondary contact among different lineages of sculpins resulting from a recent range expansion of one of them. The age of these hybrid zones postdates the appearance of “invasive” sculpins within the River Rhine in the early 1990’s. Applying a dense sampling and a highly informative set of microsatellite markers we found a genetic structure that implies an instantaneous stabilisation due to strong selection. Extremely narrow hybrid zones are coupled to sharp ecological transitions from small streams to larger rivers. The width of these is smaller than estimates of annual individual dispersal distances. The pattern cannot be explained by pre- or postzygotic reproductive barriers as numerous hybrids occur locally. This suggests strong natural selection against immigrant genotypes and the presence of adaptive traits. The view the involved lineages diverge ecologically is supported. Sculpin hybrid zones are asymmetric in that different processes are acting on alternative sides, implying a complex architecture rather than a single underlying trait.

Introduction:

Less than 20 years ago populations of the sculpin *Cottus gobio* L. 1758 (Cottidae, Scorpaeniformes; Teleostei) were discovered in the main channel of the River Rhine, artificial canals and the Ijsselmeer. These waters represent summer warm potamal habitats. This sudden range expansion was unexpected as sculpins are commonly thought to be specialised to upstream areas where they are characteristic for the trout region (rithral). Own studies have now shown that the sculpins that have appeared within potamal habitats may have resulted from a hybridization process because they carry a hybrid genepool derived from ancient phylogeographic lineages (Chapter 1). Moreover, a previously unknown tolerance to summerwarm and turbid waters within the River Rhine basin results in an exclusive habitat use. This suggests that the hybrids actually possess unique adaptations to the respective habitats

that could have arisen from hybridization events (Chapter 1). However the nature of these adaptations is not understood as yet and evidence is limited to patterns of differential colonisation. Any attempts to further explain the recent invasion and possible impacts of hybridization rely on a better understanding of the reasons why the hybrid lineage is so successful, i.e. a demonstration of differential adaptation and ultimately on the identification of adaptive traits.

In this study we analyse natural experiments to gather evidence for differential adaptation for divergent lineages. This is a promising strategy as the development of natural populations represents the most integrative measure of evolutionary success. The focus of this study is on hybrid zones in which the invasive sculpin (being of hybrid origin itself) backcrosses with representatives of one of its ancestral lineages, namely the Lower Rhine stream sculpin (Chapter 1). Secondary contact of sculpin lineages resulted from a rapid upstream range expansion of invasive sculpins within large rivers thereby approaching local populations. During the course of the expansion populations of stream sculpins remained confined to their native streams leaving downstream habitats vacant even in the absence of invasive sculpins (Chapter 1). Thus, multiple contact zones have formed in parallel with independent local populations.

Below we discuss features of these newly formed sculpin hybrid zones. Their age postdates the recent appearance of invasive sculpins in the early 1990's and offers the unique opportunity to study the initial phase of secondary contact. It is generally accepted that hybrid zones may form at ecotones but their initial moulding is an exceptional event. While the formation of secondary contact at established range boundaries of local populations already suggests a dominant role of ecological factors, it is a priori unclear if and how fast the structure of the hybrid zone itself would follow patterns expected from theory. For example a tension zone model would predict a random initial position and that the zone width would depend on individual dispersal ability (Barton and Gale 1993). In contrast ecotone models assume a tight coupling of the hybrid zone centre to environmental gradients and a width being mainly determined by patterns of exogenous selection (Barton and Gale 1993). The patterns we observe imply an instantaneous stabilisation of hybrid zones due to strong adaptive differentiation. This suggests the presence of adaptive traits and adds evidence that the hybrid lineage strongly diverges ecologically from its ancestors. Moreover, the situation found at the hybrid zones can be best explained by different processes acting on alternative sides, implying a complex architecture rather than a single underlying trait.

Methods:

Sampling:

Important findings on the population dynamics of sculpins were made in studies of the fish community of the River Sieg since 1993 (Freyhof 1997, unpublished data). Together with genetic studies (Chapter 1) the data

provided a good starting point for a study of hybrid zones. We have sampled specifically for sulpins annually between 2000 and 2004 in spring, summer and autumn. Streams and rivers given in table 1 were surveyed repeatedly at several sites. Of particular importance is a survey of the Stream Wahnbach. Since 1958, a drinking water reservoir divides this stream into upper reaches and a lower part (2.1 km) draining into the Sieg. We have sampled the lower Wahnbach as well as its upper reaches since 2000. The lower Wahnbach was also surveyed intensively on one occasion in 1995. Sampling was done with portable electroshockers. Specimens for genetic analysis were fixed in 70% ethanol. In this study we use “stream” to refer to small summercold streams that correspond largely to the salmonid region (rhitral) while “river” is used for the summerwarm cyprinid region (potamal). Initially, we have screened for morphological transitions of “skin prickling” (Chapter 1) and screened six microsatellites taken from Englbrecht et al (1999). Results from morphology and genetics were highly congruent and allow to identify areas of hybridization (Tab. 1). For an in depth genetic analysis we sampled two hybrid zones at the outlets of the Stream Broel into the River Sieg and the Stream Naaf into the River Agger (tributary to the Sieg) intensively (Tab. 1, Fig. 2, sampling sites 1-11).

Table 1: Sample reference numbers, Sampling sites, GIS data and Group N. The classification given for 1-11 is based on the complete genetic analysis on this study. Specimens from 12-17 were screened using fewer markers or based on morphological characters. The occurrence of pure populations above or below the hybrid zones was verified for all sites classified as hybrid zones.

Ref.	Site	GIS	N	Classification
1	Stream Broel between Broel and Winterscheidt, North Rhine-Westphalia, Germany	50°47'N 7°20'E	48	Stream sculpins
2	Stream Broel south of Broel, North Rhine-Westphalia, Germany	50°47'N 7°19'E	48	Stream sculpins
3	Stream Broel at Mueschmuehle, 200 m above outlet into River Sieg, North Rhine-Westphalia, Germany	50°47'N 7°18'E	130	hybrid zone
4	River Sieg at Allner, below outlet of Stream Bröl, North Rhine-Westphalia, Germany	50°46'N 7°18'E	36	hybrid zone
5	Stream Wahnbach, Outlet into River Sieg at Seligenthal, North Rhine-Westphalia, Germany	50°47'N 7°16'E	4	Invasive sculpin
6	Stream Pleis, outlet into the River Sieg at Niederpleis, North Rhine-Westphalia, Germany	50°46'N 7°12'E	5	Invasive sculpin
7	River Sieg at Muehlenbach, North Rhine-Westphalia, Germany	50°47'N 7°10'E	35	Invasive sculpin
8	River Agger at Donrath, 900 m below outlet of stream Naaf, North Rhine-Westphalia, Germany	50°51'N 7°13'E	30	Invasive sculpin
9	Stream Naaf, Outlet into River Agger, North Rhine-Westphalia, Germany	50°51'N 7°14'E	48	hybrid zone
10	Stream Naaf at Kreuznaaf, 200 m above outlet into River Agger, North Rhine-Westphalia, Germany	50°51'N 7°14'E	48	hybrid zone
11	Stream Naaf southeast of Hausdorp, North Rhine-Westphalia, Germany	50°52'N 7°16'E	48	Stream sculpins
12	Outlet of Stream Ottersbach at Eitorf, North Rhine-Westphalia, Germany	50°46'N 7°28'E	96	hybrid zone; inferred from 6 microsatellite loci and morphology
13	Outlet Stream Mengbach in the west of Eitorf, North Rhine-Westphalia, Germany	50°46'N 7°25'E	96	hybrid zone; inferred from 6 microsatellite loci and morphology
14	Outlet Stream Krabach at Bach, North Rhine-Westphalia, Germany	50°45'N 7°24'E	210	hybrid zone; inferred from morphology
15	Stream Pleisbach at Dambroich; North Rhine-Westphalia, Germany	50°44'N 7°14'E	144	hybrid zone; inferred from 6 microsatellite loci and morphology

16	Stream Flaumbach south of Treis-Karden, Tributary to the River Mosel, Rhineland-Palatinate, Germany	50°11'N 7°17'E	96	hybrid zone; inferred from 6 microsatellite loci and morphology (not shown)
17	Stream Wisper east of Lorch, Tributary to the Rhine, Taunus, Hesse, Germany	50°3'N 7°51'E	60	hybrid zone; inferred from morphology (not shown)

Populationgenetic analysis:

DNA was extracted from fin clips of specimens using a standard proteinaseK/chloroform method. PCR's were done for individual loci using Eurobio Taq polymerase (Eurobio) using buffers supplied by the vendor. Alternatively, up to 6 fluorescently labeled (Fam, Hex, Tet) primer pairs were combined and amplified using the Multiplex-PCR Kit (Quiagen) as described in Nolte *et al.* (2005). The loci were combined in a way such that all fragments could be separated in a single lane without overlap and scored unambiguously. Fragments were scored on a MegaBACE 1000 capillary sequencer (Amersham pharmacia). Allele editing was done with the GENETIC PROFILER 2.2 (Amersham Biosciences).

Besides markers from Englbrecht *et al.* (1999), we have identified additional microsatellite loci (Nolte *et al.* 2005) and typed 171 loci for samples from our target populations, namely stream sculpins from the streams Naaf and Broel and Invasive sculpins from the Lower Sieg. Such a large dataset contains microsatellites with surpassing information content for an analysis of admixture among as well as physically linked loci. In order to maximize population genetic resolution and to minimize confounding effect introduced by physical linkage we chose markers with the highest possible allele frequency differential among populations following Shriver *et al.* (1997). The loci were chosen for their particularly high information content for our study system using WHICLOCI (http://www.bml.ucdavis.edu/whichloci.htm). Furthermore, we have used a preliminary genetic map of *Cottus* (Stemshorn *et al.* 2005), to verify that our set of markers does not contain pairs of loci that are tightly physically linked. This screen allowed us to assemble a set of 45 microsatellite loci with an average between population allele frequency differential of 88%, none of which was linked more closely than 18 cM (Chapter 4 - suppl. Tab. 1).

STRUCTURE 2.1 (Falush *et al.* 2003) was used to infer the most likely population structure based only on genotypic data. We have run several chains (burnin: 20000; sampling iterations 100000, correlated allele frequency model allowing for an individual alpha and different F_{ST} for each subpopulation) for each value for K (number of populations) between 1 and 7. The log-likelihood of the data rises strongly up to a value of K=3. For K>3 there is a strong tendency to plateau (Fig. 1). Thus, we consider a value of K=3 the most sensible model following the recommendations given in the STRUCTURE manual and in Falush *et al.* (2003) and because of the perfect fit with morphological, ecological and distribution data (Chapter 1). Independently of run length, the MCMC chains converged to alternative solutions for runs with K>3 some of which resulted in slightly better LnP(D) values for K=4. In these runs a subdivision of samples from the invasive lineage according to two sampled rivers Sieg and Agger was suggested, which would be in line with geographic isolation of the two separate rivers.

Alternatively, invasive sculpins were assigned to one cluster with some foreign genetic material distributed among them. This interpretation of K=4 would imply that there are unknown populations involved, i.e. by recent immigration from additional unknown populations. Although this scenario cannot be ruled out, the two stream populations and the fraction of genetic material in admixed individuals were consistently recovered runs where K=3 and K=4. With respect to the analysis of hybrid zones of stream and invasive sculpins this suggests that the inference of stream ancestry is quite robust against alternative interpretations of the population structure of the invasive lineage. To visualize the genetic transition across hybrid zones individual ancestry estimates for K=3 from STRUCTURE (Chapter 4 - suppl. Tab. 1) were plotted against geographical distance from the stream river confluence. To estimate the centre and width of multilocus clines, as approximated by the inverse of the maximum slope (Barton and Gale 1993) a sigmoidal (3 parameter) curve was fitted using SIGMAPLOT 2000 v.6.10 (SPSS inc.).

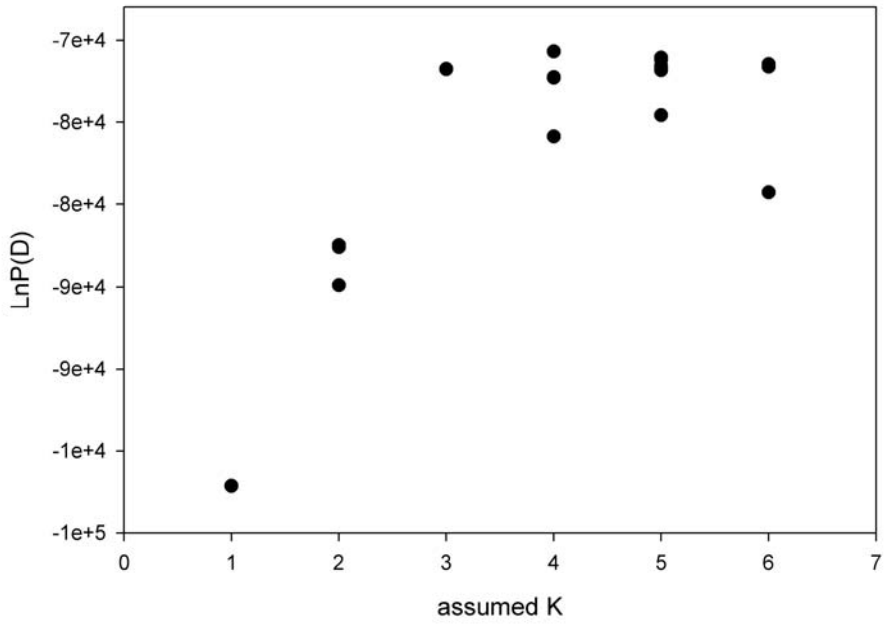


Figure 1: Plot of the likelihood of all genotypic data (LnP(D)) under a model assuming K populations (5 replicates per K). The probability rises sharply as K increases but plateaus at K=3. This as well as morphological and distributional data (see text) suggest that there are three populations (Stream Naaf, Stream Broel and Invasive sculpin) that have contributed to the total genotypic data.

NEWHYBRIDS 1.1beta (Anderson 2002) as of November 2004 was used to analyse sub samples. Here the aim was to analyse specimens that were from single localities close to the centre of the two hybrid zones. The genetic composition of these samples with respect to pure ancestral types, F_1 or F_X hybrids and backcrosses was determined. We chose well-sampled sites from both hybrid zones that span less than 50 m of continuous habitat and contain

numerous hybrids. For this purpose pure specimen from above and below the hybrid zones were included as learning samples (Anderson 2002) in agreement with results from structure to make use of the information content from the complete dataset.

Results:

A common pattern of local hybridization was found repeatedly at independent outlets of tributaries throughout the lower and middle Rhine basin (Fig. 2 and Tab. 1, Sites: 3, 9, 12-17). Upstream samples taken above the areas where hybrids occur always revealed populations of stream sculpins while main rivers were typically inhabited by invasive sculpins. The area of intense hybridization was usually restricted to the outlets of streams with the exception of Stream Pleisbach where the hybrid zone is displaced roughly 4.5 km upstream. The Stream Pleisbach (6) and the Lower Stream Wahnbach (5) represent the only known streams that were at least partially colonized by invasive sculpins (Fig. 2). While older surveys or material for the Pleisbach is unfortunately missing our survey data for the lower Wahnbach show that this area was completely free of sculpins in 1995. With respect to the macrohabitat, the lower stream Wahnbach resembles typical trout streams of our study area in size and fauna. Since stream sculpins persist in the upper reaches of stream Wahnbach we assume that they went extinct in the lower reaches due to the construction of a drinking water reservoir (finished 1958), which still isolates the lower reaches of this stream. Apparently, an upstream colonization occurred from the main Sieg as we found invasive sculpins that founded a reproducing population in the lower Wahnbach since 2000.

The genetic structure observed at stream outlets shows that the areas of secondary contact represent thoroughly admixed hybrid zones. The genetic structure of the Broel- and the Naaf hybrid zones was analysed by plotting individual ancestries in the two stream populations against river distance (Fig. 3). In both cases a steep multilocus cline is situated in the outlet area. Cline width is estimated to be 1,95 km for the stream Broel and 2,3 km for the stream Naaf. The centre of the clines is displaced 370 metres upstream of the position of outlet of the Broel and 108 m downstream of the outlet of the Naaf. Streams and rivers on alternative sides of the hybrid zones generally harbour ancestral genotypes with few immigrants in the vicinity of the hybrid zones. Hybrids as well as backcrosses are largely confined to outlet areas.

Within the hybrid zones pure ancestral genotypes as well as hybrid genotypes occur syntopically at the same sites. Intermediate individual ancestry coefficients do not fall into discrete classes but take continuous values (Fig. 3). The fact that several classes of hybrids and backcrosses are present shows that continuous admixture occurs in natural populations. Proportions of pure types and hybrid classes at selected sites are given in Table 2. Although pure genotypes representing both ancestral lineages occur in both hybrid zones F_1 hybrids are rare. F_X hybrids and backcrosses dominate among the admixed individuals. Among the backcross genotypes, backcrosses with the invasive lineage are more abundant than backcrosses with stream sculpins (Tab. 2).

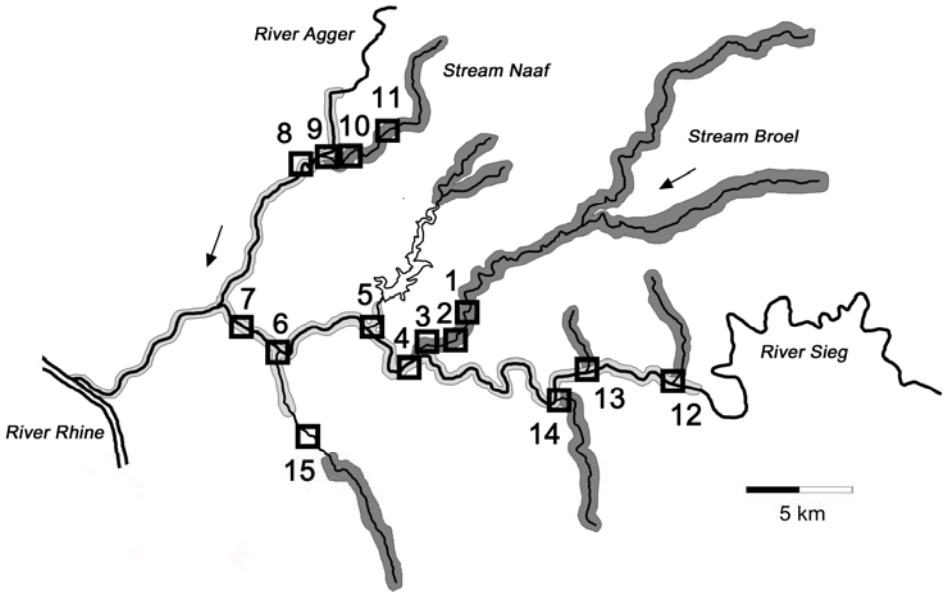


Figure 2: Map of the River Sieg and smaller tributaries as they drain into the River Rhine. The current distribution of invasive sculpins (light grey) results from an upstream invasion since 1993 while stream sculpins (dark grey) are restricted to small tributaries (see Chapter 1). Samples used for a detailed analysis of two hybrid zones are numbered from 1- 11. These and additional hybrid zones (12-14) were found to be situated at outlets of small tributaries into rivers. Still, there is one hybrid zone (15) that is displaced upstream within Stream Pleisbach. Note that the lower Stream Pleisbach and Stream Wahnbach (5) are colonized by invasive sculpins.

Table 2: Genetic composition of samples taken from selected sites of two hybrid zones. All possible hybrid classes were found, yet among the pool of admixed individuals F_1 hybrids are rare and backcrosses with the invasive sculpin are particularly abundant. Ratios demonstrate the relative abundance of pure invasive vs. stream sculpins as well as invasive sculpin backcrosses.

Category	Site 3	Site 10	Site 9
Stream-pure	0.24	0.67	0.11
Invasive-pure	0.234	0.02	0.07
F_1	0.03	0.02	0.04
F_X	0.15	0.07	0.20
Stream-Backcross	0.10	0.20	0.31
Invasive-Backcross	0.24	0.02	0.26
Ratio Stream / Stream-Backcross	2.28	3.29	0.36
Ratio Invasive / Invasive-Backcross	0.95	1.02	0.28

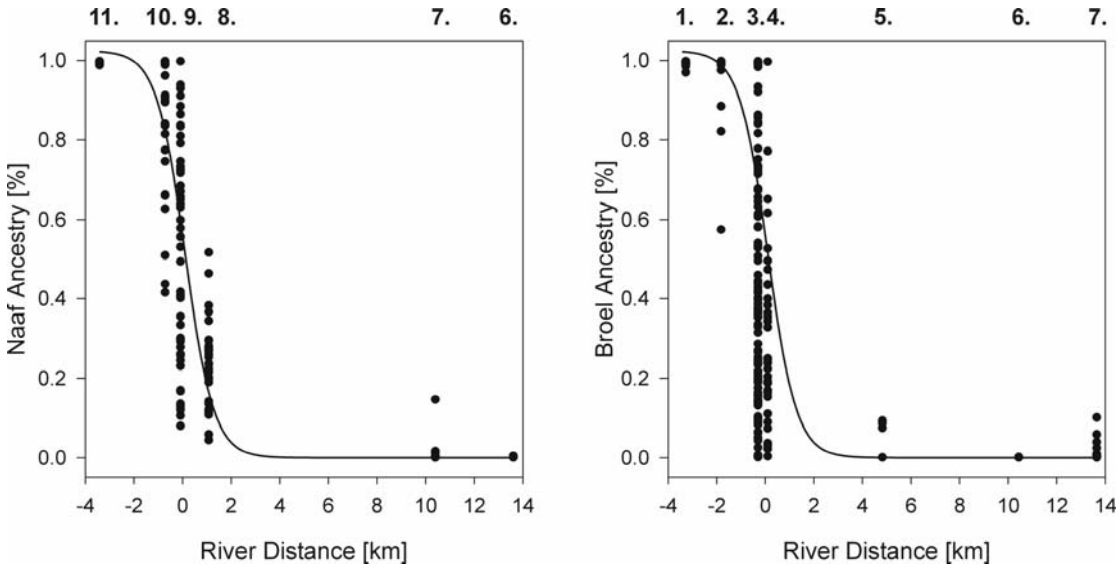


Figure 3: Individual ancestry coefficients observed across the ecological transition zones from small streams to adjacent rivers. Samples representing a transect from streams to riverine habitats are plotted from left to right. Areas of intense hybridization are restricted locally while pure stream- and invasive sculpins occur above and below the outlet areas. Ancestry coefficients in alternative streams were calculated based on a total of 480 specimens typed for 40 highly informative microsatellites. Site numbers above graphs correspond to the map (Fig. 1).

Discussion:

Provided that hybrid zone width is determined by the interplay of selection and dispersal (Barton and Gale 1993) the forces that have moulded the hybrid zone have already left obvious short-term traces. Secondary contact among the invasive sculpin and native stream sculpins is younger than 12-15 years (Chapter 1) but must have been established for several generations as seen from the presence of backcross genotypes. The average upstream dispersal rate of invasive sculpins within the River Sieg was inferred to exceed 4 km per year (Chapter 1). Yet the hybrid zones at stream outlets are amazingly narrow, extending over a distance of roughly 2 km (Fig. 3).

If only genetic incompatibilities would exist between the lineages (tension hybrid zone) (Barton and Gale 1993), one would expect hybrid zones determined primarily by the dispersal rate, which is larger than the hybrid zone width for the invasive lineage. However, the zones are narrower than would be expected if there were random admixture and dispersal. Most importantly, multiple hybrid zones have formed in a similar fashion at tributaries draining to the River Sieg, the River Mosel or the River Rhine (Tab. 1). The repeated positioning at outlets of trout streams suggests a key role of the habitat transition from the trout region (rhithral) to the cyprinid region (potamal). Even though one can only speculate about the long-term development of nascent hybrid zones the current patterns imply an instantaneous process that counteracts dispersal and admixture upon secondary contact. Furthermore, the width and shape of the genetic clines argues for strong selection against immigrant genotypes. Thus, the overall pattern corresponds to an ecotone hybrid zone model (Barton and Gale

1993). With respect to the lineages that are involved this implies the presence of alternative adaptations and provides general evidence for ecological divergence.

While the general pattern is remarkably clear those factors acting to shape the sculpin hybrid zones are still unknown. The distribution of the involved lineages with and without the influence of secondary contact suggest that different factors and processes may be important at alternative sides of the outlet areas. Stream sculpins are confined to small streams and can therefore be assumed to represent locally specialized types simply because they persist exclusively in those streams. Such statements cannot be made for the invasive lineage because it is still in a dynamic process of population expansion (Chapter 1). Neither equilibrium nor a stable range or fixed habitat requirements can be inferred for the latter. Consequently, a formation of hybrid zones at established range margins of stream sculpins suggests that the ecological factors that limit the distribution of stream sculpins are independent of competition with invasive sculpins (Chapter 1), i.e stream sculpins were not able to colonise downstream habitats even in the absence of the invasive competitor. The invasive sculpin apparently colonizes rivers successfully up to the point where secondary contact is established. Note that we have found the invasive lineage to successfully colonize a small stream (rhithral) (Lower Wahnbach) in the absence of stream sculpins. Thus secondary contact and inter lineage competition with locally adapted types appears to be an important determinant limiting the distribution of invasive sculpins in contrast to pure habitat specificity. The hybrid zones observed in sculpins are therefore asymmetric in that the ecological processes and selective forces shaping alternative sides most likely differ.

A presence of numerous hybrids and backcrosses at the centres of the hybrid zones (Fig. 2) suggests that there is no absolute premating barrier to reproduction in natural habitats. The same observation also indicates that admixture is not prevented by a general postzygotic hybrid inviability. Despite this principal viability, hybrids are largely confined to the core area of the contact zones. This spatial restriction strongly suggests that hybrids are unable to compete with either one of the parental types outside of the ecological transition zones. Thus, our data are best explained by exogenous selection that limits hybrid fitness particularly in direct competition with their ancestors in ancestral habitats. The centre of the Broel zone is displaced upstream but this is not true for the Naaf zone where the inferred centre of the cline is situated slightly downstream of the confluence. Together with the slightly wider cline observed at the Naaf zone this may indicate less pronounced selection against foreign genetic material in the River Agger.

Following Strauss (1986) there are signs of non-random mating or selective mortality in the hybrid zones studied here although this depends entirely on the assumption that no bias is introduced by migration. The genotypic classes observed at hybrid zones are skewed towards invasive sculpin backcrosses and F_1 hybrids are particularly rare in both hybrid zones (Tab. 2) relative to the abundance of the parental phenotypic classes. This would not be expected under random inter- and intra group mating. Most likely

this assortativity does not involve fixed mate choice behaviour because invasive and stream sculpins can be easily crossed in aquarium experiments (Stemshorn *et al.* 2005). Preliminary observations suggest that the onset of spawning in stream sculpins may be slightly later than in invasive sculpins, nevertheless with a large overlap in spawning time (Freyhof and Nolte unpublished) which would explain some degree of assortative mating, especially if the timing of spawning would be a dominant trait inherited by invasive sculpins. Another reason to explain non-random genotypic classes relates to selective mortality. If true, this would be most likely related to external selection as we have not observed signs of hybrid inviability in several F_1 labcrosses.

Sculpin ecotone hybrid zones are coupled with a sharp environmental gradient, namely the transition from the rithral to the potamal subregion within continuous bodies of water. Generally, a sharp transition of ecological conditions is rather uncommon in rivers but may occur linked to sharp physical transitions, for instance at confluences of small streams into larger rivers (Vannote *et al.* 1980). Available survey data on the fish community of the River Sieg (Freyhof 1997, Freyhof unpublished) and its tributaries (data from LÖBF, North Rhine Westphalia) and on the macroinvertebrate community of the rivers and streams studied here (data from STUA, Köln) suggest that outlets indeed represent a border of divergent habitats. Unfortunately, little is known about the width of these ecological transitions. Beckmann *et al.* (2004) studied tributaries to the middle Rhine and provide evidence that abrupt faunal transitions exist directly above confluences. Note that we have tentatively identified one hybrid zone (Tab. 1, Stream Wisper) that was situated at one of the confluences for which Beckmann *et al.* (2004) have also found an invertebrate faunal transition. An ecotone model would predict less pronounced clines at hybrid zones spanning areas where the habitat transition is less steep. Preliminary data based on the transition of phenotypic markers indicate that this may be the case for the Pleisbach hybrid zone (Nolte unpublished). This stream differs also in that the position of the hybrid zone is apparently displaced upstream approximately 4.5 km.

The stream-river ecotone extends practically unidirectional along the path of the river. This simple one-dimensional structure is probably not reflected in the processes stabilizing the hybrid zones. While the streamside is apparently maintained by an intraspecific competition component, the riverside may be entirely facilitated by other ecological factors. Thus, direct competition between Stream and Invasive sculpins may play a big role in stream habitats, but not in adjacent rivers. Different interactions and factors apparently limit the expansion of populations beyond hybrid zones. Consequently, a given adaptive trait may not be subject to equally strong selection on both sides of the hybrid zone. Therefore the key adaptations that convey ecological success on alternative sides of the hybrid zones may differ and may not necessarily be functionally linked.

Acknowledgements:

B. Nürnberger has made helpful suggestions on earlier versions of the manuscript and U. Heckes helped in the interpretation of faunal transitions from streams to rivers. C. Schütz (LÖBF, North Rhine Westphalia) and G. Eckartz-Nolden (STUA Köln) have provided data on the presence of fish and invertebrates within the study area. This work was possible only with benevolent support and permissions from A. Mellin, T. Heilbronner, W. Fettweis, L. Jörgensen and C. Köhler. Collaboration with I. Steinmann, J. Ennenbach and E. Korte has facilitated much of the fieldwork. This study was financed by the Deutsche Forschungsgemeinschaft and Verband der Chemischen Industrie.

Literature:

Anderson, E. C. Thompson, E. A. (2002) A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160: 1217–1229.

Barton, N. Gale, K. S. (1993) Genetic analysis of hybrid zones. In: Hybrid zones and the evolutionary process, edited by R. G. Harrison, Oxford University Press (1993), New York.

Beckmann, M. C. Schoell, F. Matthaei, C. D. (2004) Effects of increased flow in the main stem of the River Rhine on the invertebrate communities of its tributaries. Freshwater Biology, doi:10.1111/j.1365-2427.2004.01289.x

Englbrecht, C. C. Largiader, C. R. Hänfling, B.Tautz, D. (1999) Isolation and characterization of polymorphic microsatellite loci in the European bullhead *Cottus gobio* L. (Osteichthyes) and their apllicability to related taxa. Molecular Ecology, 8, 1966 - 1969.

Falush, D. Stephens, M. Pritchard, J. K. (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics, 164: 1567-1587.

Freyhof, J. (1997) Ökologische Zusammenhänge zwischen Lebensraumstrukturen und der Fischartengemeinschaft der Sieg, Untersuchung im Auftrag des Ministeriums für Umwelt, Raumordnung und Landwirtschaft des Landes Nordrhein Westphalen, pp. 1-189.

Nolte, A. W. Stemshorn, K. C. and Tautz, D. (2005). Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Molecular Ecology Notes (*in press*).

Shriver, M. D. Smith, M. W. Jin, L. Marcini, A. Akey, J. M. Deka, R. Ferrell, R. E. (1997) Ethnic-affiliation estimation by use of population-specific DNA markers. American Journal of Human Genetics, 1997, 60(4): 957-964.

Stemshorn, K. C. Nolte, A. W. Tautz, D. (2005). A Genetic Map of *Cottus gobio* (Pisces, Teleostei) based on microsatellites can be linked to the Physical Map of *Tetraodon nigroviridis*. J. Evol. Biol (*in press*)

Vannote, R. L. Minhall, G. W. Cummins, K. W. Sedell, J. R. Cushing, C. E. (1980) The river continuum concept. Can. J. Fish. Aquat. Sci. 37: 130-137.

Appendix – Chapter 4

Chapter 4 - Supplementary Table 1: Table containing individual genotypic data for 480 sculpins from sampling sites as described in Table 1. “SiteNumber” refers to Ref. in Table 1. Individual ancestry estimates from STRUCTURE are given for all three populations. Microsatellite alleles with locus names for 45 Loci are numbered, “0” indicates missing data. Data format: Table; saved row by row with fields separated by semicolons. Ends of rows are marked by the insertion of “XXX”.

Chapter 5: Shape based assignment tests reveal transgressive phenotypes in natural sculpin hybrids (Teleostei, Scorpaeniformes, Cottidae)

Arne W. Nolte and H. David Sheets

Abstract

Hybridization receives attention because of the potential role that it may play in generating evolutionary novelty. An explanation for the emergence of novel phenotypes is given by transgressive segregation, which, if frequent, would imply an important evolutionary role for hybridization. This process is still rarely studied in natural populations as samples of recent hybrids and their parental populations are needed. Further, the detection of transgressive segregation requires phenotypes that can be easily quantified and analysed. We analyse variability in body shape of divergent populations of European sculpins (*Cottus gobio* complex) as well as natural hybrids among them. A distance-based method is developed to assign unknown specimens to known groups based on morphometric data. Apparently, body shape represents a highly informative set of characters that parallels the discriminatory power of microsatellite markers in our study system. Populations of sculpins are distinct and “unknown” specimens can be correctly assigned to their source population based on body shape. Furthermore, recent hybrids are indermediate along the axes separating their parental groups but display additional differentiation that is unique and coupled with the hybrid genetic background. There is a specific hybrid shape component in natural sculpin hybrids that can be best explained by transgressive segregation. This inference of how hybrids differ from their ancestors provides basic information for further evolutionary studies. Furthermore, our approach may serve to assign candidate specimens to their source populations based on morphometric data and help in the interpretation of population differentiation.

Introduction

Although hybridization has long been considered important in the diversification of plants zoologists often considered it detrimental and thus unimportant (Burke and Arnold 2001). The debate of the relative importance of hybridization has received recent attention because the advance of molecular techniques has resulted in a surge of data suggesting that hybridization is taking place rather frequently in animal kingdom as well. This in turn has revived questions surrounding the potential role that hybridization may play in the generation of evolutionary novelty in animals (Seehausen

2004, Schliewen and Klee 2004). A simple explanation for novel phenotypes of hybrids is available through the process of transgressive segregation. Briefly, transgressive segregation is a phenomenon specific to segregating hybrid generations and refers to individuals that exceed parental phenotypic values in any direction. This could be caused by heterosis, which is most pronounced in first generation hybrids, or alternatively by the complementary action of parental alleles dispersed among divergent parental lineages. If this is frequent, then an important evolutionary role for hybridization is more easily explained (Rieseberg *et al.* 1999).

In fact, there is abundant evidence that transgressive segregation is common in both plants and animals and that the genetic architecture for it is rather commonplace than exceptional (Rieseberg et al 2003). Given these findings, it is astonishing, that relatively few studies have evaluated transgressive segregation in natural systems (Rieseberg et al 1999). On the one hand this results from the paucity of study systems where sufficiently large samples are readily available and from the simple fact that quantitative genetics experiments are best conducted in controlled environments in order to separate environmental from genetic effects. If one searches for transgressive segregation one would ideally study traits, which are determined by several genes and that display a hidden divergence of the underlying genetic network (Rieseberg *et al.* 2003). Finally, one has to study direct hybrids and not lineages of hybrid origin because otherwise secondary evolutionary processes will have reshaped any hybrid lineage and secondarily modified transgressive traits. On the other hand many evolutionary studies will ultimately have to incorporate natural populations in real ecosytems if the effects and outcome of hybridization are to be analysed. As an example, hybrid zones among divergent lineages are viewed as natural laboratories and offer interesting study systems (Harrison 1993). In these, the fitness of hybrids is a key component to understand the dynamics of the hybrid zone as a whole (Barton 2001). Transgressive segregation may affect hybrid fitness as it is a mechanism that would make hybrids different and thus produces the raw material upon which selection can act.

We have recently identified hybrid zones of European sculpins belonging to the *Cottus gobio* complex (Scorpaeniformes, Cottidae) that fulfil the above requirements. Sculpins are small, benthic freshwater fish that occur in streams throughout Europe, with closely related species distributed throughout the northern hemisphere. Previous studies have revealed a high cryptic diversity of this group across the entire distribution range (Englbrecht et al 2000). Our focus area is the River Rhine System, where divergent lineages of sculpins are known to occur in parapatry and have come into secondary contact (Englbrecht *et al.* 2000, Volckaert *et al.* 2002). Small tributaries to the Lower Rhine drainage are inhabited by isolated populations of ‘stream’ sculpins, a lineage endemic to the River Rhine (Englbrecht *et al.* 2000, Chapter 1). Intriguingly, a new ‘invasive’ lineage, has recently appeared within the main channels of larger rivers that where previously free of *Cottus* (Chapter 1). The invasive sculpins represent a genetically distinct population that differs from stream populations in the Rhine area in that its lateral body is largely covered by modified scales vs. an almost complete absence of such

modified scales. Invasive sculpins come into secondary contact with local stream populations where small tributaries disembogue into the main channel of larger rivers. In these areas individuals belonging to both parental populations as well as hybrids among them occur syntopically. With respect to transgressive segregation the above prerequisites are fulfilled. First, sufficiently divergent lineages come into contact and produce recent hybrids. Secondly, these hybrids can be readily identified using genetic data. Finally, variation in body shape provides a well suited character complex as sophisticated methods are available for analyses of shape (Zelditch *et al.* 2004). Furthermore, previous studies on body shape show that this complex trait is generally determined by multiple genes (Klingenberg and Leamy 2002, Albertson *et al.* 2003, Moraes *et al.* 2004). Below we combine genetic and phenotypic approaches to study body shape in sculpin hybrid zones and present data suggesting that transgressive body shape phenotypes occur in natural sculpin hybrids.

In order to study variation in shape, parental groups and their hybrids were classified to establish how their phenotypes and genotypes were related. Since shape was of key interest, we relied on model based population genetic approaches (Falusch *et al.* 2003) to independently cluster and assign specimen to their populations of origin or to determine their hybrid status. Such model based clustering is not possible in the analysis of shape because a powerful “theory of population shape” comparable to population genetic theory is lacking that would allow to independently infer population affinity.

However, simple assignment methods can contribute much in the sense of the first assignment approaches in genetics that were employed for much more basic questions (Paetkau *et al.* 1995) namely the problem that distances alone are biologically and conceptually hard to interpret. As an alternative to an abstract distance one may ask whether a given character is sufficiently informative to be diagnostic at the individual, population or at higher levels to help in assessing the significance of results. This general problem also applies to quantitative morphometric studies, especially when multivariate analyses are used. We have developed a distance based assignment approach with statistical tests that parallel population genetic approaches (Cornuet *et al.* 1999). The method is not intended to give a measure of the absolute distance among groups but may help to interpret the differences among groups. One purpose of this paper is to introduce shape based assignment as a multivariate measure of distinctness and to employ this approach to study the relationships of genotypes and phenotypes at natural hybrid zones.

Methods

Implementation of shape based assignment

Landmark-based geometric morphometric methods were used to capture information about shape, by measuring the x and y coordinates of homologous landmarks in the configuration shown in Figure 1. Differences among specimen in the sets of coordinates due to scaling, rotation and

translation were removed using the typical geometric morphometric approach (Bookstein 1991, Rohlf and Marcus 1993, Dryden and Mardia 1998, Rohlf 1999, Zelditch *et al.* 2004) of placing the specimens in Partial Procrustes Superimposition (Dryden and Mardia 1998, Rohlf 1999, Slice 2001) on the iteratively estimated mean reference form, using the Generalized Procrustes Analysis procedure. This procedure places the shapes of specimens in a linear tangent space to Kendall's shape space (Kendall 1977), allowing the use of linear multivariate statistical methods (Rohlf and Marcus 1993, Bookstein 1996, Dryden and Mardia 1998, Zelditch *et al.* 2004).

After superimposition, the data were converted from cartesian coordinate form into components along the eigenvectors of the bending energy matrix (Principal Warp Axes) of the thin-plate spline model of deformations of the reference (Bookstein 1989, 1991] and along the uniform axes of deformation due to shear and dilation (Bookstein 1996b). Use of these linearly transformed variables (referred to as Partial Warp plus Uniform Component scores), produces a convenient set of variables (using a basis set called the Principal Warp axes) for use with standard multivariate statistical methods, since the Partial Warp and Uniform component scores have the same number of variables per specimen as degrees of freedom. No information is lost during this linear transformation of variables.

A canonical variates analysis (CVA) is then used to determine the set of axes which best discrimate among pre-defined groups of specimen, by determining the linear combinations of the original variables which display the greatest variance between groups relative to the variance within the groups (Krzanowski 1988, Zelditch at al. 2004). A simple Mahalanobis distance-based approach is then used to determine which group each specimen belongs to, based on the canonical variate scores. The predicted group membership of each specimen based on the CVA scores is determined by assigning each specimen to the group whose mean is closest (under the Mahalanobis distance) to the specimen.

To obtain a measure of the quality of the assignment of each specimen to a group, an assignment test was developed. The CVA axes can always be used to assign any given specimen to some group, since a minimum distance can always be found but a measure of whether the quality of the assignment is similar to that expected for specimens known to be in that particular group is desirable. The assignment test presented here is modeled on the genetic distance-based assignment test (Cornuet *et al.* 1999). The distribution of distances produced by a Monte Carlo simulation (see discussion in Manly 1997) is used to determine if the observed distance of a given specimen is consistent with the null model of random variation around the mean of the group to which the specimen is assigned to. The distance from a specimen to a group mean can then be assigned a p-score which describes how likely the specimen is to be a member of a group (under the null model used in the Monte-Carlo simulation). If the p score is smaller than 5%, then we can assert that there is a less than 5% chance that random variation could have produce a distance as large as that of the particular specimen from the group mean, and hence that the assignment of that specimen to the group is in doubt. It should be noted that in a study with many specimens, a number of them will have low p-scores by chance (the Bonferroni problem), and so to assess the validity of the assignments of the set as a whole, the researcher should

assess the number of specimens expected to have p values less than 5%. It will then be possible to determine if the observed number of low p values exceeds that expected by chance.

The model used in the Monte Carlo simulation of the distribution of Mahalanobis distances of specimens within a group around the group mean (the average specimen within the group) is based on a normal model of the distribution of the CVA scores of each group about the mean of that group. For a given group, it appears probable that the CVA scores along each CVA axes for the specimens within the group are correlated, thus there exists within each group a covariance structure to the CVA scores of specimens within the group. In carrying out the Monte-Carlo simulation of the distribution of specimens within the group about the mean, it is necessary to preserve this covariance structure to produce a valid model of the distribution. An eigenvalue decomposition of the variance-covariance matrix of within group CVA scores is used to find the principal component axes of the within group variance. This yields the same number of variables as the CVA scores, but now with uncorrelated axes (the eigenvectors), each of which has a variance given by the corresponding eigenvalue. The model used for the distribution of the CVA scores of the specimens assumes the group has an independent random normal distribution along each of these eigenvectors (principle component of variation axes), with amplitudes given by the square roots of the eigenvalues (the eigenvalues are the variances of the group along the corresponding eigenvectors), so that the square root of the eigenvalue is the standard deviation of the population along that eigenvector.

An independent, normal distribution with a known amplitude (the square root of the eigenvalue) is assumed along each eigenvector. This allows generation of a Monte Carlo population of specimens, assuming the independent normal distribution along these principal component axes. Each simulated specimen is generated using a random number generator to compute locations along the eigenvectors, which are then translated back into CVA axes scores. The Monte Carlo generated CVA axis scores will have the same mean and variance-covariance structure as the original population did. The Mahalanobis distance from the group mean is calculated for each simulated specimen. The Monte Carlo distribution of Mahalanobis distances of specimens about the group mean under the null model of random variation can then be used as an estimate of the distribution of distances about the group mean in the original data.

Based on the estimated distributions of Mahalanobis distances about the group means expected for specimens in the group, p-values may be determined for the assignment of all specimens, with either initially known or unknown group affinities. Based on an alpha level of p=0.05, all assignments of specimens to groups can be scored as either statistically significant or not.

As a test of the performance of the assignment test, a cross validation or jack-knife procedure (Manly 1997, Efron and Tibishirani 1993) procedure was implemented. Under the jackknife, a variable percentage of individuals from a dataset are left out during the CVA procedure, and then assigned to groups as “unknown” specimens. The specimens treated as unknowns during the jack-knife procedure are also assigned an assignment p-value during this procedure. High success rates under the jackknife resampling indicates that the differentiation among the involved groups is sufficient to be diagnostic.

This implies that the discriminant axes capture enough information to assign individuals of the given groups, and form a reasonable estimate of the distribution of distances based on the Monte Carlo procedure. The jack-knife procedure also allows estimation of the number of individuals needed to obtain meaningful CVA axes, and distance distribution estimates.

The sculpin data set

Here we employ the methods outlined above to study the differentiation in shape that occurs among divergent populations and their naturally occurring hybrids. Hybrid status is independently derived from genetic markers. Note that the specimen are taken from natural populations and occur syntopically in the hybrid zones studied below (Tab. 1). Sculpins were sampled across an area of secondary contact of invasive and stream sculpins, which is situated at the confluence of the Stream Broel with the River Sieg (Tab. 1). An extra population of stream sculpins was sampled from the stream Naaf (also tributary to the River Sieg drainage).

All specimen were genotyped for 45 microsatellite loci (Chapter 4, Nolte *et al.* 2005) The loci were chosen for their particularly high information content for our study system following the approach of Shriver *et al.* (1997) using WHICLOCI (http://www.bml.ucdavis.edu/whichloci.htm). We have used a preliminary genetic map of *Cottus* (Stemshorn *et al.* 2005), to verify that our set of microsatellite markers does not contain pairs of loci that are tightly physically linked. The genotypic data allow to unambiguously classify individuals to belong to pure populations or to identify them as hybrids with a mixed ancestry using the methods outlined in Falush *et al.* (2003). The program STRUCTURE2.1 (Falusch *et al.* 2003) yielded consistent results in independent runs (burnin: 20000; sampling iterations 100000, correlated allele frequency model allowing for an individual alpha and different F_{ST} for each subpopulation) according to which the genetic ancestry of individual could be determined (as of Chapter 4; see Chapter 5 suppl. Tab. 1 for genotypic data of those specimen included here). The classification based on genotypic data was highly congruent with data from distribution and morphology. Two independent populations of stream sculpins confined to separate streams (Stream Naaf, Stream Broel) both being devoid of skin prickling were recovered. A third population can be identified, which represents the invasive sculpin. Invasive sculpins generally occur within the main channel of the River Sieg and display pronounced skin prickling. Hybrids among Stream Broel sculpins with the invasive sculpins were only found at the confluence where the Broel merges with the River Sieg (Tab. 1). A detailed study of these hybrid zones, particularly on the geographic extension, is currently in preparation (Chapter 4). Of particular relevance in this context is the fact that hybrid sculpins occur syntopically with their parental populations within the hybrid zones (Tab. 1; Sites 2,3,4). For the morphometric analysis we grouped specimens identified as pure populations from the genotypic data into those corresponding to the invasive sculpins (invasive) and to the two stream sculpin populations (Streams Naaf and Broel). To allow for some uncertainty in the estimates we used those specimens that were found to be at least 97% pure in the STRUCTURE analysis. These populations are represented here by 117 Stream Broel sculpins, 76 Stream Naaf sculpins and 40 Invasive sculpins

(Tab. 1). In contrast, hybrids represent a somewhat inhomogeneous group consisting of various degrees of ancestry. To restrict this analysis to those specimens that have a pronounced hybrid genotype and to exclude later generation backcrosses that might have been subject to repeated rounds of selection we decided to exclude hybrids with less than 25% ancestry in one of the ancestral populations. Based on genotypic data, we were able to identify 62 BI-hybrids (mixed Stream Broel/Invasive ancestries, less than 75% pure ancestry) (Chapter 5 – suppl. Tab. 1).

Table 1: Sampling sites and number of specimens in the morphometric study. Sampling Sites, total number of genotyped specimens and numbers in genotypic classes used for this study. Specimens excluded from the analysis include later generation backcrosses or those for which morphometric data could not be obtained. Note that there are sampling sites at which all genotypic classes occur syntopically.

Number	Sampling Site	N (Genotyped)	N (Genotypic Classes)
1	Stream Broel between Broel and Winterscheidt, North Rhine-Westphalia, Germany; 50°47'N 7°20'E	48	48 Stream Broel sculpins
2	Stream Broel south of Broel, North Rhine-Westphalia, Germany; 50°47'N 7°19'E	48	42 Stream Broel sculpins; 1 BI Hybrid
3	Stream Broel at Mueschmuehle, 200 m above outlet into River Sieg, North Rhine-Westphalia, Germany; 50°47'N 7°18'E	130	26 Stream Broel Sculpins; 45 BI Hybrids; 13 Invasive Sculpins
4	River Sieg at Allner, below outlet of Stream Bröl, North Rhine-Westphalia, Germany; 50°46'N 7°18'E	36	2 Stream Broel Sculpins; 16 BI Hybrids; 2 Invasive Sculpin
5	Stream Wahnbach, Outlet into River Sieg at Seligenthal, North Rhine-Westphalia, Germany; 50°47'N 7°16'E	4	1 Invasive sculpin
6	Stream Pleis, outlet into the River Sieg at Niederpleis, North Rhine-Westphalia, Germany; 50°46'N 7°12'E	5	5 Invasive sculpin
7	River Sieg at Muehlenbach, North Rhine-Westphalia, Germany; 50°47'N 7°10'E	35	19 Invasive sculpin
8	Stream Naaf, Outlet into River Agger, North Rhine-Westphalia, Germany; 50°51'N 7°14'E	48	1 Stream Naaf sculpins;
9	Stream Naaf at Kreuznaaf, 200 m above outlet into River Agger, North Rhine-Westphalia, Germany; 50°51'N 7°14'E	48	30 Stream Naaf sculpins
10	Stream Naaf southeast of Hausdorp, North Rhine-Westphalia, Germany; 50°52'N 7°16'E	48	45 Stream Naaf sculpins

Images of specimen were taken with a digital camera fixed on a stage so that the midsagittal body plane was as much as possible perpendicular to the image plane. 14 morphological landmarks were used to capture the shape of each individual (Chapter 5 suppl. Tab. 2) using TPSdig (Rohlf 2003). The positions of the tip of the nasal (#1), nares (#2), interorbital pores (#3), dorsal fin I origin (#4), dorsal fin II origin (#5), dorsal fin II end (#6), upper caudal fin

origin (#7), lower caudal fin origin (#8), anal fin end (#9), anal fin origin (#10), ventral fin origin (#11), upper origin of the gill opening (#12), opercular spine (#13) and posterior end of the maxilla (#14) were used as landmarks (Fig. 1). The morphometric analysis was conducted using the IMP package according to the methods outlined above (Sheets 2002). Shape based assignment tests were conducted with CVAgen6N (part of IMP). In order to estimate possible confounding effects of allometric growth and sexual dimorphism these variables were determined individually. A scale bar was photographed besides each specimen as a size reference and sex was determined for individuals larger than 45mm by examination of the genital papilla (Chapter 5 - suppl. Tab. 2).

Results

Morphometric differentiation

The invasive sculpins and the two populations of stream sculpins form distinct clusters in a canonical variates analysis (CVA) along the first two axes, which display the greatest separation of the groups relative to within group variance. Both populations of stream sculpins separate from the invasive sculpins along CVAaxis1 (Axis 1 Lambda= 0.0678 chisq=777.6114 df=72 $p<0.001$). The Naaf and the Broel population are separated along axis2 (Axis 2 Lambda= 0.2792 chisq=368.6997 df=46 $p<0.001$). BI-hybrids overlap with their ancestral populations and take somewhat intermediate positions along CVAaxis1 (Fig. 2). However, if used as a predetermined group in the CVA, hybrids are further characterised by CVAaxis3 (Axis 3 Lambda= 0.7369 chisq=88.2487 df=22 $p<0.001$). They separate partially from invasive sculpins and stream sculpins along CVAaxis3 (Fig. 3) and take, on average, more extreme phenotypic values than both parental populations along this axis. The differentiation in shape as captured by CVA axes can be visualized as displacement vector for each landmark on a deformation grid relative to a reference (Fig. 1). Invasive sculpins differ from both populations of stream sculpin in that they have a larger head and anterior trunk as well as a shorter tail (Fig. 1; CVAaxis1). The two populations of stream sculpins differ most in their head length and the positions of their anal and dorsal fin landmarks (Fig. 1; CVAaxis2). While the deformation implied by the first two axes can be expressed in terms of inflation or compression of body parts, the hybrid specific shape change appears to be less balanced although this is hard to objectify (Fig. 1; CVAaxis3).

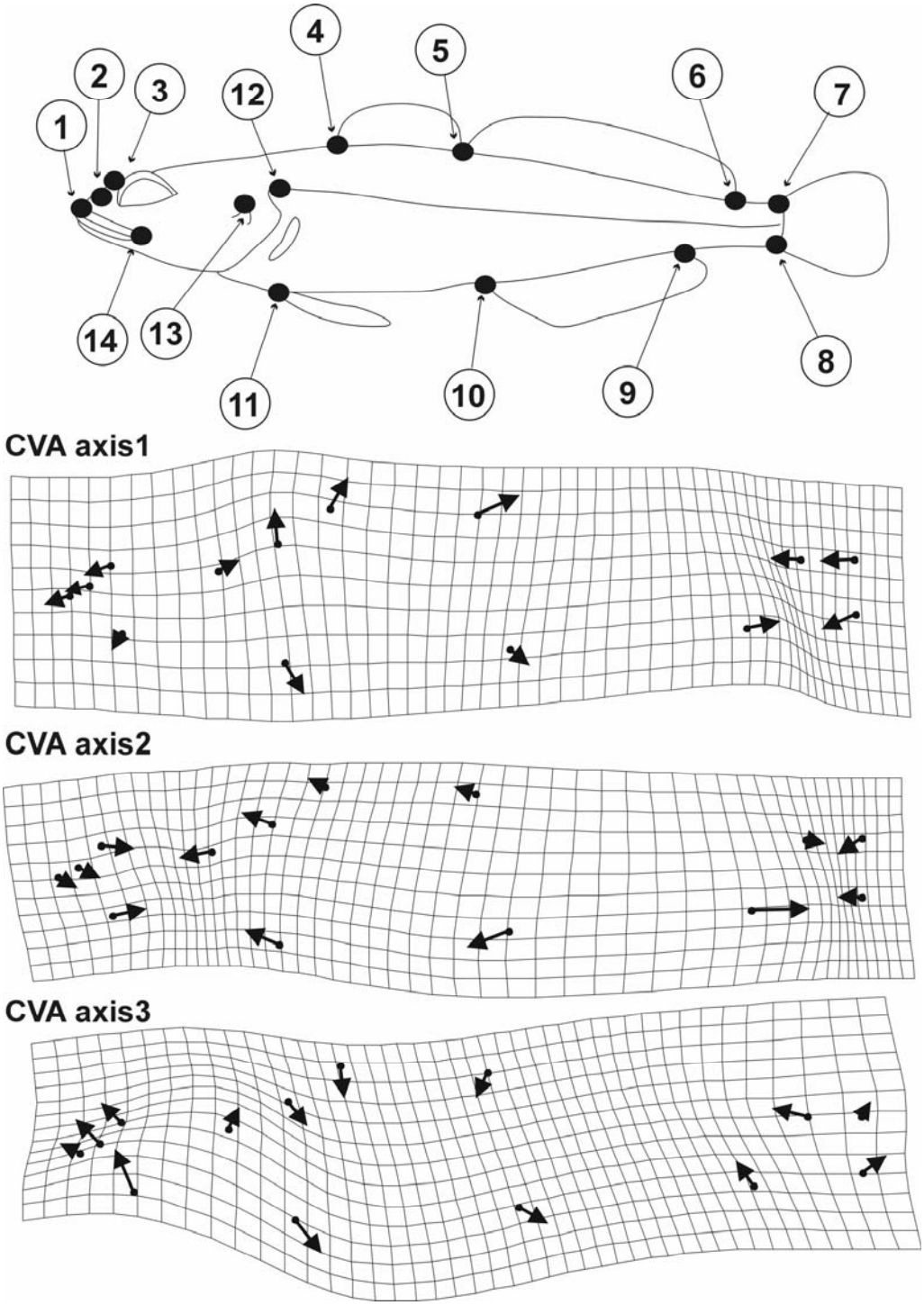


Figure 1: Landmark configuration and displacement vectors that distinguish groups of sculpins.14 Landmarks were chosen to analyse variability in sculpin body shape (top). CVA was used to identify axes along which different groups can be discriminated based on the relative position of landmarks to a reference. The shape change captured by these axes can be visualized as relative displacement vectors for each landmark on a deformation grid. Axis 1 separates invasive sculpins from all stream sculpins and axis 2 further separates two populations of stream sculpins. Axis 3 captures the shape component that is unique to recent hybrids. While the deformation along axes 1 and 2 can be expressed in terms of inflation or compression of body parts, the hybrid specific shape change appears to be less balanced.

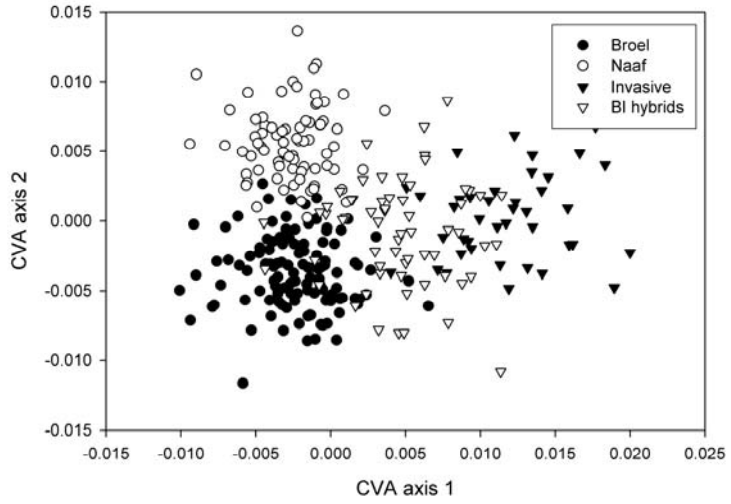


Figure 2: Differentiation of ancestral populations and hybrid intermediacy. Invasive sculpins separate from all stream sculpins along CVAaxis1. Sculpin populations from Stream Broel and Stream Naaf separate along CVAaxis2. BI hybrids form an intermediate group between their parental populations. Distance based assignment based on these two axes correctly identifies pure candidates while a majority of BI hybrids are wrongly assigned to one of the parental groups with which they overlap.

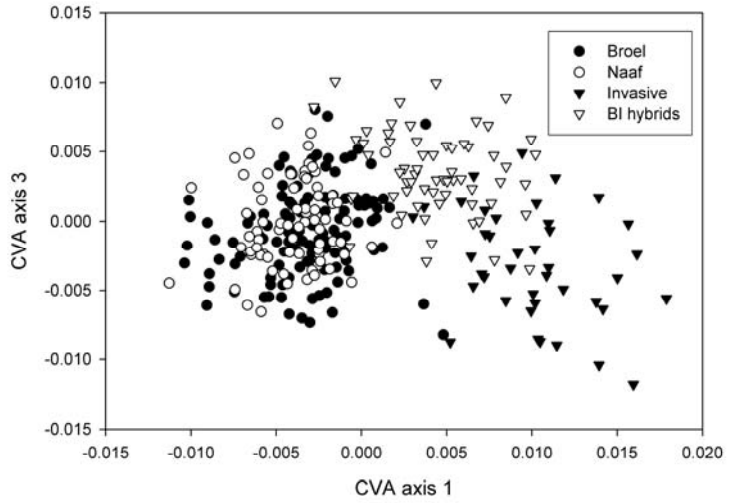


Figure 3: Extreme phenotypic values indicate a hybrid shape component. BI Hybrids are, on average, not intermediate along CVAaxis3 and may occupy extreme values relative to their parental populations. The parental populations as well as stream Naaf sculpins display little differentiation along CVAaxis3. An inclusion of this hybrid specific shape component in distance based assignment increases the power to correctly identify hybrids more than two fold.

In order to evaluate whether the observed differentiation was biased due to imbalanced sampling CVA scores were regressed on centroid size and sex for all specimens. A linear regression reveals for all axes that correlations coefficients were low at most (CVAaxis1 vs. size r2=0.03; CVAaxis1 vs. sex r2< 0.01; CVAaxis2 vs. size r2<0.01; CVAaxis2 vs. sex r2<0.01; CVAaxis3 vs. size r2=0.11; CVAaxis3 vs. sex r2=0.019). Therefore, neither size nor gender can explain a considerable amount of variance of the CVA axes that distinguish the groups.

Assignment and cross validation

To evaluate the utility of the derived axes to discriminate among groups and to determine a given specimens group affinity, distance based assignment tests were performed. In a first approach the model was based on the differentiation as observed among the pure populations but hybrids were not included. The clear differentiation of pure populations facilitates that single “unknown” specimen removed from the complete dataset can be correctly assigned to their population of origin with high success (Tab. 2). Approx. 93 % of pure sculpins were assigned correctly. This number is slightly lower than the expected 95% due to false positive assignments because of an overlap of parental phenotypic values. The number of outliers corresponds well to amount expected from the significance criterion. In this approach hybrids were used as “unknowns” and could only be identified as outliers relative to the pure sculpins (non significant assignment test). Only 34.8% of the BI hybrids were correctly classified while the majority was misassigned to one of the pure populations.

Table 2: Assignment success under alternative CVA models. Assignment of sculpins to their population of origin based on body shape. A model based only on the differentiation of parental populations is very effective in identifying pure sculpins but assigns the majority of hybrids to pure populations as false positives. A more complex model that includes the shape components specific to hybrids correctly identifies the majority of all hybrids. The overall success of parental group assignment is decreased when hybrids are taken into account as they overlap with parental phenotypic values.

Assigned Group	Broel	Naaf	Invasive	BI Hybrids	CVA model
Broel	**90.6**	1.3	2.5	25.8	Based on parental populations
Naaf	4.3	**93.4**	0.0	11.3	
Invasive	0.9	0.0	**92.5**	25.8	
n.s.	4.3	5.3	5.0	**37.1**	
Broel	**87.2**	5.3	0.0	1.6	Including parental groups and hybrids
Naaf	2.6	**89.5**	0.0	1.6	
Invasive	0.9	0.0	**85.4**	6.5	
BI hybrid	6.8	1.3	12.2	**83.9**	
n.s.	2.6	3.9	0.0	6.5	

In an alternative approach assignment was based on a model that includes the differentiation among pure populations but also takes into account the shape component specific to hybrids as captured by CVAaxis3 (Fig. 3). The assignment success of pure populations was decreased to 82.9-89.4% because of the partial overlap with the group of hybrids. In sharp contrast to the above, 84.1% of the BI hybrids were classified correctly with only relatively few false positive assignments to the parental groups. None of the BI hybrids were assigned to the stream Naaf population (Tab. 2).

A jack knife test of assignment was performed for both assignment approaches to evaluate the robustness of the CVA axes and assignment model (Tab. 3). The cross validation procedure revealed a very consistent signal, inherent to even small partitions of the whole dataset. Roughly half of the specimens can be removed from the data without much loss of information for the CVA axes. Even when 80% of the whole dataset are left out in the CVA procedure the general outcome remains unchanged although the number of correct assignments decreases. As evident from the individual assignment tests (Tab. 2) the overall assignment success is lower when the more complex model including hybrids is employed.

Table 3: Jack – knife estimates of assignment performance. Jack – knife test of assignment. Percent correct and false assignments when fractions of 1% to 80% of the specimens are left out in the CVA procedure and then assigned to groups in the remaining dataset. Large fractions of the data can be removed without loss of the discriminatory power of CVA axes.

% left out in Jack – knife (500 replicates)	1	10	50	80	CVA model
% correct	**93.6**	**92,7**	**90.3**	**72.1**	Based on parental groups
% correct ns.	0.5	0.5	1.3	10.5	
% false	5.9	6.8	8.3	14.8	
% false ns.	0.0	0.0	0.1	2.6	
% correct	**84.7**	**84.3**	**80.4**	**65.3**	Including parental groups and hybrids
% correct ns.	0.1	0.0	0.4	4.2	
% false	15.3	15.6	19.2	28.9	
% false ns.	0.0	0.0	0.1	1.6	

Discussion

Transgressive phenotypes in natural hybrids

The CVA recovers significant differentiation that separates sculpin populations as verified by independent genetic data. The differentiation of invasive sculpins and the two populations of stream sculpins is furthermore in line with their distribution and divergence in skin prickling patterns (Chapter 1). Cross validation confirms that these results are based on a signal inherent to the whole dataset as a removal of a large fraction of the specimens will not notably alter the axes as determined in the CVA (Tab. 3). This differentiation

is sufficient to assign unknowns to either one of the known groups with high confidence. Thus the groups are distinctly different in their multivariate signal even though no single diagnostic morphometric character can be found. In contrast to the ancestral populations, hybrids cannot be distinguished completely from all of the ancestral phenotypes (Fig. 2) and are more or less intermediate in the characters that discriminate their parental populations. This is expected for a character like body shape that is most likely determined by multiple genes. Yet, there are properties of the hybrids that could not be attributed to hybrid intermediacy. Hybrids display a unique differentiation that distinguishes them from their parental populations (Fig. 3). Altogether it is not a strong effect thus additional evidence to evaluate the biological significance of this result are desirable. To address sampling artefacts, we have tested for possible effects representing confounding variables in morphometric studies. Regressions show that the amount of total variance of individual CVAaxis scores that can be explained by size or sex is small. Therefore the influence of allometry or sexual dimorphism is most likely not important for the differentiation we observe. Despite the large overlap with the parental populations, the hybrid shape constitutes a considerable amount of variation, which results in an increased overall assignment success when hybrid shape is considered specifically (Tab. 2). Moreover, cross validation has shown that all axes are robust to removal of specimens, which suggests that the signal is inherent to all hybrids (Tab. 3).

Two alternative explanations remain. One assumes an involvement of genetic factors that interact to produce novel phenotypes, in contrast, the second proposes that the genetic background is not important. According to the latter hypothesis extreme hybrid phenotypes should be determined by the environment. Our genetic data demonstrate that hybrids occur syntopically with the parental populations within the hybrid zones (Tab. 1). This excludes the possibility that hybrids would be exclusively subjected to environmental factors that could induce the observed phenotypes. Phenotypic plasticity cannot be fully excluded in heterogeneous environments but this process alone is not likely to explain our results. After possible confounding variables were found to play a minor role, it seems reasonable to assume the differentiation is real. In contrast to the above explanations differentiation due to the underlying genetic background is strongly supported. This includes, that a specific shape effect can be identified that is linked to a hybrid genetic background. These findings can be easily explained by transgressive segregation, which is expected to be common and has already been described for morphometric traits (Rieseberg *et al.* 1999, Rieseberg *et al.* 2003). The hybrid shape component recovered by the CVA appears to be somewhat distorted and imbalanced relative to the differentiation among the pure populations (Fig. 1). This could be interpreted as a more or less random appearance of different transgressive effects in different parts of the body as the hybrid phenotypes we observe cannot be expected to represent perfectly functional phenotypes as obtained at the end of an evolutionary process. This relates directly to evolutionary questions surrounding hybridization, still, functional studies and measurements of fitness would have to complement the mere observation of transgressive phenotypes in the future.

Information content of shape markers

A drawback as compared to population genetic model – based assignment is that our shape distance based method needs a priori defined groups as input. If such groups can be provided hypotheses regarding their differentiation and distinctness can be tested. For example, an attractive application of genotype based assignment procedures is to detect outliers that belong to source populations that were not sampled (Primmer *et al.* 2000, Cornuet *et al.* 1999). Unfortunately this is not straightforward in our implementation of phenotype based assignment. If a candidate does not belong to one of the expected groups, the exclusion of source populations is not predictable because assignment based on discriminant axes is conditioned on the specific case being studied. We find this for the hybrids among stream Broel and Invasive sculpins if they are used as unknowns and not as a separate group in the CVA. Hybrids take more or less intermediate phenotypic values but largely overlap with the parental groups (Fig. 2). Similar results were already obtained by Strauss (1986) in a study of phenotypic variation in hybridizing North American sculpins. However, we have a sufficiently large sample of verified hybrids that could be used as extra group in the CVA. Only this revealed significant differentiation along an extra axis that is specific to the hybrids. The differentiation specific to hybrids adds information to the group assignment. As a result the assignment success of pure and hybrid specimen was raised notably despite the tremendous overlap of the hybrids with both parental populations (Tab. 2).

The assignment procedure based on morphometric data as implemented here allows to unambiguously assign sculpins to their population of origin. Morphometric differentiation of European sculpins was studied before Riffel and Schreiber (1998) using a set of landmarks that was largely identical to the ones used here (note that these authors did not study the same evolutionary lineages, see Englbrecht *et al.* 2000). Groups of sculpins as defined by different tributaries to the Rhine were found to differ significantly in shape but formed largely overlapping clusters. Our system differs in that we have not compared assemblages of populations but separate more or less panmictic populations as defined by a currently shared genepool. These form distinct clusters in the CVA (Fig. 2). Such differentiation would have escaped the approach of Riffel and Schreiber (1998) as they pooled specimens from different subpopulations for their analysis. This by no means negates their results but demonstrates an even higher information content of shape data, namely the power to discriminate separate populations. Although a comparison among genetic markers and body shape seems arbitrary the resolution as compared population genetic markers goes beyond recognition of ancient lineages as resolved by mitochondrial haplotypes (Englbrecht *et al.* 2000) and parallels that of microsatellites in that genetically well separated populations are also distinctly differentiated in body shape. Apparently, shape represents a character with a fast evolutionary divergence that occurs and becomes fixed even among closely related populations. Thus, in our example morphometric data resolve to the lowest possible level above the individual.

Acknowledgements

Thanks to D. Tautz for critical comments on the manuscript. A.N. thanks D. Tautz for his support of the *Cottus* project. The *Cottus* project was funded by the DFG (Ta99/20). We have obtained permissions to sample and take specimens from A. Mellin, T. Heilbronner and W. Fettweis, whom we thank for their benevolent support.

Literature

Albertson, R. C. Streelman, J. T. Kocher T. D. (2003) Genetic Basis of Adaptive Shape Differences in the Cichlid Head. Journal of Heredity, 94 (4):291-301.

Barton, N. H. (2001) The role of hybridization in evolution. Mol. Ecol. 10, 551-568.

Bookstein, F. L. (1989) Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 567–585.

Bookstein, F. L. (1991) Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press.

Bookstein, F. L. (1996a) Combining the tools of geometric morphometrics. In Advances in Morphometrics (L. F. Marcus, M. Corti, A. Loy *et al.* eds) pp. 131–152. Plenum Press.

Bookstein, F. L. (1996b) Standard formula for the uniform shape component in landmark data. InAdvances in Morphometrics (L. F. Marcus, M. Corti, A. Loy *et al.* eds) pp. 153–168. Plenum Press.

Burke, J. M. Arnold, M. L. (2001) Genetics and the fitness of hybrids. Annu. Rev. Genet. 35, 31–52

Cornuet, J-M. Piry, S. Luikart, G. Estoup, A. Solignac, M. (1999) New Methods Employing Multilocus Genotype to Select or Exclude Populations as Origins of Individuals. Genetics, 153:1989-2000.

Dryden, I. L. Mardia. K. V. (1998) Statistical shape analysis. John Wiley & Sons, New York.

Englbrecht, C. C. Freyhof, J. Nolte, A. Rassmann, K. Schliewen, U. Tautz, D. (2000) Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-pleistocene origin of the major central European populations. Mol. Ecol. 9, 709-722.

Efron, B. and Tibshirani, R. J. (1993) An Introduction to the Bootstrap. Chapman & Hall.

Falush, D. Stephens, M and Pritchard, J. K. (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics, 164: 1567–1587.

Harrison, R. G. (1993) Hybrids and hybrid zones. In: Hybrid zones and the evolutionary process, edited by R. G. Harrison, Oxford University Press, New York.

Kendall, D. G. (1997) The diffusion of shape. Advances in Applied Probability, 9:428-430.

Klingenberg, C. P. Leamy, L. J. (2002) Quantitative genetics of geometric shape in the mouse mandible. Evolution, 55 (11): 2342-2352.

Krzanowski, W. J. (1988) Principles of Multivariate Analysis. Oxford University Press.

Manly, B. F. (1997) Randomization, Bootstrap and Monte Carlo Methods in Biology. Chapman & Hall.

Moraes, E. M. Manfrin, M. H. Laus, A.C. Rosada, R. S. Bomfin, S. C. Sene, F. M. (2004) Wing shape heritability and morphological divergence of the sibling species Drosophila mercatorum and Drosophila paranaensis. Heredity, 92, 466–473.

Nolte, A. W. Stemshorn, K. C. Tautz, D. (2005) Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Mol. Ecol. Notes (*in press*).

Paetkau, D. Calvert, W. Stirling, I. Strobeck, C. (1995) Microsatellite analysis of population structure in Canadian polar bears. Mol. Ecol. 4, 347–354.

Primmer, C. R. Koskinen, M. T. Piironen, J. (2000) The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. Proceedings of the Royal Society of London Series B -Biological Sciences 267 (1453): 1699-1704.

Rieseberg, L. H. Archer, M. A. & Wayne, R. K. (1999) Transgressive segregation, adaptation, and speciation. Heredity, 83, 363–372.

Rieseberg, L. H. Widmer, A. Arntz, A. M. Burke, J. M. (2003) The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. Phil. Trans. R. Soc. Lond. B, 358, 1141–1147

Riffel, M. Schreiber, A. (1998) Morphometric differentiation of sculpin (*Cottus gobio*), a fish with deeply divergent genetic lineages. Can. J. Zool. 76, 876-885.

Rohlf, F. J. Marcus, L. F. (1993) A Revolution in Morphometrics. Trends in Ecology and Evolution, 8:129-132.

Rohlf, F. J. (1999) Shape Statistics: Procrustes Superimpositions and Tangent Spaces. Journal of Classification, 16:197-223.

Rohlf, F. J. (2003) TpsDig, digitize landmarks and outlines, version 1.39. Department of Ecology and Evolution, State University of New York at Stony Brook.

Schliewen, U. K. Klee, B. (2004) Reticulate sympatric speciation in Cameroonian crater lake cichlids. Frontiers in Zoology, 1:5

Seehausen, O. (2004) Hybridization and adaptive radiation. Trends Ecol. Evol. 19 (4): 198-207

Sheets, H. D. (2002) IMP-Integrated Morphometrics Package. Deptartment of Physics, Canisius College, Buffalo, New York.

Shriver, M. D. Smith, M. W. Jin, L. Marcini, A. Akey, J. M. Deka, R. Ferrell, R. E. (1997) Ethnic-affiliation estimation by use of population-specific DNA markers. American Journal of Human Genetics, 1997, 60(4):957-964.

Slice, D. E. (2001) Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space. Systematic Biology, 50, 141–149.

Stemshorn, K. C. Nolte, A. W. Tautz, D. (2005) A Genetic Map of *Cottus gobio* (Pisces, Teleostei) based on microsatellites can be linked to the Physical Map of *Tetraodon nigroviridis*. J. Evol. Biol (*in press*)

Strauss, R. E. (1986) Natural hybrids of the freshwater sculpins *Cottus* bairdi and *Cottus* cognatus (Pisces: Cottidae): Electrophoretic and morphometric evidence. Am. Midl. Nat. 115, 87-105.

Volckaert, F. A. M. Hänfling, B. Hellemans, B. Carvalho, G. R. (2002) Timing of the population dynamics of bullhead *Cottus gobio* (Teleostei: Cottidae) during the Pleistocene. J. Evol. Biol. 15, 930-944.

Zelditch M. L. Swiderski D. L. Sheets H. D. Fink W. L. (2004) Geometric Morphometrics: A primer. Academic Press, London.

Appendix – Chapter 5

Chapter 5 - Supplementary Table 1: Inferred group affinity and individual genotypic data. Genotypes of all specimens for 45 microsatellite loci (0 = missing data, alleles numbered according to size, but not necessarily repeat size) with group affinity and sampling site as of Tab. 1. Data format: Table; saved row by row with fields separated by semicolons. Ends of rows are marked by the insertion of “XXX”.

Chapter 5 - Supplementary Table 2: Individual landmark data, centroid size and sex. Cartesian coordinates (X - Y format) for 14 landmarks, with individual group affinity and sampling site as of Tab. 1 as well as sex (0 = female; 1 = male) and centroid size. Data format: Table; saved row by row with fields separated by semicolons. Ends of rows are marked by the insertion of “XXX”.

Chapter 6: The distribution of genomic regions associated with habitat and divergent morphology across sculpin hybrid zones

Arne W. Nolte and Diethard Tautz

Abstract:

Hybridization among wild organisms receives attention as a natural experiment on selection and adaptation. The theory that explains the dynamics of hybridization rests to a great extent on how different traits are subject to selection. Unfortunately, most studies on hybridization are restricted to anonymous markers to date simply because of a lack of knowledge about the genetics of nonmodel organisms. In principle, QTL mapping can be used to identify specific genetic factors but this methodology is largely confined to model organisms. Natural hybrid zones represent a possible alternative because far-reaching associations between states of ancestry at physically linked loci are generated in groups of segregating hybrids. This ancestry association is exploited by admixture mapping, which is applied for the first time in a wild population of non-model organisms. 134 microsatellite markers covering a large fraction of the genome were typed in 480 specimens of sculpins (*Cottus gobio*) from natural hybrid zones. Besides first insight into the genetic architecture of divergent morphological characters, admixture mapping provides access to genomic regions that are associated with habitat use. Candidate regions are scattered throughout the genome but affect more than one trait in the majority of cases. This implies physical linkage of divergent morphotypes with fitness relevant traits. The findings shed light on the genetic architecture of divergence across hybrid zones and the history of past hybridization of the invasive sculpin.

Introduction:

Hybridization among wild organisms receives attention as a natural experiment on selection and adaptation. The theory that explains the dynamics of hybridization rests to a great extent on how different traits are subject to selection (Barton 2001). Unfortunately, most studies on natural hybridization are restricted to anonymous markers to date simply because of a lack of knowledge about trait genetics of the vast majority of animal and plant life. In principle, QTL mapping (MacKay 2001) can be used to identify the genetic factors that affect a phenotypic trait but to date this methodology is still largely confined to model organisms. As a result, quantitative genetic studies in natural populations are still rare and basically absent in truly wild

populations (Slate 2005). Three reasons are traditionally thought to prohibit QTL studies in non model organisms. These are the combined lack of genetic markers, a genetic map and of laboratory strains for controlled experiments. On the other hand, it is obvious that many fascinating biological phenomena are not observed in laboratory strains but in wild populations of largely unexplored species. We are currently using European sculpins of the *Cottus gobio* complex (Teleostei, Scorpaeniformes, Cottidae) to study the genetics of divergence, adaptation and hybridization. The first two issues were addressed with the development of a large number of microsatellite markers (Nolte *et al.* 2005) and construction of a preliminariy genetic map from F_1 crosses (Stemshorn *et al.* 2005). In this paper the last and key step, namely the localization of candidate genes, is approached using material from entirely wild populations.

Admixture mapping in natural hybrids

An approach to map QTL in the wild is based on pedigree data (Slate *et al.* 2002) with a focus on the identification of traits affecting phenotypes within populations. If one aims to dissect the genetics of traits that are associated with the divergence of evolutionary lineages, pedigree data are most likely not available. Rieseberg and Buerkle (2002) already noted that natural hybrid zones offer a way to identify QTL since hybrids are quite common in animals and plants and are particularly abundant in hybrid zones. These could principally be employed to locate QTL because hybridization creates linkage disequilibrium that is absent in nonadmixed populations (Stephens *et al.* 1994). Under admixture, gene flow produces chromosomal segments that have ancestry in different populations. These segments carry genes that should be identifiable based on the ancestry of linked neutral markers (McKeigue 2005). McKeigue (1998) and Shriver *et al.* (2003) have introduced “admixture mapping” to identify significant correlations of phenotypic traits and genetic markers. The key principle in their approach is to control for individual admixture in tests of association to remove spurious associations of phenotypes and ancestry that are introduced by population substructure.

Admixture mapping confers principles of classical QTL mapping in experimental crosses to the level of admixture among differentiated lineages. Candidate traits are required to be heritable and differ in their incidence among source populations and secondly, marker loci are needed that are informative of ancestry in alternative source populations. Within a group of segregating hybrids associations between states of ancestry at different loci are generated that are strongest in physically linked loci. These ancestry associations can be identified by modeling genotypic data (McKeigue 2005). The power and precision to identify candidate regions depends on the degree to which recombination has broken down the ancestry association in admixed individuals, i.e. time since admixture. Simulations (Briscoe *et al.* 1994) and empirical studies (Collins-Schramm *et al.* 2003) indicate that admixture linkage disequilibrium between loci is most pronounced between 5 to 12 generations after hybridization and extends over 5-20 centimorgans. Furthermore the degree to which both the candidate trait and the neutral genetic variability are diagnostic among source populations is of key importance as this determines the information content of the dataset. Thus,

admixture mapping offers an approach to identify candidate QTL in entirely wild populations but empirical studies have been restricted to humans to date.

Mapping traits across hybrid zones

European sculpins provide us with a study system to identify the genetic basis of ecological success, the genomic architecture of divergence and to learn more about the role of past hybridization (Chapter 1, Chapter 4). Briefly, the so called “Invasive” sculpin has recently colonized lowland habitats of the River Rhine that were free of sculpins before while other populations referred to as stream sculpins have remained confined to small headwater streams within the same drainage. Admixture mapping is applicable because the explosive range expansion of invasive sculpin resulted in the formation of hybrid zones with Rhine stream sculpins since approx. 10-15 years (Chapter 4). This translates into an equal number of generations and rounds of recombination at most. Secondly, the two lineages are interfertile so that a continuous range of ancestry is realized within narrow hybrid zones (Chapter 4). Sculpin hybrid zones are particularly interesting because they are strongly determined by environmental factors. They are situated at outlets of small streams, which also represent a pronounced ecological transition within continuous watercourses. The genetic structure observed at these transition zones can only be explained by a divergence in fitness traits in the two types of sculpin (Chapter 4). Besides this, invasive and steam sculpins are differentiated in life history and morphological traits but the interrelationships of all of these are not known.

In this paper, admixture mapping is used to determine the genomic distribution of genetic factors affecting morphology and habitat specificity. Quantitative genetic aspects of differentiation of divergent lineages of sculpins are integrated with the processes that mould hybrid zones. The results allow for first insights into the genetic architecture of divergence and suggest a genomic correlation of traits affecting the divergent morphology with factors that contribute to the unique ecological potential of the invasive sculpin.

Methods:

Sampling scheme

Invasive sculpins are represented by material from the River Sieg and a major tributary the River Agger. Pure stream sculpins were taken from two populations, one in the Stream Naaf and the other one from the Stream Broel. Hybrid zones are situated at the confluences of streams with the next larger rivers (Tab. 1, Fig. 1). Sampling was done with a portable electroshocker in 2003-2004. All specimens were deafened with CO_2 and subsequently fixed in 100% ethanol for 10 minutes, which straightens the body. To prevent excessive dehydration, the specimens were then transferred to 70% ethanol for storage. This procedure will result in massive distortion of specimens in many species but this effect is relatively weak in sculpins. As a result samples could be analysed morphologically and genetically as the sculpin tissues still contain DNA that is suitable for genotyping. The genetic structure, extension and ecological context of these hybrid zones are subject to a detailed study that is treated separately (Chapter 4). Here, only those results that are

relevant in the context of admixture mapping are reported. This includes the number of involved genetic clusters and estimates of individual admixture.

Genotyping and Inference of population structure

In addition to markers from Englbrecht *et al.* (1999), further microsatellites were identified of which 171 were typed (Chapter 6 – suppl. Tab.1) for all samples as described in Nolte *et al.* (2005). The first goal was to test whether the inference of one invasive population in secondary contact with two separate stream populations is reflected by genetic data. The dataset was reduced with the goal to maximize population genetic resolution and to minimize confounding effects introduced by physical linkage. Markers with the highest possible allele frequency differential among pure populations were determined following Shriver *et al.* (1997) using WHICHLOCI (http: //www.bml.ucdavis.edu/whichloci.htm). Furthermore, a preliminary genetic map of *Cottus* (Stemshorn *et al.* 2005) served to exclude markes that are tightly physically linked. This yielded a set of 45 microsatellite loci (as of Chapter 4; Chapter 6 – suppl. Tab. 2) with an average between population allele frequency differential of 88%, none of which was linked more closely than 18 cM (Chapter 4). STRUCTURE 2.1 (Falush *et al.* 2003) was used to infer the most likely population structure based only on genotypic data. Several chains (burnin: 20000; sampling iterations 100000, correlated allele frequency model allowing for an individual alpha and different F_{ST} for each subpopulation) were run for each value of K between 1 and 7. The log-likelihood of the data rises strongly up to a value of K=3. For K>3 there is a strong tendency to plateau (See Chapter 4). A value of K=3 represents the best estimate of the number of involved populations following the recommendations given in the STRUCTURE manual and in Falush *et al.* (2003) and because of the perfect fit with morphological, ecological and distribution data (Chapter 1, Chapter 4).

Characters

The goal of admixture mapping is to detect QTL carrying genomic regions that underlie the observed differentiation of a given trait. Ancestry association may occur both with a given population and with individual marker loci. The former would be informative about the incidence of a trait in a population while the latter would identify candidate regions carrying QTL. A prerequisite for admixture mapping is that candidate traits are heritable and differ in their incidence among the involved source populations. This was explored by plotting estimates of individual ancestry as obtained in the STRUCTURE analysis against individual phenotypic values.

The morphological traits studied here include skin prickling and the total number of vertebrae, which were previously found to be particularly divergent among invasive and stream sculpins (Chapter 6 – suppl. Tab. 3). Modified spinelike scales called skin prickling (size 1-2 mm) vary in the degree to which they cover the lateral sides of the sculpin body (Fig. 1). In order to quantify this trait, degrees of skin prickling were classified into five categories: 0 = prickles absent; 1 = less than ten prickles present beneath pectoral fin; 2 = more than ten prickles but all covered by the pectoral fin; 3 = prickling extends beyond pectoral fin but ends anterior to the middle of the second dorsal fin; 4 = prickling extends back beyond the middle of the second

dorsal fin (Fig. 1). The number of total vertebrae was counted for a subset of 324 specimens (Chapter 6 - suppl. Tab. 3). For this purpose, radiographs of individual specimens were taken. All vertebrae were counted from the first vertebra behind the cranium to the last vertebra not in contact with the hypural plates in the tail.

Table 1: Sampling sites with geographical coordinates, the number of sampled specimens and a rough classification into pure populations or hybrid zones (compare Fig. 1). All pure samples that represent the ancestral lineages are included in the admixture mapping analysis to model ancestral allele frequencies and character states.

Ref.	Site	GIS	N	Classification
1	Stream Broel between Broel and Winterscheidt, North Rhine-Westphalia, Germany	50°47'N 7°20'E	48	Stream sculpins
2	Stream Broel south of Broel, North Rhine-Westphalia, Germany	50°47'N 7°19'E	48	Stream sculpins
3	Stream Broel at Mueschmuehle, 200 m above outlet into River Sieg, North Rhine-Westphalia, Germany	50°47'N 7°18'E	130	Hybrid zone
4	River Sieg at Allner, below outlet of Stream Bröl, North Rhine-Westphalia, Germany	50°46'N 7°18'E	36	Hybrid zone
5	Stream Wahnbach, Outlet into River Sieg at Seligenthal, North Rhine-Westphalia, Germany	50°47'N 7°16'E	4	Invasive sculpin
6	Stream Pleis, outlet into the River Sieg at Niederpleis, North Rhine-Westphalia, Germany	50°46'N 7°12'E	5	Invasive sculpin
7	River Sieg at Muehlenbach, North Rhine-Westphalia, Germany	50°47'N 7°10'E	35	Invasive sculpin
8	River Agger at Donrath, 900 m below outlet of stream Naaf, North Rhine-Westphalia, Germany	50°51'N 7°13'E	30	Hybrid zone
9	Stream Naaf, Outlet into River Agger, North Rhine-Westphalia, Germany	50°51'N 7°14'E	48	Hybrid zone
10	Stream Naaf at Kreuznaaf, 200 m above outlet into River Agger, North Rhine-Westphalia, Germany	50°51'N 7°14'E	48	Hybrid zone
11	Stream Naaf southeast of Hausdorp, North Rhine-Westphalia, Germany	50°52'N 7°16'E	48	Stream sculpins

Fitness and adaptation are key characters in evolutionary inference, which by their nature, are much harder to grasp than the above. However, the specific properties of the sculpin hybrid zones can be exploited because two alternative habitats directly abut and some dispersal into alternative habitats occurs. Genetic data suggest for these habitat transitions that selection counteracts the establishment of immigrant genotypes in both rivers and streams (Chapter 4). Based on the fundamental concept of ecological genetics that the adaptive value of genotypes determines association with ecological factors (Arnold *et al.* 1999, Lowe *et al.* 2004) „habitat“ can serve as a surrogate to integrate adaptationally important genes. Specimen were considered to originate from river-habitats when they were sampled downstream of a confluence of a tributary stream (Stream Broel; Stream Naaf) with a larger river (River Sieg, River Agger) and from stream-habitats when they were sampled from a stream (Chapter 6 - suppl. Tab. 3). Habitat was not classified for a small set of five specimens collected at site 6 (Table 1) as this site corresponds more to a seperate lowland sandbottom type stream as opposed to all other streams (Broel, Naaf, Stream Wahn (6)) that are gravel bottom upland streams.

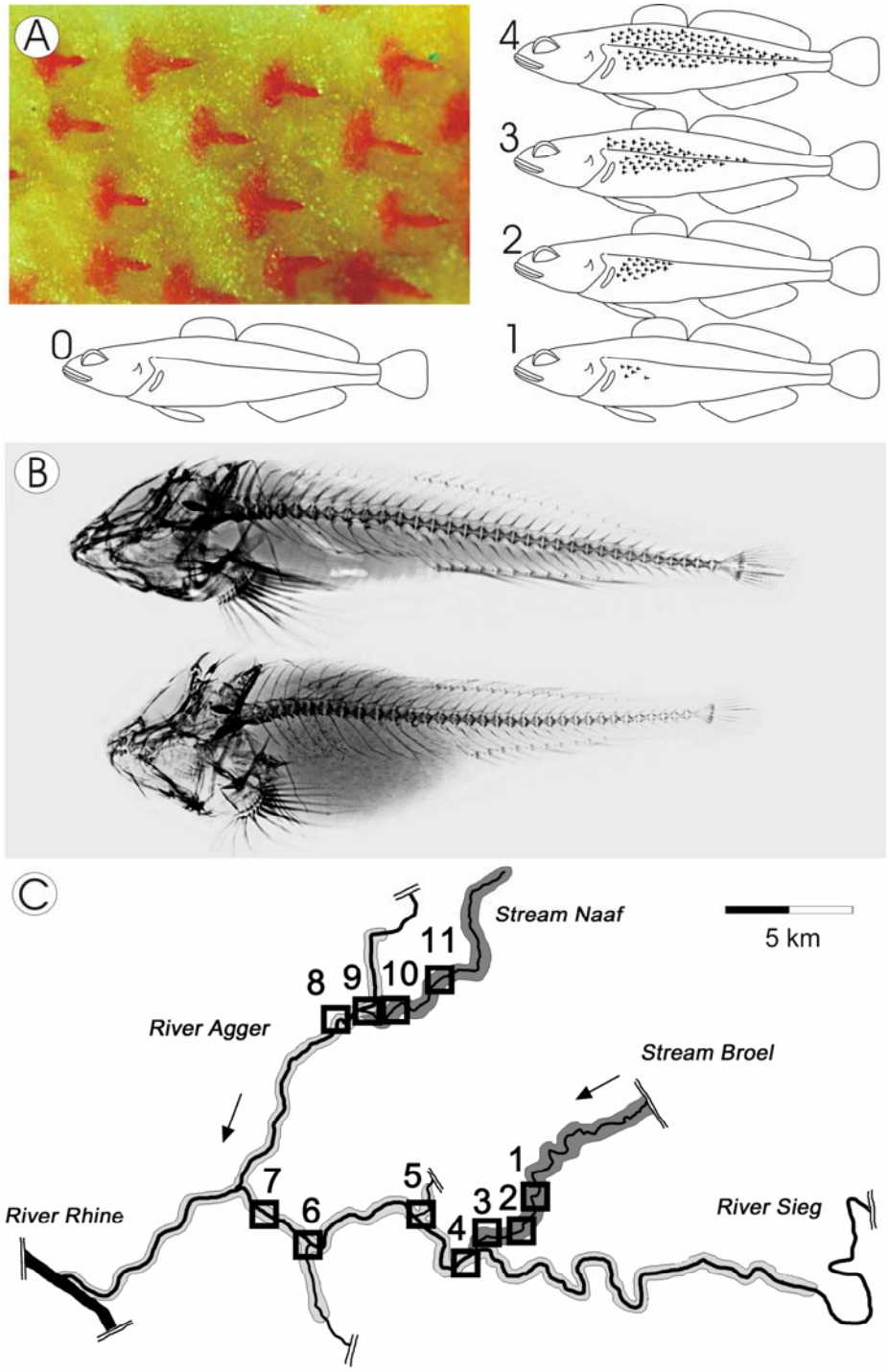


Figure 1: Characters examined for ancestry association and the geographical distribution of sculpin hybrid zones. A) Prickles (stained with alizarin red) are modified scales embedded in the skin. Sketches illustrate the different degrees (0-4) to which these were found to cover the sculpin body. B) The total number of vertebrae was counted from radiographs (upper: stream sculpin; lower: invasive sculpin) and ranges from 28-33. C) Stream sculpins (dark grey) are confined to small tributaries while the invasive sculpin occurs downstream in larger rivers (light grey) within the River Sieg drainage. Individuals of mixed ancestry occur in the proximity of stream outlets. These areas were sampled densely (numbered squares) to obtain recent hybrids exploited for admixture mapping.

Admixture mapping

134 microsatellite loci (Chapter 6 – supp. Tab. 2) were part of known linkage groups (Stemshorn *et al.* 2005) and were used here with the goal to identify candidate genomic regions for the traits described above using ADMIXMAP 1.6.0 (McKeigue *et al.* 2000). ADMIXMAP allows in principle to include linkage information to raise the amount of information that can be extracted from closely linked loci. However, here linkage information was not included because the preliminary genetic map of *Cottus* may include artefacts like falsely joined linkage groups (Stemshorn *et al.* 2005). Further, this has the advantage that statistically significant associations of physically linked markers would in fact reconfirm the validity of single locus results. Finally, fine mapping was not the goal in this screen. Larger genomic regions will be studied in the future using a refined approach and denser marker spacing that can exploit the full power of admixture mapping. All inference was based on a model assuming one population of invasive sculpin and two of stream sculpins. ADMIXMAP produces posterior estimates of ancestry specific allele frequencies for three populations based on genotypic data alone. To save computational time, we have used these estimates as prior values to model ancestry association of traits with the following settings: burnin 5000, iterations 20000, sampling every 10 iterations and prior allele frequencies for three populations. A significance threshold of $p<0.01$ was used. All calculations were based on the complete set of specimens. This was done to base estimates of ancestral allele frequencies on the largest possible sample size to account for the variability of microsatellite markers. ADMIXMAP permits to include possible effects of covariates on trait values into the regression model. Therefore, the values of traits described above (except for the trait under study) as well as individual sex and size (Chapter 6 – suppl. Tab. 3) were included as covariates to determine ancestry association. This takes into account previous findings that suggest: 1) that the distribution of genotypes is highly dependant on habitat (Chapter 4); 2) that the involved populations of sculpins display rather different demography which is accounted for by size (Chapter 1) and 3) the fact that allometry and gender may have an effect. For instance, prickling is most pronounced in juveniles and strongly decreased after first reproduction in males.

To determine, whether candidate regions associated with habitat were genetically correlated with those associated with divergent morphology nearest neighbour map distances were examined. For this purpose, the distance in cM from map positions associated with habitat to both the nearest possible positions associated with and not associated with a morphological trait was determined within linkage groups. A Wilcoxon test in SPSS 12.0.1 (SPSS inc.) was used to examine, if the map distance to a morphotype associated trait is on average lower than the distance to the next non-associated marker.

Results.

Plots of trait value distribution against ancestry estimates taken from STRUCTURE show that more than 60% of the total phenotypic variance is explained by ancestry (Fig. 2) for both morphological characters. Pure invasive sculpins display higher degrees of skin prickling and a lower number of vertebrae as compared to both populations of stream sculpins. Of particular relevance for this study is the fact that intermediate degrees of ancestry are on average coupled with intermediate phenotypic values, which documents phenotypic segregation in recent hybrids.

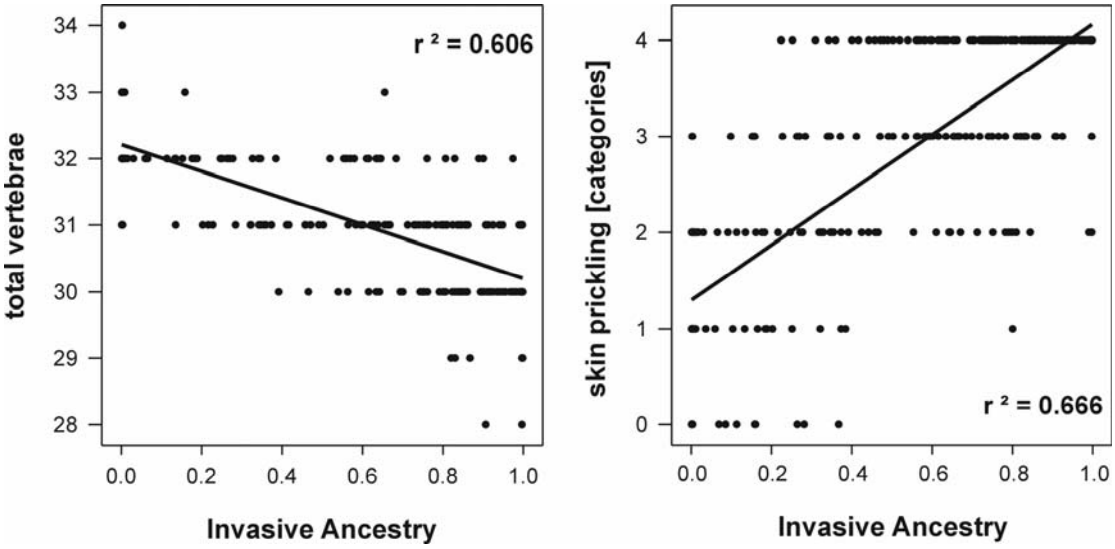


Figure 2: Association of individual phenotypic values with overall ancestry. Individual ancestry is expressed as proportion of genetic material derived from the invasive genepool (Invasive Ancestry). Regression of the number of vertebrae and skin prickling categories shows that a considerable amount of phenotypic variance (>60%) can be explained by ancestry. Extreme ancestries on the left and right of the x-axis represent non-admixed individuals, which are well differentiated in their phenotypes. The intermediate range of ancestries is on average associated with intermediate phenotypic values. These hybrids carry the information that is exploited by admixture mapping.

Tests for ancestry association among loci and characters yields 34 candidates for skin prickling, 21 for vertebrae, and 28 loci that are associated with habitat (Tab. 2). Candidate loci for one trait are often linked physically and thus confirm each other in defining candidate regions. 22 of all skin prickling loci are grouped into regions with a distance of less than 20 cM between significant markers. The same holds for 15 and 12 of the candidates for habitat and vertebrae respectively (Tab. 2). Independent candidate regions are distributed across several linkage groups of which 13 show association with skin prickling, 10 with the total number of vertebrae, and 12 with habitat.

Table 2: Mapped Microsatellite loci with a significant ancestry association with morphological and ecological characters. Several loci are associated with more than one character across sculpin hybrid zones. Note that physically linked loci tend to reveal the same signal. Groups of loci often represent continuous candidate regions based on linkage group and map position. The extracted information varies due to variation in ancestry information content of loci and depending on the trait (P=prickling, V=vertebrae, H=Habitat).

Locus	Traits	Linkage Group	Map Position	Percent Info (P/V/H)	Ancestry association p-value (P/V/H)
CottES10	H	1	0	35	2.20E-03
Cott164	V	2	0	72	3.10E-03
CottES2	V, H	2	0	36, 30	6.60E-03; 8.80E-06
CottES21	H	2	0	59	1.70E-03
LCB18	V, H	2	0	40, 20	3.40E-03; 1.80E-08
Cott194	P	3	261.12	46	9.60E-03
LCE42	P	3	366.54	55	2.70E-04
LCE43	P, V	3	768.79	32, 46	1.20E-07; 5.30E-03
LCE89	V, H	3	768.79	16, 29	1.40E-05; 3.10E-03
Cott255	V	3	825.53	10	3.70E-03
CottE9	P, H	3	825.53	47, 28	2.40E-03; 2.00E-04
CottES6	P	3	948.25	37	6.90E-03
CottE23	P, H	3	1019.18	45, 21	3.30E-08; 1.40E-05
LCE21	P	3	1019.18	39	9.10E-14
Cott146	P	3	1029.07	28	8.00E-03
Cgo33	P	3	1038.14	41	2.20E-06
Cott105	P, H	3	1038.14	54, 24	8.40E-05; 7.00E-04
Cott152	P, H	3	1038.14	20, 17	2.50E-04; 2.20E-03
LCE64	P	3	1045.87	34	8.30E-04
CottE30	V	3	1308.11	15	7.40E-04
LCE2	P	3	1315.47	30	5.20E-03
Cgo42	P, V, H	3	1531.92	40, 11, 12	4.10E-04; 7.90E-03
LCE30	P	3	1569.68	41	1.10E-04
Cott170	V, H	3	1646.81	13, 41	7.00E-09; 2.60E-03
Cott348	V	5	0	13	8.80E-03
LCE51	V	5	0	37	4.00E-03
Cott128	H	5	18.02	22	2.20E-03
CottE2	P, H	5	85.38	18, 12	4.10E-05; 9.70E-04
Cott582	H	5	98.12	41	2.80E-03
Cgo1033	P, V	6	0	28, 18	4.20E-04; 6.00E-03
LCE40	P	6	13.88	17	1.70E-03
LCE27	P, H	7	0	22, 32	3.40E-04; 3.70E-03
LCE35	P	7	0	45	2.30E-05
LCE13	H	8	0	45	2.50E-05
LCE37	H	8	49.76	21	2.30E-05
LCE44	P	8	69.86	29	1.40E-03
LCE69	V, H	8	69.86	14, 30	2.10E-03; 4.50E-03
LCE105	P, H	9	0	25, 32	2.80E-04; 9.40E-04
Cgo56	P	9	7.61	16	6.60E-03
LCE24	H	9	7.61	60	1.80E-04
Cott130	P	11	0	19	4.10E-04
LCE66	H	11	0	84	1.50E-03
LCE275	H	13	23.98	35	2.00E-03
Cott78	P	13	48.23	20	8.80E-04
Cott635	P, V, H	14	0	68, 22, 54	1.20E-03; 1.60E-05
LCE26	P, V, H	14	9.03	56, 19, 21	3.00E-03; 2.30E-05
LCB16	V	15	0	10	4.00E-10
Cott138	P	15	8.13	11	5.50E-03
CottE12	P	15	8.13	59	4.10E-04
Cott700	P	15	19.28	49	1.30E-03
LCE11	H	16	0	29	2.30E-04
Cott580	H	16	16.35	50	9.00E-06
CottE11	H	16	16.35	28	5.50E-04
LCE126	V	16	256.98	33	2.30E-05
Cgo22	V, H	17	0	34, 66	2.30E-05; 1.00E-04
Cott619	V	17	0	78	5.70E-03
CottE7	V	17	0	25	2.80E-03
Cott570	P	17	99.22	33	6.10E-03
Cott210	P	19	15.96	23	5.60E-04
Cott163	P, V	20	8.47	27	4.90E-03; 2.00E-03
Cott205	P	20	8.47	13	4.90E-05

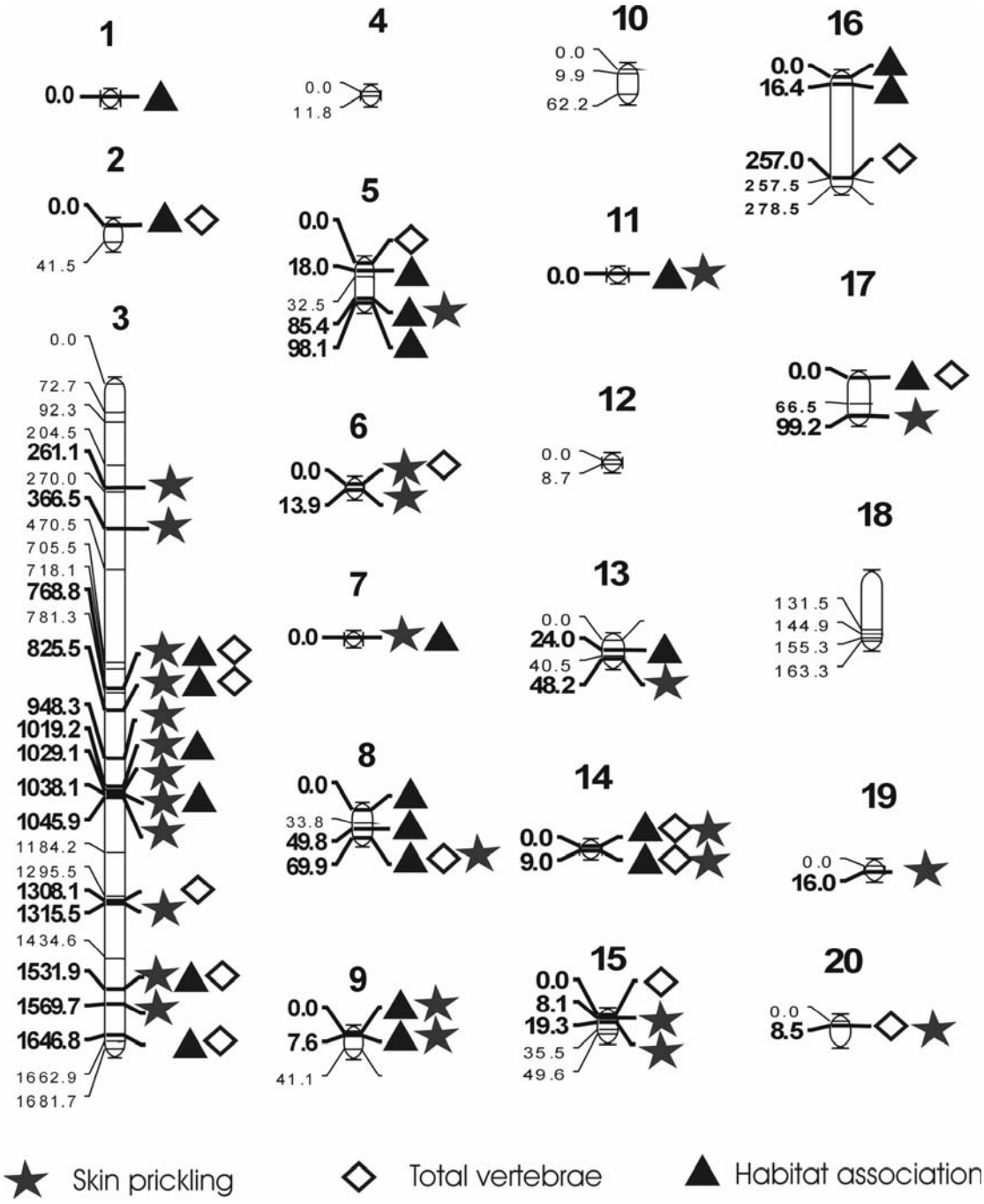


Figure 3: The relative genomic distribution of candidate regions and their association with different characters. A preliminary genetic map of *Cottus* (Stemshorn *et al.* 2005) serves as a backbone to visualize the genomic localization of all analysed loci (Map position in cM) and those positions, for which significant ancestry association was detected (bold) on known linkage groups (numbered). Individual sites may contain multiple loci (see Tab. 2). Continuous candidate regions often affect more than one character as indicated by different symbols implying physical linkage of the underlying genetic factors.

Besides within trait correlations many candidate regions are associated with more than one character. This is true for single marker loci (Tab. 2) but particularly evident when the overall distribution of positive loci is plotted on the preliminary genetic map of *Cottus* (Fig. 3). The genomic distribution of trait association reveals candidate regions that apparently affect more than one of the characters studied here. 48 % of the continuous candidate regions are associated with two characters and 20 % are even associated with both morphological characters as well as with habitat (Fig. 3). Several of these regions span distances between 10 and 25 cM with particularly large ones located on linkage groups 3 and 16.

The distribution of habitat associated candidate map positions relative to candidate map positions associated with a morphological trait was determined for 8 linkage groups which carry both habitat and morphology associated candidates together with markers showing no significant association (Fig. 3). The average map distance of habitat candidate position to the nearest morphotype candidate position was 33.7 cM (range: 0-257 cM) whereas the average distance to the nearest non-candidate marker was 76.6 cM (range: 12.5-257.5 cM) based on 18 paired comparisons. This difference is significant at $p = 0.001$ (Wilcoxon test) implying genomic proximity of habitat and morphology candidates studied here.

Discussion

A strong dependence of trait values upon individual admixture proportions (Fig. 2) documents nicely that the phenotypes scored here are indeed heritable and thus confirms the validity of admixture mapping to study these traits. Representatives of ancestral populations form well-differentiated groups and hybrid individuals display intermediate phenotypes on average. In the admixture mapping approach used in this study, a comparison of ancestry association with classical QTL studies is not straightforward. However, individual genetic factors are hard to identify even with much more sophisticated approaches and would at least require denser sampling of the genome and certainty about map positions of markers (MacKay 2001). The discussion is therefore focussed on the relative genomic positions of candidate regions regardless of the actual map resolution. Candidate loci appear to be clustered both for the same as well as for different characters implying linkage or possibly identity of some of the underlying genes. The following discussion will outline how this genomic correlation of characters is relevant for the genetic architecture of divergence across hybrid zones and the history of hybridization that has shaped the invasive sculpins genepool.

Candidate regions and their genomic distribution

Significant ancestry associations with characters were found for a number of markers but the corresponding genomic regions are unstudied in sculpins to date. Thus, the best confirmation that we are not dealing with artefacts comes from the fact that tightly linked genetic markers often reveal

the same signal in terms of a significant association with the same trait (Tab. 2). With a candidate locus at hand the genome architecture of *Tetraodon nigriviridis* provides additional evidence since it is possible to transfer positional information to the *Cottus* genome (Stemshorn *et al.* 2005). *Hox* genes are thought to affect the number of vertebrae in fishes (Anand *et al.* 2003, Ahn and Gibson 1999). In agreement with the position of *Hox* genes on *Tetraodon* chromosomes the corresponding linkage groups in *Cottus* harbour significant candidate regions associated with the number of vertebrae (*Tetraodon* chr.1/*Cottus* linkage group 3&16; 2/3; 8/2; 9/5; 21/8; *Tetraodon* - Ensembl v31.1c). A potential candidate gene for skin prickling is represented by the *Ectodysplasin* gene that was recently shown to affect the lateral plate reduction in *Gasterosteus* (Colosimo *et al.* 2005), a trait that is likely to be homologous with skin prickling and shows parallel reduction phenotypes. In *Tetraodon* this gene occurs on chromosome 1, which may correspond to parts of *Cottus* linkage group 3 that also carries a number of candidate regions associated with prickling. Such crude associations are by no means conclusive, but do not exclude the possibility that known candidates may play a role in *Cottus* as well.

A suite of markers is associated with habitat more than expected from the ancestry of specimens. Depending on the validity of the assumptions these genomic regions would carry genetic factors that contribute to ecological differentiation across hybrid zones. An alternative explanation would be given by traits that are related to intrinsic genomic incompatibilities of the mixing genepools, however, this is harder to reconcile with the continuous admixture observed at sculpin hybrid zones (Chapter 4). Still, it remains a challenge to disentangle purely environmental from intrinsic genetic effects at hybrid zones (Shaw *et al.* 1993). Ancestry association of habitat and a marker is not as clear as it is with morphological traits for several reasons. First, it rests on the assumption that the rough classification of habitats sufficiently reflects the situation in nature. The River vs. Stream classification does not consider the distance to the outlet and thus assumes an abrupt ecological transition. This is, however, in line with observations on faunal transitions from streams to rivers in the study area (Chapter 4) and also implied by the abrupt physical habitat border as described in the River Continuum Concept (Vannote *et al.* 1980). Furthermore, “habitat” does not represent an individual measurement but a classification. The invasive sculpins prevalence in riverine habitats results from a historical process (recent immigration) and could thus be interpreted as a mere correlation of genotypes with habitat unless determinism is demonstrated. The assumption that “habitat” integrates a genetically determined component rests on the observation that immigrant genotypes are apparently selected against (Chapter 4). Despite these complications, a test of association of locus ancestry with habitat would reveal those traits in which ancestry best explains the presence of individuals in a given habitat, which in turn would be likely to be related to differential adaptation.

Note that ancestry based estimates of habitat association for a given genomic region differs from approaches that are based on measurements of geneflow as approximated by F_{ST}. Habitat association as inferred here does not make pooled estimates per sampling site as association with locus

ancestry is conditioned for individually. Moreover ancestry based estimates of habitat association would be less biased by the total variability of markers (compare Hedrick 1999) as long as markers are sufficiently informative of ancestry.

Genomic correlations in an Evolutionary Perspective

The most conspicuous pattern in the genomic distribution of candidate regions is that a majority of them is associated with more than one character (Tab. 2, Fig. 3). Several of these span map distances of 20 cM including particularly large ones on linkage groups 3 and 16. Theory and empirical studies suggest that admixture linkage disequilibrium extends over 5-20 cM (Briscoe *et al.* 1994, Collons-Schramm *et al.* 2003) for the timeframe that we are dealing with in the sculpin system. Under the assumption of equal recombination rates, this would suggest the involvement of several factors within larger regions while smaller regions could be explained by a single gene. It is impossible to predict numbers of divergent traits for a pair of divergent lineages except for the notion that divergence should accumulate with time (Orr 2001). Here, it needs to be emphasized that the genepool of invasive sculpins traces back to old phylogeographic lineages that have been separated for about 1 Myr (Englbrecht *et al.* 2000) before they have recently mixed to give rise to the invasive sculpin. With respect to the genetic architecture across hybrid zones it is therefore plausible to assume that divergence of traits could trace back to older lineages that gave rise to the invasive sculpin (Chapter 1).

Irrespective of the number of underlying genes the genomic distribution of candidate regions suggests physical linkage of factors that affect morphology and association with habitat. If the latter were indeed involved in adaptation, all others would be affected equally by environmentally induced selection. It would even be possible that the morphological characters contribute to adaptation themselves but this remains to be demonstrated. A genetic architecture of adaptive traits physically linked to particularly divergent morphological traits could contribute to the overall divergence of specialists selected for in different environments. Hawthorne and Via (2001) describe such a situation for pea aphids, yet their example also includes mate choice, which seems to have little effect on the dynamics in the sculpin hybrid zones (Chapter 4). Still, the morphological transition across sculpin hybrid zones can now be assumed to be enhanced by linkage of the underlying genetic factors to ecologically relevant traits. Unfortunately, what facilitates admixture mapping at the same time prohibits far reaching analysis of introgression patterns as the hybrid zones studied here are probably too young to allow for pronounced differential introgression.

The fact that the invasive sculpin morphology resembles one of its ancestors (Sheldt sculpins) seems paradoxical given that a considerable amount of its genepool originates from the morphologically different Lower Rhine lineage (Chapter 1). The observed cohesion of traits during the course of past hybridization requires genomic cohesion of genetic factors that determine the observed Scheldt – like morphology. These would have had to resist the past hybridization process that has affected the present invasive

sculpin genepool. This study provides first evidence that cohesive genomic regions may exist, which carry both morphologically and ecologically relevant traits. If the latter were selected for in large river environments, then the prevalence of genomically linked morphological traits in a hybrid genepool would be easier to explain. This scenario is supported by the pattern of colonization within the Lower Rhine, which suggests that invasive sculpins have a fitness advantage relative to their ancestors (Chapter 1 and 4). Accordingly, the candidate regions found in this study may have been subject to environmentally induced selection. This process would have set an evolutionary trajectory of the hybrid genepool in trapping neighboring genomic regions.

On the one hand this underlines a contribution of the Scheldt sculpins to the invasive genepool, especially with respect to adaptive traits. On the other hand, it has now become clear that this effect may be restricted to fractions of the genome. In fact, this makes the observed admixture of the remaining genome even more conspicuous with respect to hybrid speciation processes. A better understanding of the consequences of hybridization for the invasive sculpin will rely on a better knowledge of the relevant adaptations and how the underlying genetic architecture in the hybrids relates to both ancestral lineages.

Acknowledgements:

We thank A. Mellin, T. Heilbronner and W. Fettweis for permissions to sample and take specimens. B. Harr has written a tremendously helpful perl script to produce input files for ADMIXMAP. The radiographies of sculpins were made with support of K. Lackner and B. Krug (Institut und Poliklinik für Radiologische Diagnostik, University of Cologne). R. Girnus and K. Stemshorn helped with taking the radiographies of sculpins. Susanne Kip has helped with the morphological analysis of sculpins.

Literature:

Ahn, D. Gibson, G. (1999) Expression patterns of threespine stickleback Hoxgenes and insights into the evolution of the vertebrate body axis. Dev. Genes Evol. 209:482–494.

Anand, S. Wang, W. C. H. Powell, D. R. Bolanowski, S. A. Zhang, J. Ledje, C. Pawashe, A. B. Amemiya, C. T. Shashikant, C. S. (2003) Divergence of Hoxc8 early enhancer parallels diverged axial morphologies between mammals and fishes. PNAS, 100. 26. 15666–15669.

Arnold, M. L. Bulger, M. R. Burke, J. M. Hempel, A. L. Williams, J. H. (1999) Natural Hybridization: how low can you go and still be important? Ecology, 80, 2, 371–381

Barton, N. H. (2001) The role of hybridization in evolution. Mol. Ecol. 10, 551-568.

Briscoe, D. Stephens, J. C. O'Brien. S. J. (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. Journal of Heredity, 85:59–63.

Collins-Schramm, H. E. Chima, B. Operario, D. J. Lindsey A. Criswell, L. A. Seldin, M. F. (2003) Markers informative for ancestry demonstrate consistent megabase-length in the African American population. Hum. Genet. 113: 211–219.

Colosimo, P. F. Hosemann, K. E. Balabhadra, S. Villarreal, G. Dickson, M. Grimwood, J. Schmutz, J. Myers, R. M. Schluter, D.Kingsley, D. M. (2005) Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles.Science 307, 1928-1933.

Englbrecht, C.C. Largiader, C.R. Hänfling, B. and Tautz, D. (1999) Isolation and characterization of polymorphic microsatellite loci in the European bullhead Cottus gobio L. (Osteichthyes) and their apllicability to related taxa. Molecular Ecology, 8, 1966 - 1969.

Englbrecht, C. C. Freyhof, J. Nolte, A. Rassmann, K. Schliewen, U. Tautz, D. (2000) Phylogeography of the bullhead Cottus gobio (Pisces: Teleostei: Cottidae) suggests a pre-pleistocene origin of the major central European populations Mol. Ecol. 9, 709-722.

Falush, D. Stephens, M. and Pritchard, J.K. (2003) Inference of population Structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics, 164: 1567-1587.

Hawthorne, D. J. Via, S. (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature, 412, 904-907

Hedrick, P. W. (1999) Perspective: Highly variable loci and their interpretation in evolution and conservation. Evolution, 53, 313-318.

Lowe, A. Harris, S. Ashton, P. (2004) Ecological genetics: design, analysis, and application. Blackwell Publishing, Oxford, UK. vii-x+320pp.

MacKay, T. F. C. (2001) The genetic architecture of quantitative traits. Annu. Rev. Genet. 35:303–39.

McKeigue, P.M. (1998) Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. Am. J. Hum. Genet. 63:241–251.

McKeigue, P. M. Carpenter, J. Parra, E.J. Shriver, M. D. (2000) Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. Annals of Human Genetics. 64: 171-86.

McKeigue, P.M. (2005) Prospects for Admixture Mapping of Complex Traits. Am. J. Hum. Genet. 76:1–7.

Nolte, A. W. Stemshorn, K. C. and Tautz, D. (2005). Direct cloning of microsatellite loci from Cottus gobio through a simplified enrichment procedure. Molecular Ecology Notes (in press).

Orr, H. A. (2001) The genetics of species differences. Trends in Ecology and Evolution. 16, 343–350.

Rieseberg, L. H. Buerkle, C. A. (2002) Genetic mapping in Hybrid Zones. The American Naturalist, 159, Supplement. S36–S50.

Shaw, D. D. Marchant, A. D. Contreras, N. Arnold, M. Groeters, F. Kohlmann, B. C. (1993) Genomic and Environmental Determinants of a Narrow Hybrid Zone: Cause or Coincidence. In: Hybrid zones and the

evolutionary process, edited by R. G. Harrison, Oxford University Press, New York.

Shriver, M. D. Parra, E. J. Dios, S. Bonilla, C. Norton, H. Jovel, C. Pfaff, C. Jones, C. Massac, A. Cameron, N. Baron, A. Jackson, T. Argyropoulos, G. Jin, L. Hoggart, C. J. McKeigue, P. M. Kittles, R. (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. Human Genetics 112:387-399.

Shriver, M. D. Smith, M. W. Jin, L. Marcini, A. Akey, J. M. Deka, R. Ferrell, R. E. (1997) Ethnic-affiliation estimation by use of population-specific DNA markers. American Journal of Human Genetics, 60:957-964.

Slate, J. Visscher, P. M. MacGregor, S et al. (2002) A genome scan for QTL in a wild population of red deer (Cervus elaphus). Genetics, 162, 1863–1873.

Slate, J. (2005) Quantitative trait locus mapping in natural populations: progress, caveats and future directions. Molecular Ecology, 14, 363-379.

Stemshorn, K. C. Nolte, A. W. Tautz, D. (2005). A Genetic Map of Cottus gobio (Pisces, Teleostei) based on microsatellites can be linked to the Physical Map of Tetraodon nigroviridis. J. Evol. Biol (in press)

Stephens, J. C. Briscoe, D. O'Brien, S. J. (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am J Hum Genet 55:809–824.

Vannote, R. L. Minhall, G. W. Cummins, K. W. Sedell, J. R. Cushing, C. E. (1980) The river continuum concept. Can. J. Fish. Aquat. Sci. 37, 130-137.

Appendix – Chapter 6

Chapter 6 - Supplementary Table 1: Individual genotypic data for 171 Microsatellite Loci. Table with Individuals, Microsatellite Genotypes of all specimens (? = missing data, alleles numbered according to size, but not necessarily repeat size) with sampling site as of Tab. 1. Data format: Table; saved row by row with fields separated by semicolons. Ends of rows are marked by the insertion of "XXX".

Chapter 6 – Supplementary Table 2: Joint table containing 45 most informative Microsatellite Loci used to infer population structure and those 134 Loci of the total dataset with known map positions (Stemshorn *et al.* 2005) used to explore candidate regions.

Chapter 6 – Supplementary Table 3: Individual ancestries with phenotypic and habitat data. Individual ancestry in the invasive sculpin lineage as inferred in STRUCTURE, centroid size, sex (0 = female; 1 = male), the total number of vertebrae, degrees of skin prickling and the habitat classification. Missing data are indicated by (?).

Abstract

Fish abundance surveys have shown that the Lower Rhine was invaded by sculpins (*Cottus* sp.) within two decades. These fish are found in habitats that are untypical for *Cottus gobio*. In order to find source populations and reasons for this invasion an evolutionary genetic analysis of populations from the Rhine and surrounding rivers was conducted. A combined analysis of nuclear and mitochondrial markers suggests that invasive sculpins are hybrids between lineages from the Rivers Scheldt and Rhine. The hybrids form a distinct genetic group and possess a unique ecological potential, attributed here to the process of hybridization.

To explore the genetics of invasive sculpins 177 microsatellite markers were developed and a linkage map for *Cottus* was constructed from F_1 crosses. BLAST searches with microsatellite flanking sequences yield significant hits with the *Tetraodon nigroviridis* genomic sequence for 45% of the *Cottus* loci. Comparisons of the map locations between the genomes reveal extensive conserved synteny, suggesting that the *Tetraodon* genomic sequence will serve as an excellent reference for *Cottus*.

The novel markers and linkage information are applied in population genetic studies and to initiate first QTL approaches. Evidence that invasive sculpins have a unique ecological potential to colonize large rivers is derived from recently formed hybrid zones where secondary contact postdates the appearance of invasive sculpins in the early 1990's. Hybrid zones are moulded by exogenous selection in the absence of intrinsic reproductive barriers.

Hybridization receives attention because of the role that it may play in generating evolutionary novelty, which is tested in recent hybrids. Body shape is analysed with a new distance based assignment method using morphometric data. Shape is highly informative in that populations are sufficiently different to assign "unknowns" to their source. Recent hybrids are intermediate between parental groups but display additional differentiation. This suggests transgressive segregation and genetic divergence in developmental processes underlying body shape.

Natural hybrid zones offer an alternative for QTL mapping because far-reaching associations between states of ancestry at physically linked loci are generated in segregating hybrids. Ancestry association is exploited by admixture mapping, which is applied for the first time in wild populations. The *Cottus* genome was screened for association with morphological characters as well as with habitat as a surrogate of ecological differentiation. Candidate regions often affect more than one trait. This implies linkage of traits affecting divergent morphology and fitness. Selection at hybrid zones likely affects linked morphological traits. Likewise, the differentiation of invasive sculpins despite past hybridization could be explained by linkage of morphological with adaptive traits.

Zusammenfassung

Fischbestandserhebungen in Rheingebiet haben gezeigt, dass das Niederrheingebiet in den letzten zwei Jahrzehnten von Groppen (*Cottus* sp.) neu besiedelt wurde. Diese Fische wurden in für *Cottus gobio* als untypisch geltenden Lebensräumen gefunden. Um die Quellpopulationen der Besiedlung zu identifizieren und die Gründe zu verstehen, die diesen Prozess ermöglicht haben, wurde eine evolutionsgenetische Analyse von Groppenpopulationen im Rheingebiet und angrenzenden Flüssen durchgeführt. Eine kombinierte Analyse nukleärer und mitochondrialer Marker hat gezeigt, dass die „Invasiven Groppen“ Hybride darstellen, die auf Vorfahren aus dem Schelde- und Rheingebiet zurückgehen. Diese Hybride bilden eine distinkte und homogene genetisch Gruppe über ihr gesamtes Verbreitungsgebiet. Sie besitzen offensichtlich ein besonderes ökologisches Potential, welches hier auf Hybridisierung zurückgeführt wird.

Um die Genetik der Invasiven Groppe weiter zu untersuchen, ist ein genomweiter Zugang zu deren Genpool nötig. Mit Hilfe eines vereinfachten Anreicherungsverfahrens wurden 177 neue Mikrosatelliten Marker entwickelt. Diese wurden dann für die Erstellung der ersten Kopplungskarte für die Groppe verwendet, wobei F_1 Kreuzungen benutzt wurden. BLAST Vergleiche von flankierenden Sequenzen der Mikrosatelliten Loci mit der genomischen Sequenz von *Tetraodon nigroviridis* ergeben signifikante Treffer für 45% der Marker aus *Cottus*. Vergleiche der Kartenpositionen zwischen beiden Arten zeigen, dass weitreichend konservierte Syntenie zwischen den Genomen besteht. Damit stellt *Tetraodon* eine hervorragende genomische Ressource für weitere Untersuchungen an der Groppe dar. Die neuen Marker, zusammen mit den Kopplungsinformationen, wurden für weitere populationsgenetische Analysen und Ansätze von QTL Analysen verwendet.

Hinweise, dass die Invasiven Groppen ein besonderes ökologisches Potential für die Besiedlung grösserer Flüsse besitzen, konnten an neu gebildeten Kontaktzonen gefunden werden. Sekundärer Kontakt mit typischen Bachgroppen ist erst nach dem Erscheinen der Invasiven Groppen in den neunziger Jahren des letzten Jahrhunderts entstanden. Die genetische Struktur dieser jungen Kontaktzonen legt umweltinduzierte natürliche Selektion bei gleichzeitigem Fehlen intrinsischer Fortpflanzungsbarrieren nahe. Eine ökologische Unterlegenheit der jungen Hybride äußert sich darin, dass diese nicht erfolgreich mit ihren Elternpopulationen konkurrieren.

Durch transgressiven Segregation kann Hybridisierung eine wichtige Rolle bei der Entstehung evolutiver Neuheiten spielen. Hier wird dieses Phänomen mit einer neuen distanzbasierten Zuordnungsmethode für morphometrische Daten an jungen Hybriden untersucht. Die Körperform stellt einen sehr informativen Merkmalskomplex dar, der es ermöglicht, „Unbekannte“ korrekt ihren Quellpopulationen zuzuordnen. Hybride sind intermediär entlang der Achsen, die Gruppen ihrer Elternpopulationen trennen, aber zeichnen sich durch eine zusätzliche, ihnen eigene Formkomponente aus. Diese erklärt sich am besten durch transgressive Segregation, was eine genetische Divergenz von Entwicklungsprozessen der Elternpopulationen nahelegt.

Natürliche Hybridzonen bieten eine Alternative zu klassischen QTL Kartierungsmethoden, da in segregierenden Hybriden weitreichende genetische Assoziationen zwischen physikalisch gekoppelten Genorten entstehen. „Ursprungsassoziationen“ werden in der „admixture mapping“ Methode hier erstmals genuzt, um in Wildpopulationen von Nicht - Modellorganismen genetische Faktoren zu kartieren. Ein großer Teil des *Cottus* Genoms wurde auf Assoziation mit morphologischen Merkmalen, sowie dem Habitat, als Näherung für ökologische Differenzierung untersucht. Kandidatenregionen sind weit über das Genom verteilt und haben oft einen Effekt auf mehr als ein Merkmal. Dies legt eine physikalische Kopplung von divergenten morphologischen Merkmalen und fitnessrelevanten Merkmalen nahe. Damit würde Selektion an Kontaktzonen einen direkten Effekt auf morphologische Differenzierung haben. Weiterhin könnte die morphologische Differenzierung der Invasiven Groppen, trotz vergangener Hybridisierung, durch Kopplung von adaptiven und morphologischen Merkmalen erklärt werden.

Erklärung

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 dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. D. Tautz betreut worden.

Arne Nolte,

Köln den 17.05.2005

Teilpublikationen

Die folgenden für die Publikation akzeptierten Manuskripte basieren auf Teilen dieser Arbeit und entsprechen den Kapiteln 2 und 3.

Nolte, A. W. Stemshorn, K. C. and Tautz, D. (2005). Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Molecular Ecology Notes (*in press*).

Stemshorn, K. C. Nolte, A. W. Tautz, D. (2005). A Genetic Map of *Cottus gobio* (Pisces, Teleostei) based on microsatellites can be linked to the Physical Map of *Tetraodon nigroviridis*. Journal of Evolutionary Biology (*in press*).

Lebenslauf

Name: Arne W. Nolte

Geburtstag, Geburtsort: 13.07.1973 in Oldenburg

Adresse: Berrenrather Str. 409, 50937 Köln

Familienstand: ledig

Schulbildung:

1981-1985: Grundschule Sandkrug
1985-1987: Orientierungsstufe, Waldschule Sandkrug
1987-1994: Graf Anton Günther Gymnasium Oldenburg
1990-1991: Moss Point High School, Moss Point, Mississippi, USA
1991-1994: Graf Anton Günther Gymnasium Oldenburg – Abitur

Wehrdienst:

1994-1995: Grundwehrdienst bei der Luftwaffe.

Studium:

1995-1997: Carl von Ossietzky Universität Oldenburg – Vordiplom Biologie
1997-2000: Carl von Ossietzky Universität Oldenburg – Hauptstudium mit den Schwerpunkten Zoologie, Ökologie und Genetik
2000-2001: Diplomarbeit bei Prof. D. Tautz am Institut für Genetik der Universität Köln – Titel: „Populationsgenetische und morphologische Differenzierung von Groppen (*Cottus* sp. Scorpaeniformes, Teleostei) im Gebiet des Niederrheins“.
seit Dez. 2001: Beginn der Promotion bei Prof. D. Tautz am Institut für Genetik der Universität Köln – Titel: „Evolutionary genetic analysis of an invasive population of sculpins in the Lower Rhine“.

Juli 2005: voraussichtlicher Abschluss der Promotion.

Köln, den 17.05.05

Arne Nolte