A Systematic Assessment of Signatures of Positive Selection Events in Natural Populations of the House Mouse

Inaugural – Dissertation

Zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Meike Thomas
aus Traben-Trarbach

Köln, 2006

Berichterstatter:	**Prof. Dr. Diethard Tautz**
	Prof. Dr. Thomas Wiehe
Tag der letzten mündlichen Prüfung:	**01. Juni 2006**

List of Contents

Danksagung

Bedanken möchte ich mich bei all jenen, die mich im Laufe meiner Dissertation unterstützt haben und es somit ermöglichten, dass die Arbeit nun in dieser Form vorliegt.

Mein Dank gilt vor allem Herrn Prof. Dr. Diethard Tautz, der es mir ermöglichte meine Promotion in seiner Arbeitsgruppe durchzuführen. Seine ständige Diskussionsbereitschaft und seine Ideen haben den Erfolg dieser Arbeit maßgeblich mitbestimmt. Ohne Zweifel hat mich die Zeit in seinem Labor stark beeinflusst und mein Verständnis für die Biologie enorm gesteigert.

Besonders möchte ich Prof. Dr. Thomas Wiehe danken, der – neben der Übernahme des zweiten Gutachtens – zu allen Teilen meiner Arbeit einen wesentlichen Beitrag geleistet hat. Er war stets bereit, unzählige Fragen zur Datenanalyse und Interpretation meiner Ergebnisse ausführlich zu diskutieren und Lösungen zu erarbeiten.

Zudem danke ich Herrn Prof. Dr. Harmut Arndt für die freundlich Übernahme des Prüfungsvorsitzes und Dr. Bettina Harr als Beisitzerin in meinem Prüfungskommitee.

Einen großen Dank möchte ich auch Prof. Dr. Miguel Vences aussprechen. Ohne ihn wäre mir der Schritt ins Tautz-Labor sicherlich nicht gelungen. Ausserdem habe ich dank ihm den Anschluss an die herpetologische Arbeit während der gesamten Zeit nicht verloren. Durch seine stetigen Arbeitsanregungen habe ich es geschafft, neben der eigentlichen Dissertation auch der Froschforschung eine erhebliche Datenmenge beizutragen.

Mein Dank für unzählige Unterstützungen, Erklärungen und Denkanstöße gilt Dr. Bettina Harr, Dr. John Baines und Dr. Arne Nolte. Die vielen Diskussionen mit ihnen haben mein Verständnis für die Zusammenhänge in der Evolutionsgenetik erheblich gefördert und somit einen großen Beitrag zum Gelingen dieser Arbeit geleistet.

Der gesamten Tautz Gruppe möchte ich für die schöne Arbeitsatmosphäre danken, besonders Ruth Rottscheidt für die vielen guten Gepräche und den täglichen Spaß in den Pausen. Dank auch an Fabian Staubach, Kathryn Stemshorn, Chris

Voolstra und Till Bayer mit denen ich neben der schönen Zeit bei der Arbeit auch noch eine Menge Spaß auf der Piste hatte.

Neben den vielen hilfreichen Programmier-Arbeiten mit denen Lars Teschke mir bei der Durchführung meiner Promotion unendlich geholfen hat, möchte ich mich bei ihm dafür bedanken, dass er immer ein offenes Ohr für all die vielen kleinen Probleme hatte, die eine Doktorarbeit so mit sich bringt. Ohne seine ständige Motivation und seine Ruhe hätte ich die Zeit sicherlich nicht so ‚entspannt' durchlebt.

Meiner Mutter, Ruth Glauer, möchte einen ganz besonderen Dank dafür aussprechen, dass sie es mir ermöglichte, meinem Interesse an der Biologie nachzugehen, mich immer auch darin unterstützt hat weiterzumachen und den Verlauf dieser Arbeit stets mit großem Interesse verfolgte.

Bedanken möchte ich mich für die finanzielle Förderung des Maus-Projektes durch die Volkswagen Stiftung.

Zusammenfassung

Frequenz und Struktur von positiven Selektionsereignissen in natürlichen Populationen sind von großer Wichtigkeit, gehören aber zu den noch wenig verstandenen Parametern in der Evolutionsbiologie. Ziel dieser Studie ist die Untersuchung dieser Basisparameter in einem populationsbasierten Ansatz. Positive Selektionsereignisse hinterlassen populationsgenetische Spuren ('Selective Sweeps'), welche in molekularen Studien systematisch identifiziert werden können. Positive Selektion kann zur Fixierung der vorteilhaften Mutation in einer Population führen. Aufgrund des so genannten 'hitchhiking effects', geht die Variabilität in den neutralen, flankierenden Regionen, welche physisch mit dem unter Selektion stehenden Locus verbunden sind, verloren. Diese 'Fußabdrücke' reduzierter neutraler Variabilität können genutzt werden, um systematisch nach positiv selektionierten (adaptiven) Mutationen zu suchen und erlauben eine Abschätzung ihrer Frequenz in einer gegebenen Population.

Unterschiede in der Rekombinationsrate entlang der Chromosomen können die Evolution neutraler Loci durch den 'hitchhiking effect' auch über einen langen Zeitraum beeinflussen. Generell sollten solche Effekte in niedrig rekombinierenden stärker als in hoch rekombinierenden Regionen sein. Detailierte Informationen der physikalischen und genetischen Karte der Hausmaus erlauben es, die Korrelation zwischen neutraler Variabilität und Rekombinationsrate in einer bestimmten chromosomalen Region zu untersuchen. Ich habe Mikrosatelliten aus chromosomalen Regionen unterschiedlicher Rekombinationsraten getestet und innerhalb von fünf beprobten Mauspopulationen (*Mus musculus musculus* and *M. m. domesticus*) keinen Hinweis auf eine Korrelation zwischen Mikrosatellitenvariabilität und Rekombinationsrate gefunden. Das deutet darauf hin, dass die hohe durchschnittliche Mutationsrate von Mikrosatelliten in Säugetieren die Effekte von weitreichendem 'hitchhiking' im Mausgenom ausgleicht.

Analysen, in denen die Variabilität neutraler Marker genutzt wird, um Regionen, die kürzlich unter positiver Selektion standen, zu identifizieren, werden als 'hitchhiking mapping' bezeichnet. Da die Genome hoeherer Eukaryoten ungefähr 40,000 selektierbare Loci enhalten und die Detektion von polymorphen Varianten des Testens von mindestens 20 Individuen aus verschiedenen Populationen bedarf, würde

ein kompletter Genome-Screen eine Bestimmung von Millionen von Genotypen bedeuten. Ich präsentiere hier eine Pool-Strategie, die es erlaubt, die Anzahl an Genotypisierungsreaktionen signifikant zu reduzieren. Die vorgestellte Hochdurchsatzroutine ermöglichte es mir, fast 1000 Mikrosatelliten in verschiedenen Populationen der Hausmaus zu untersuchen. Aus den durch diese Methode produzierten, spezifischen Mustern ist es möglich, visuell die Loci auszuselektieren, die eine populationsspezifische Reduktion in der Varibilität aufweisen. In einem zweiten Schritt werden diese Kandidaten erneut typisiert, diesmal für einzelne Individuen. Die Loci werden statistisch auf Signifikanz überprüft werden.

Eine detailierte Analyse der Kandidatenloci aus einem paarweisen Vergleich zweier Populationen ergab Resultate zur Frequenz von 'Selective Sweeps', der Stärke von Selektion in natürlichen Populationen und der Herkunft der selektierten Varianten. Die zwei untersuchten Populationen der Hausmaus (*M. m. domesticus*) trennten sich vor etwa vor 3000 Jahren nach deren Ankunft in Mitteleuropa. Die massive Invasion von Hausmäusen nach Europa ist, basierend auf fossilen Daten, gut dokumentiert. Starke, unabhängige, nachfolgende 'Bottlenecks' sind wegen den vergleichbaren hohen genom-weiten Variabilitäten in beiden Populationen wenig wahrscheinlich. Aus diesem Grund können komplexe demographische Einflüsse auf die Ergebnisse ausgeschlossen werden. Die identifizierten Kandidatenregionen zwischen diesen beiden Populationen werden durch die Genotypisierung flankierender Mikrosatelliten weitergehend charakterisiert. Die signifikante Abweichung der Kandidatenregionen von dem neutralen Zustand wird durch verschiedene statistische Analysen belegt. Basierend auf diesen Ergebnissen folgere ich, dass es mindestens ein positives Selektionsereignis je 100 Generationen in jeder Linie gegeben haben muss. Da keines der detektieren Sweep-Täler sehr breit ist, im Schnitt sind sie ca. 50 kb, kann gefolgert werden, dass positive Selektion im allgemeinen von Allelen vorangetrieben wird, die einen geringen selektiven Vorteil haben. Errechnete Selektionskoeffizienten variieren zwischen 0,0007 und 0,021. Weiterhin scheint es, dass die vorteilhaften Varianten im Allgemeinen aus der bestehenden Variabiltät hervorgehen und dass positive Selektion als kontinuierlich wirkender Hintergrundeffekt in allen Populationen abläuft.

Abstract

The frequency and the structure of positive selection events in natural populations are of central importance, but one of the least known variables in evolutionary biology. The aim of this study is to investigate these basic parameters in a population based approach. Positive selection events leave population genetical signatures (selective sweeps) behind, which can be systematically identified with molecular studies. Positive selection can lead to the fixation of a favorable mutation in a population. Due to an effect called hitchhiking, variability in the neutral flanking regions which are physically linked to the target of selection is lost. This footprint of reduced neutral variability can be employed to systematically screen for positively selected (adaptive) mutations and allows to estimate their frequency in a given population.

Differences in recombination rates along the chromosomes can influence the evolution of neutral loci via hitchhiking effects even on a large time scale, which would influence the results. Generally, these effects should be stronger in regions of low recombination than in regions of high recombination. The detailed information on physical and genetic maps in the house mouse allows now to assess the correlation between neutral variability and recombination rates at given chromosomal regions. I have tested microsatellite loci from chromosomal regions which show differences in recombination rates and found no evidence for a correlation between microsatellite variability and recombination rates in samples from five wild mice populations (*Mus musculus musculus* and *M. m. domesticus*). This suggests that the high average mutation rate of microsatellites in mammals counter balances the effects of long range hitchhiking in the mouse genome.

Approaches, in which the variability of neutral markers is used to identify regions which have recently been under positive selection, is termed hitchhiking mapping. Since higher eukaryotic genomes may contain about 40,000 selectable loci and the detection of polymorphic variants requires testing of multiple individuals (at least 20) for several populations, a complete genome scan would require millions of genotypes to be determined. I present here a pooling strategy that allows to reduce the number of genotyping reactions significantly. The presented high throughput routine enabled me to investigate almost 1,000 microsatellite loci in different populations of

the house mouse. Among the composite patterns that are obtained in this way, it is possible to visually select those with population specifical reduced variability. In a second step, these candidates were then re-typed in individuals of a carefully chosen population background and statistically tested for significance.

A detailed analysis of the candidate loci identified by a single comparison yield results on the frequency of selective sweeps, the strength of selection acting in natural populations and the origin of selected variants. The two investigated populations of the house mouse (*M. m. domesticus*) have split upon arrival in Middle Europe about 3,000 years ago. The massive invasion of house mice into Europe is well documented based on fossil records. Strong independent subsequent bottlenecks can be ruled out because of a comparable high genome wide variability in both populations. Thus, complex demographic influences on the results can be excluded. Identified candidate regions between the two focal populations were further characterized by genotyping additional microsatellites in the flanking regions of the identified candidates. The significant deviation of the candidate regions from the neutral state is supported by several statistical tests. Based on these results, I find that there was at least one positive selection event per 100 generations in each lineage. Since none of the detected sweep valleys is broad, on average they are about 50 kb, I conclude that positive selection in general is driven by alleles providing weak beneficial impact. Estimated selection coefficients vary between 0.0007 and 0.021. Furthermore, it seems that beneficial variants are generally taken from the standing variation and that positive selection is a continuously acting background effect in all populations.

Declaration

The design of the whole project was done together with Diethard Tautz. Practical laboratory work as well as the major parts of the data analysis was conducted by me. In the different chapters of this thesis I profited from contributions of several co-authors:

Chapter 1

Sonja Ihle introduced me in the laboratory methods, and together with Susanne Krächter she collected ot the mouse samples used in this approach. Some of the microsatellite data was generated by Iary Ravaoarimanana. Thomas Wiehe provided major parts of the background knowledge on data analysis.

Chapter 2

The software pipeline to identify occurrences of dinucleotide repeats in the upstream regions of annotated genes was written by Friedrich Möller together with the help of Thomas Wiehe.

Chapter 3

Diethart and I designed the project together, and the interpretations of all the different results were aquired during numerous discussions. Laboratory work as well as the major parts of data analysis were done by me.

1 General Introduction

1.1 Molecular Evolution – A Short Introduction

Since "All living things have evolved" (Stearns and Hoekstra 2005), a main theme in biology is to understand the processes and mechanisms that facilitate evolutionary change. Irrespective of their function, such changes must have a molecular basis in order to be inherited from generation to generation. Evolutionary changes manifest as mutations in the DNA sequence of single organisms and, at a later stage, these mutations spread in gene pools of populations and species. This study focuses on the appearance and spread of new genetic variants in order to obtain a better insight of the prevalence of these key evolutionary processes.

There are two forces, which drive molecular changes and therefore cause molecular evolution: natural selection and random genetic drift. Drift describes the changes in allele frequency due to the random drawing of gametes that will form the next generation (Hartl and Clark 1997). Thus, drift acts randomly on genetic variation and its impact is larger in smaller populations. In a mutation-drift equilibrium population the probability of fixation is the observed frequency of the allele, and the expected fixation time of new alleles equals four times the effective population size measured in generations.

Natural selection describes all forms of directional changes in the allele frequencies of populations. The direction in these processes is induced from the environment. Two major forms of natural selection are acting in populations: positive selection and negative selection. In practice negative selection is often discussed in the context of background selection (e.g. Nachman 2001), purifying selection (e.g. Hardison 2003, Khaitovich et al. 2004) or selective constraints (e.g. Eyre-Walker et al 2002, Bush and Lahn 2005). In general negative selection refers to the removal of those genetic variants that reduce the fitness of organisms. Positive selection is the force that increases the frequency of beneficial variants and therefore enables adaptive evolution. The fixation time of a favourable allele is not only dependent on the

effective population size but also on the fitness advantage of the beneficial allele. The degree of advantage is usually expressed as the selection coefficient. The value of the selection coefficient describes the decrease in allele frequency of other than the beneficial allele from one generation to the next. If the product of four times the effective population size and selection coefficient is larger than one ($4N_es > 1$) positive selection is acting in a population. Since the relative effect of drift is stronger in small populations, positive selection acts more effectively in large populations. Thus, positive selection events with selection coefficients of e.g. 10^{-3} would be classified as strong positive selection in large populations, like in *Drosophila*, but as weak selection in small populations, like in mice.

Data on the occurrence of positive selection is indispensable in understanding how adaptive molecular evolution acts. Thus, traces of positive selection in genomes are an intensively studies subject (e.g. Harr et al. 2002, Beisswanger et al. 2005). Apart of the gene, which gained a beneficial mutation, the physically linked flanking regions are also affected by the consequences of selection. Together with the beneficial allele of the target gene the flanking regions get a lift in frequency. This effect is called 'hitchhiking effect' (Maynard Smith and Haigh 1974). By the hitchhiking effect the neutral variability in flanking regions is wiped out during the fixation process (see Figure 1.1). Such a positive selection event is termed 'Selective Sweep'. Thus, positive selection leaves characteristic footprints of reduced variability at linked neutral loci in the genome (Maynard Smith and Haigh 1974, Slatkin 1995). The size of the region that is affected by a selective sweep depends on the local recombination rate and the selection coefficient (Maynard Smith and Haigh 1974). The lower the recombination rate and the higher the selection coefficient the larger is the region which is expected to exhibit the footprint of reduced variability (Figure 1.2). In the absence of recombination, hitchhiking eliminates all linked variation. In the presence of recombination, hitchhiking is incomplete since not all variation is removed (Fay and Wu 2000). Following this logic, the often observed positive correlation between levels of nucleotide diversity and recombination rate is widely interpreted as evidence of recurrent selective sweeps (Begun and Aquadro 1993, Nachman 1997, Nachman et al. 1998, Stephan and Langley 1998, Andolfatto and Prezeworski 2001, Betancourt and Presgraves 2002). For loci with very high mutation rates, genetic hitchhiking is expected to produce a correlation between variability

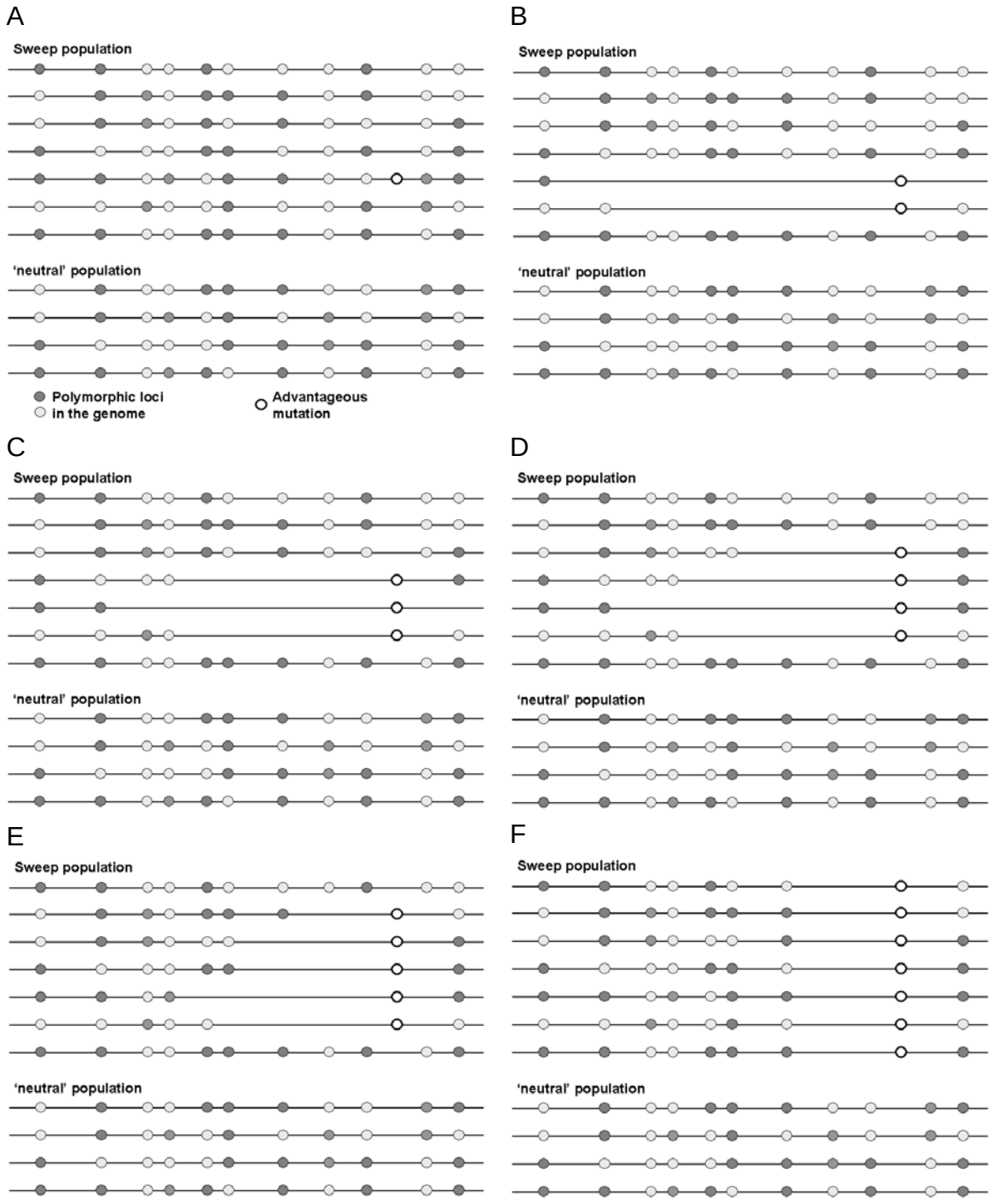


Figure 1.1 The hitchhiking effect. The figure depicts two populations, horizontal lines are chromosomes, the light and dark gray dots represent variable sites. Both populations show many neutral polymorphic sites between the different individuals. If an advantagous mutation appears in one of the two populations (A) this mutation will rise in frequency (C-E). Due to the hitchhiking effect (described by Maynard Smith and Haigh in 1974) the physically linked flanking regions will get a lift in frequency together with the advantageous allele, and thus variability between the individuals is lost in this region. During the time of fixation, the distal flanking regions recover variability because of recombination events. After such a selective sweep, a certain chromosomal area around the gene that carries the beneficial mutation exhibits reduced variability (H). Such reduced variability footprints of positive selection can be identified by comparing variability levels between distinct populations.

and recombination rate only if selective sweeps are very frequent. High mutation rates might counter balance the effects of low recombination rates because of accumulation of new mutations (Slatkin 1995, Wiehe 1998). Knowledge of the correlations between recombination rate and levels of polymorphism in natural populations is essential for projects in which signatures of selective sweeps are used to identify genes experiencing positive selection (Schlötterer 2002). During a selective sweep the increase in frequency of a favourable allele in time follows a sigmoid curve. A beneficial mutation appears in a population and starts at a low frequency. If it does not get lost due to drift, the frequency shows a rapid increase at a certain point in time, until the speed is reduced again before the frequency reaches a plateau, i. e. the allele is fixed. Recombination is expected to take place in the first phase of such a fixation process (Figure 1.3).

With time after the sweep event, the footprint is gradually lost. The recovery pattern is characterized by an excess of new mutations at low frequency. Thus, on the SNP level genetic hitchhiking is expected to produce a skew in the frequency distribution of segregating variants towards an access of rare polymorphisms in the population (Braverman et al. 1995). In microsatellites this recovery pattern can be observed as an increase of new alleles, which occur in a one- or few-repeat-steps distance around the sweep allele.

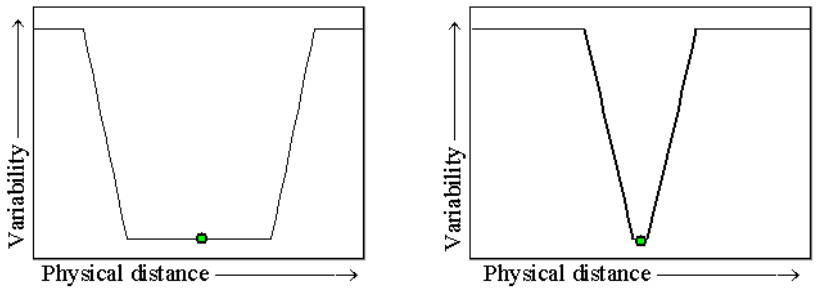


Figure 1.2 The size of the window is dependent on the local recombination rate. The left figure depicts the expected valley in a low recombining area, and the right figure shows a very narrow window, which is expected in a high recombination area for the same selection coefficients.

Screening for signatures of selective sweeps by comparing variability levels between populations is termed ‘hitchhiking mapping’ (Harr et al. 2002, Schlötterer 2003). The timeframe within which such footprints of positive selection are

observable in a hitchhiking mapping approach depends on the mutation rate of the investigated neutral marker. Suitable neutral marker systems for hitchhiking mapping are microsatellites (short tandemly repeated sequences of 1-6 bp in length) and single-nucleotide polymorphisms (SNPs). Although the substitution rates vary between different genomic regions, over the whole genome SNPs follow a more or less constant mutation rate of about 2.5 x 10^{-8} in humans (Nachman and Crowell 2000), 2.1 x 10^{-8} in mice (Nachman 1997), and 3 x 10^{-9} in insects (Andolfatto and Przeworski 2000). Whereas the mutation rate of microsatellites is highly locus specific. It is strongly correlated with the number of repeats a microsatellite carries. The longer a microsatellite, i.e. the more repeats it has, the faster it mutates (see Ellgren 2004 for review). The mutation rate of microsatellites is several orders of magnitude higher than of SNPs. Microsatellites of *Drosophila melanogaster* are short (on average up to 12 repeats) and this lowers their mutation rate. Schlötterer et al. (1998) estimated an average mutation rate of 6.3 x 10^{-6} per locus per generation in flies whereas for the longer microsatellites in humans, mice, rats and pigs mutation rates were estimated in a range of 10^{-2}-10^{-5} (see Schug et al. 1998). Thus, especially in mammals the signatures of selective sweeps are expected to be blurred more quickly by new mutations in microsatellites than on the SNP variability level.

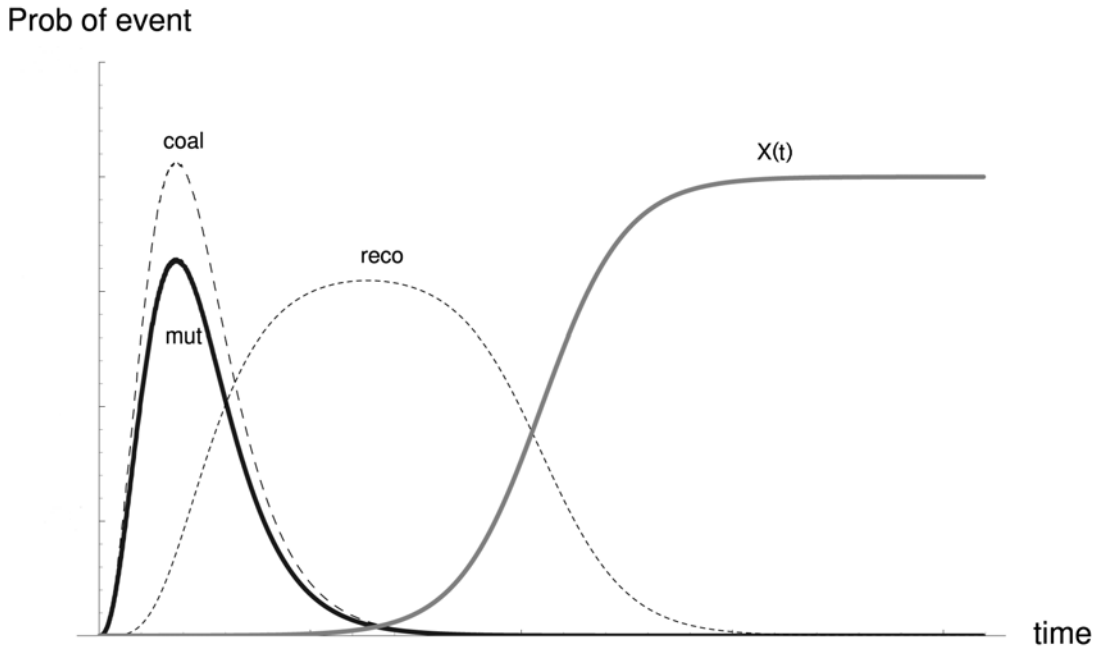


Figure 1.3 The probability that recombination, mutation or coalescence happen during the fixation time of a beneficial allele (Pennings and Hermisson in prep.).

In the genomic era the detection of positive selection on the genome scale is a challenge. Several studies on the theoretical (e.g. Maynard Smith and Haigh 1974, Wiehe 1998), as well as on the comparative genomic level (e.g.Smith and Eyre-Walker 2002, Birne and Eyre-Walker 2004, Bazykin et al. 2004, Keightley et al. 2005) and empirical data from natural populations (e.g. Payseur et al. 2002, Kauer et al. 2003, Kayser et al 2003, Storz et al 2004) have been published. There are two possible types of approaches, first, candidate gene approaches, in which the investigated genes are *a priori* expected to be under positive selection (e.g. resistance evolution (Kohn et al. 2000, Wootton et al. 2002), artificial selection (Vigouroux et al. 2002)), and second, whole genome approaches, where randomly selected genes throughout the genome are investigated. The first kind of approach gives insights into the general pattern of selective sweeps and can provide examples, whereas the second type leads to results concerning the frequency and intensity of positive selection in general. It is still unknown to which degree positive selection acts in nature. Do many genes change a little bit, or do just a few genes undergo great evolutionary changes? This question can only be answered by empirical data from natural populations. The two major model organisms for such population-based approaches are *Drosophila* flies and humans. The focus in experimental studies, so far, is on adaptive changes connected to range expansion (Harr et al. 2002, Kauer et al. 2003, Kayser et al. 2003, Storz et al. 2004). For humans and fruitflies this is the out of Africa event. In this context, again an *a priori* expectation for positive selection is given. Here, selective sweeps are not expected on a certain class of genes but within the derived populations because of large changes in their environmental background. From recently published data by Haddrill et al. (2005) it is known that demographic events can lead to an overestimate of positive selection. Bottlenecks, like the one in *Drosophila* during the out of Africa event can produce reduced variability patterns similar to those of selective sweeps. Thus the differentiation between footprints of positive selection and bottleneck artefacts in regions identified by reduced variability in a derived population compared to an ancestral one is difficult. The selection of a suitable model system for hitchhiking mapping approaches should therefore – beside other important factors – also take care of the demographic background of the populations to be investigated.

1.2 The house mouse

The house mouse has become one of the major model organisms in bio-medical science. It provides many advantages enabling a variety of applications especially in the field of genetics (for review see Guénet and Bonhomme 2003). For several decades the mouse phylogeny and history (e.g. Boursot et al. 1993, Boursot et al. 1996), as well as its behaviour (e.g. Reimer and Petras 1967, Lidicker 1976) was intensively studied. In addition, the complete genome sequence is available since 2002 (Mouse Genome Sequencing Consortium 2002). These data combined provide a perfect background for the study of molecular evolution.

The house mouse evolved on the Indian subcontinent. Less than one million years ago the species *Mus musculus* split into three major sub-species (Figure 1.4). The nominate subspecies *M. m. musculus* which is nowadays found all over northern Asia as well as in Eastern Europe, *M. m. domesticus*, which has its todays range in Western Europe, the Near East, Northern Africa, and recently introduced by humans into the New World, Subsaharan Africa and Australia, and a third subspecies, *M. m. castaneus,* spread all over South East Asia.

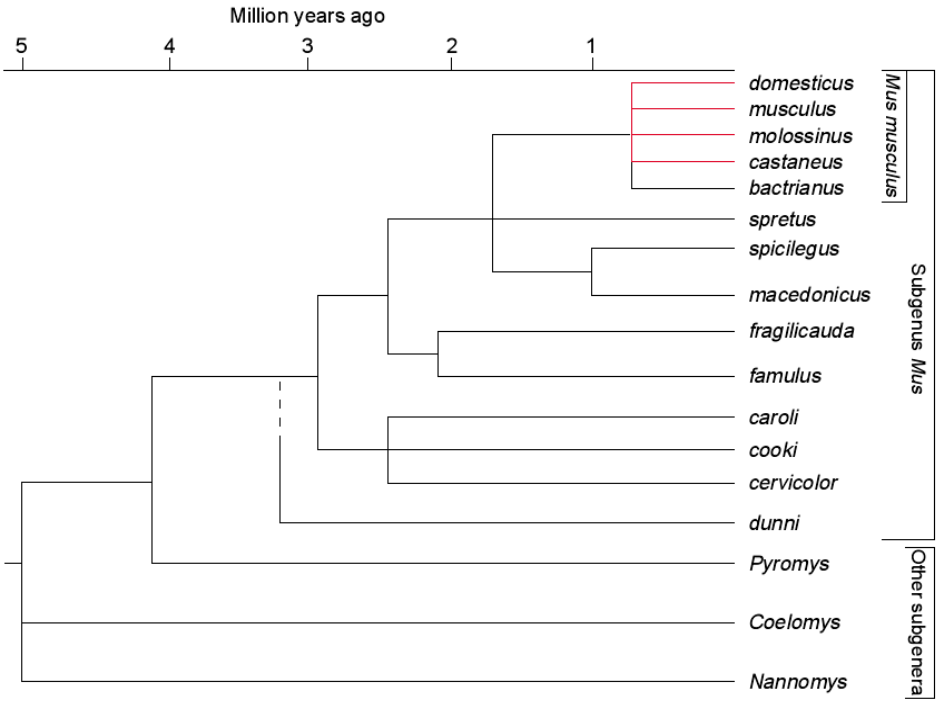


Figure 1.4 Evolutionary tree of the genus *Mus*. The time scale is based on single copy nuclear DNA hybridization studies and is calibrated with the separation of *Mus* and *Rattus*, estimated at 10 Myr ago (taken from Guénet and Bonhomme 2003).

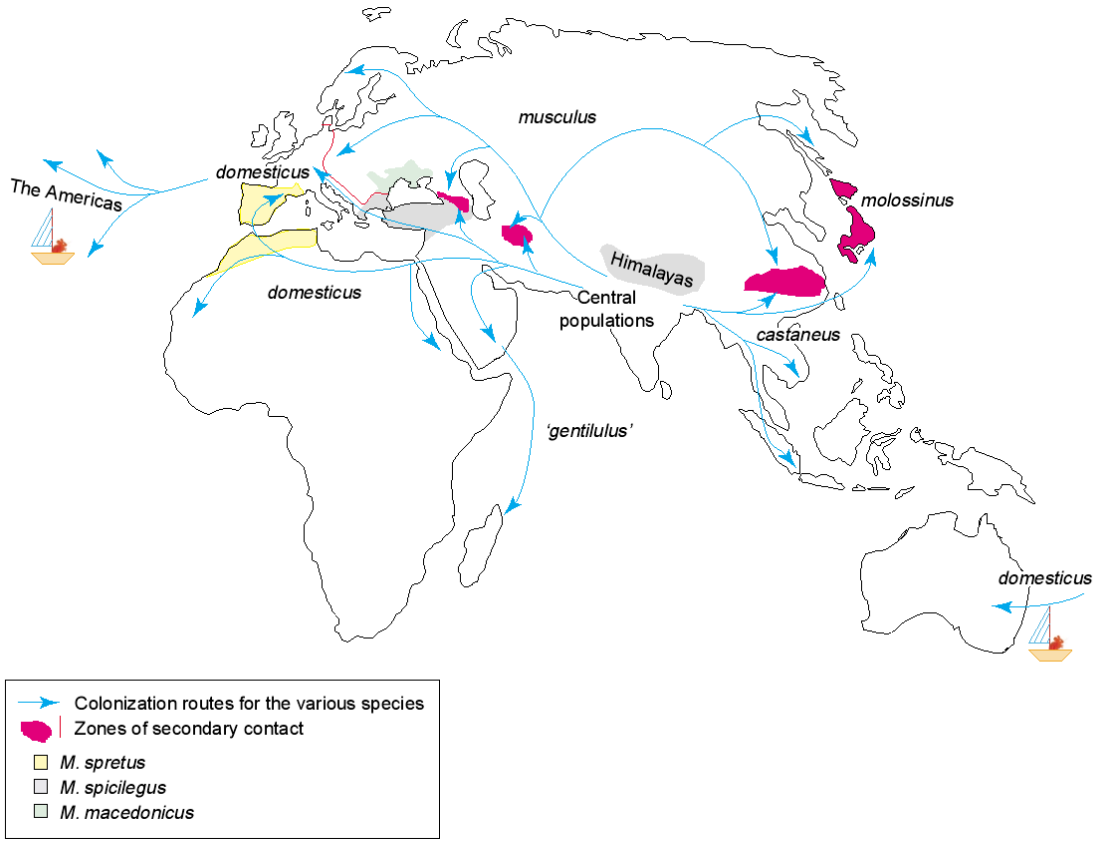


Figure 1.5 Geographical distribution of the different species of the genus *Mus* and routes of colonization. Mice of the American and Australian continents were imported by humans during colonization (taken from Guénet and Bonhomme 2003).

None of these subspecies are completely isolated genetically. There are several natural hybrid zones in the contact areas of their ranges (Figure 1.5). *M. m. musculus* and *M. m. domesticus* meet in Europe, in the Caucasus, and in a region southeast of the Caspian Sea. *M. m. musculus* and *M. m. castaneus* have a contact zone in China. In Japan, these two subspecies have hybridized extensively, giving rise to an unique population often referred to as *M. m. molossinus* (Yonekawa et al. 1988). The two distal subspecies *M. m. domesticus* and *M. m. castaneus* do not have a natural hybrid zone, but they can produce fertile offspring in the laboratory.

The available genome sequence of the house mouse is a mosaic of the three major lineages. About two third of the genome is of '*domesticus*'-origin and the remaining third is of non-*domesticus* origin, i.e. of '*musculus*' and '*castaneus*'- origin (Wade et al. 2002, Wade and Daly 2005). In a recently published paper, Sakai et al. (2005) showed that the '*dometicus*' background of most common laboratory mouse strains (one of which is C57BL/6J, the strain used for the genome sequence assembly)

is mainly derived from the Western European lineage. Thus, the available genome sequence is most similar to mice from these populations.

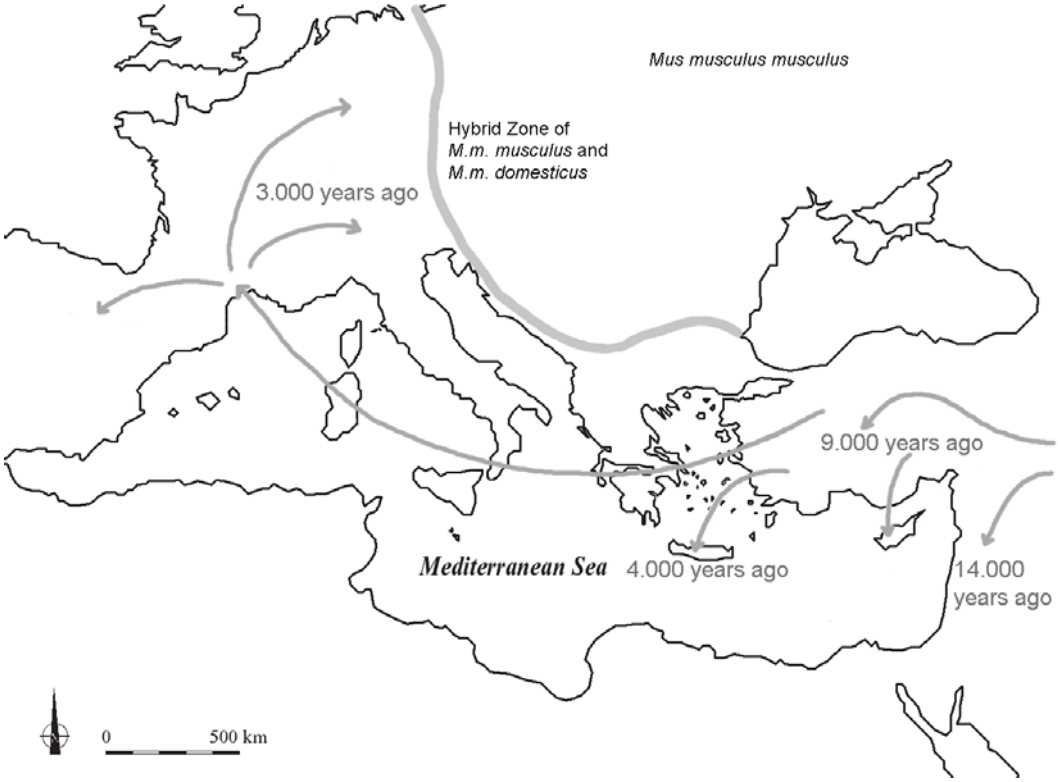


Figure 1.6 Colonization of Middle Europe by the house mouse *Mus musculus domestius* (Figure is based on data by Cucchi et al. (2005)). The mice followed the Neolithic in a massive invasion through Anatolia and then came into Southern France, most likely with Phoenician trading ships, from where they spread all over Western Europe.

The Western European mice have their origin in the Near East. The house mouse, *M. m. domesticus* moved westward from the Near East and entered Europe on the Mediterranean route. Cucchi et al. (2005) published details about the colonization of Western Europe by *M. m. domesticus* based on an analysis of palaeontological records. Their results clearly show that the house mouse invaded Europe not longer than 3,000 years ago via the Near East (in contrast to the route depicted in Figure 1.5 via northern Africa). While humans settled in Europe already 6,000 BC, the mice followed the Neolithic diffusion about 5,000 years later (Figure 1.6). Cucchi et al. (2005) mention three possible reasons for this delayed invasion. First, maritime exchanges were very limited between the Eastern and Western Mediterranean Basin until 1,000 BC, which consequently leads to a very weak migratory flow, probably too weak to maintain stable pioneer populations. Second, the absence of suitable

ecological niches. Since the villages and cities were strikingly different in the early Neolithic, commensal niches were poorly represented and less stable in Western Europe than in the Near East until 1,000 BC. And third, the available commensal niches in Western Europe were favoured by the wood mouse until larger and more stable commensal environments developed. Since the house mouse is not able to live independent of human settlements, especially under non-Mediterranean climate, it was in an inferior position to the wood mouse until the human pressure on the environment increased. After that the house mouse was able to overwhelm the wood mouse in the commensal niches and colonized Europe in its entirety in a massive invasion. Thus, the populations of house mice in Western Europe represent a system which is about 9,000 generations (Karn et al. 2002) old and has its origin in the Near East.

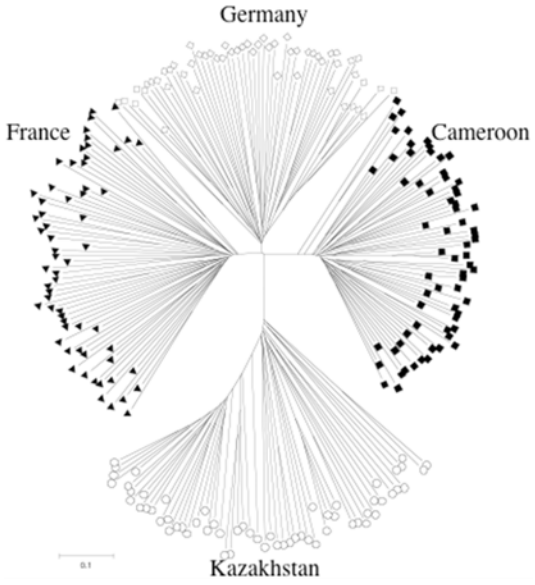


Figure 1.7 Allele sharing tree based on more than 200 microsatellites. The two western European populations, as well as samples from Cameroon can be clearly separated by this data. Samples from a Kazakhstan population, representative of the subspecies *M. m. musculus*, are clearly distinct from the three *M. m. domesticus* populations by a longer branch (Ihle et al. 2005).

The study here is based on samples from natural populations. In the major analysis I focus on two different populations of the Western Eurpean house mouse, one from the Cologne-Bonn-Area and the other one from the Massif Central. Additionally to these two focal samples three other populations were included in different parts of the study, one very young ‚*domesticus*' population from the recently colonized Subsaharan Africa and two populations of the subspecies *M. m. musculus* (one from Kazakhstan, a presumably old population, and one from the Czech Republic). In a previous study we have shown that the investigated populations are all

clearly distict, there is no significant gene flow between them (Figure 1.7, the population from the Czech Republic is not included) (Ihle et al. 2005). Thus, the samples allow the investigation of population specific selective sweeps. In the presence of geneflow between the analyzed poplations, adaptive events would potentially sweep over all populations which have a reasonable exchange.

To summarize, the investigated populations harbour several advantages for the study of evolutionary biology: based on fossil records the history is well documented, the populations are genetically distinct, their maximum diverence time is known, both samples represent derived populations, and a genome sequence, which is in its major parts directly deduced from mice of the same geographical region is avialable.

1.3 Aim of the study

In my study, I am investigating the molecular evolution in natural populations of the house mouse.

- In Chapter 2 I tested and excluded a correlation of recombination rate with microsatellite variability in natural populations. The results suggest that the high average mutation rate of microsatellites in mammals counterbalances the effects of long-range hitchhiking in the mouse genome. Thus, a long term influence of the recombination rate on microsatellite variability can be excluded in mice.

- For systematically genome wide scans for signatures of positive selection a high throughput routine is indispensable. Variability comparisons between natural populations require large amounts of polymorphism data, i.e. an enormous experimental effort has to be performed. In Chapter 3 I introduce a new high throughput routine for genome screens using variability measurements of microsatellites. The routine enables a quick and reliable selection of candidate loci for selective sweeps and therefore reduces the experimental effort to a minimum.

- Applying the described routine I systematically screened the genome for signatures of selective sweeps by comparing variability levels of microsatellites between natural populations. The results enabled me to calculate the minimal frequency with which positive selection events occur in natural populations. In a second step, I studied in detail the structure of the detected footprints of positive selection. The patterns observed in the different regions allowed me to draw conclusions on different basic parameters of selective sweep in general.

1.4 Literatur

Andolfatto, P. and Prezeworski, M. 2000. A Genome-Wide Departure From the Standard Neutral Model in Natural Populations of Drosophila. Genetics **156**: 257-268.

Andolfatto, P. and Prezeworski, M. 2001. Regions of lower crossing over habor more rare variants in African populations of *Drosophila melanogaster*. Genetics **158:** 657–665.

Bazykin, G.A., Kondrashov, F.A., Ogurtsov, A.Y., Sunyaev, S. and Kondrashov, A.S. 2004**.** Positive selection at sites of multiple amino acid replacements since rat-mouse divergence. Nature **429**: 558-562.

Begun, D.J. and Aquadro, C.F. 1993. African and North American populations of Drosophila melanogaster are very different at the DNA level. Nature **365**: 548-550.

Beisswanger, S., Stephan, W. and De Lorenzo, D. 2005. Evidence for a selective sweep in the *wapl* region of *Drosophila melanogaster*. Genetics Oct 3; [Epub ahead of print].

Betancourt, A.J. and Presgraves, D.C. 2002. Linkage limits the power of natural selection in *Drosophila*. Proceedings of the National Academy of Sciences, USA **99**: 13616–13620.

Birne, N. and Eyre-Walker, A. 2004. The genomic rate of adaptive amino acid substitutions in Drosophila. Molecular Biology and Evolution **21**(7): 1350-1360.

Boursot, P., Auffray, J.-C., Britton-Davidian, J. and Bonhomme, F. 1993.

The Evolution of house mice. Annu. Rev. Ecol.Syst. **24**: 119-52

Boursot, P., Din, W., Anand, R., Darviche, D., Dod, B., VonDeimling, F., Talwar, G.P. and Bonhomme, F. 1996. Origin and radiation of the house mouse: Mitochondrial DNA phylogeny. Journal of Evolutionary Biology **9** (4): 391-415.

Braverman, J.M., Hudson R.R., Kaplan, N.L., Langley, C.H. and Stephan, W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140**: 783–796.

Bush, E.C. and Bruce T. Lahn, B.T. 2005. Selective Constraint on Noncoding Regions of Hominid Genomes PLoS Computational Biology **1**(7): e73.

Cucchi, T., Vigne, J.-D. and Auffray, J.-C. 2005. First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. Biological Journal of the Linnean Society **84**: 429-446.

Ellgren, H., 2004 Microsatellites: simple sequences with complex evolution. Nature reviews **5**: 435-443.

Eyre-Walker, A., Keightley, P.D., Smith, N.G.C. and Gaffney, D. 2002 Quantifying the slightly deleterious mutation model of molecular evolution. Molecular Biolology and Evolution **19**(12): 2142–2149.

Fay, J. and Wu, C.-I. 2000. Hitchhiking under positive Darwinian selection. Genetics **155**: 1405-1413.

Guenet, J.L. and Bonhomme, **F.** 2003. Wild mice: an ever-increasing contribution to a popular mammalian model. Trends in Genetics **19**: 24-31.

Haddrill, P.R., Thornton, K.R., Charlesworth, B. and Andolfatto, P. 2005 Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Genome Research **15**: 790-799

Hardison, R.C. 2003. Comparative Genomics. PLoS Biology **1**(2): e58.

Harr, B., Kauer, M. and Schlötterer, C. 2002. Hitchhiking mapping – a population based fine scale mapping strategy for adaptive mutations in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences, USA **99**: 12949-12954.

Hartl, D.L. and Clark, G.C. 1997 Principals of population genetics Ed. 3. Sinauer Associates, Sunderland, Mass. 542 pp.

Ihle, S., Ravaoarimanana, I., Thomas, M. and Tautz, D. 2006 Tracing signatures of selective sweeps in natural populations of the house mouse. Molecular Biology and Evolution **23**(4): 790-797.

Karn, R.C, Orth, A., Bonhomme, F. and Boursot, P. 2002. The Complex History of a Gene Proposed to Participate in a Sexual Isolation Mechanism in House Mice. Molecular Biology and Evolution **19**: 462-471.

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

Kayser, M., Bauer, S. &Stoneking, M. 2003. A genome scan to detect candidate regions influenced by local natural selection in human populations. Molecular Biology and Evolution **20**: 893-900.

Keightley, P.D., Lercher, M.J. and Eyre-Walker, A. 2005. Evidence for widespread degradation of gene control regions in hominid genomes. PLOS **3**(2): e42.

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W. and Pääbo, S. 2004. A Neutral Model of Transcriptome Evolution. PLoS Biology **2**(5): e132.

Kohn, M.H., Pelz, H.-J. and Wayne, R.K. 2000. Natural selection mapping of the warfarin-resistance gene. Proceedings of the National Academy of Sciences, USA **97**(14): 7911-7915.

Lidicker, W.Z. 1976. Social Behaviour and Density Regulation in House Mice Living in Large Enclosures. Journal of Animal Ecology **45**(3): 677-697.

Maynard Smith, J. and Haigh, J. 1974. The hitch-hiking effect of a favorable gene. Genetic Research **23**: 23-35.

Mouse Genome Sequencing Consortium 2002. Initial sequencing and comparative analysis of the mouse genome. Nature **420**(6915): 520-62.

Nachman M.W. 1997. Patterns of DNA variability at X-linked loci in *Mus domesticus*. Genetics **147:** 1303–1316.

Nachman, M.W. 2001. Single nucleotide polymorphisms and recombination rate in humans Trends in Genetics **17**(9): 481-485.

Nachman, M.W, Bauer, V.L., Crowell, S.L. and Aquadro, C.F. 1998. DNA variability and recombination rates at X-linked loci in humans. Genetics **150**: 1133-1141.

Nachman, M.W. and Crowell, S.L. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics **156**: 297-304.

Payseur, B.A., Cutter A.D. and Nachman, M.W. 2002. Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. Molecular Biology and Evolution **19**: 1143-1153.

Pennings, P.S. and J. Hermisson, J. The title of the paper will be Soft Sweeps III - The signature of positive selection from recurrent mutation. in preparation.

Reimer, J. D. and Petras, M.L. 1967. Breeding Structure of the House Mouse, Mus musculus, in a Population Cage. Journal of Mammalogy **48**(1): 88-99.

Sakai, T., Kikkawa, Y., Miura, I., Inoue, T., Moriwaki, K., Shiroishi, T., Satta, Y., Takahata, N. and Yonekawa, H. 2005. Origins of mouse inbred strains deduced from whole-genome scanning by polymorphic microsatellite loci. Mammalian Genome **16**: 11-19.

Schlötterer, C. 2002. A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics **160:** 753–763.

Schlötterer, C. 2003. Hitchhiking mapping functional genomics from the population genetics perspective. Trends in Genetics **19**: 32-38.

Schlötterer, C, Ritter, R., Harr, B. and Brem, G. 1998. High mutation rate for a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. Molecular Biology and Evolution **15:** 1269–1247.

Schug, M.D., Hutter, C.M., Noor, M.A.F. and Aquadro, C.F. 1998. Mutation and evolution of microsatellites in *Drosophila melanogaster*. Genetica

102/103: 359–367.

Slatkin, M. 1995. Hitchhiking and associative overdominance at a microsatellite locus. Molecular Biology and Evolution **12:** 473–480.

Smith, N. G. and Eyre-Walker, A. 2002 Adaptive protein evolution in Drosophila. Nature **415**: 1022-1024.

Stearns and Hoekstra 2005. Evolution an introduction. Oxford University Press. 575 pp.

Stephan, W. and Langley, C.H. 1998. DNA polymorphism in lycopersicon and crossing-over per physical length. Genetics **150**: 1585–1593.

Storz, J.F., Payseur, B.A. and Nachman, M.W. 2004. Genome scan for DNA variability in humans reveal evidence for selective sweeps outside of Africa. Molecular Biology and Evolution **21**: 1800-1811.

Vigouroux, Y., Mullen, M., Hittinger, C.T., Houchings, L. and Schlulz, L. 2002 Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proceedings of the National Academy of Sciences, USA **99**: 9650-9655.

Wade, C.M., Kulbokas, E.J., Kirby, A.W., Zody, M.C., Mullikin, J.C., Lander, E.S., Lindblad-Toh, K. and Daly, M.J. 2002. The mosaic structure of variation in the laboratory mouse genome. Nature **420**: 574-578.

Wade, C.M. and Daly, M.J. 2005. Genetic variation in laboratory mice. Nature Genetics **37**(11): 1175-80.

Wiehe, T. 1998. The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. Theoretical Population Biology **53**: 272–283.

Wootton, J.C., Feng, X., Ferdig, M.T., Cooper, R.A, Mu, J., Baruch, D.I., Magill, A.J. and Su, X.Z. 2002. Genetic diverstiy and chloroquine selective sweeps in *Plasmodium falsiparum*. Nature **418**: 320-323.

Yonekawa, H., Moriwaki, K., Gotoh, O., Miyashita, N., Matsushima, Y., Shi, L.M., Cho, W.S., Zhen, X.L. and Tagashira, Y. 1988. Hybrid origin of Japanese mice Mus musculus molossinus: evidence from restriction analysis of mitochondrial DNA. Molecular Biology and Evolution **5**: 63–78.

2 Microsatellite variability in wild populations of the house mouse is not influenced by differences in chromosomal recombination rates

Meike Thomas, Sonja Ihle, Iary Ravaoarimanana, Susanne Krächter, Thomas Wiehe and Diethard Tautz

2.1 Abstract

Differences in recombination rates along the chromosomes can influence the evolution of neutral loci via hitchhiking effects. Generally, these effects should be stronger in regions of low recombination than in regions of high recombination. The detailed information on physical and genetic maps in the house mouse allows now to assess the correlation between neutral variability and recombination rates at given chromosomal regions. We have chosen 29 microsatellite loci from chromosomal regions which show differences in recombination rates and have tested their variability in samples from five wild populations of *Mus m. musculus* and *Mus m. domesticus*. Our results provide no evidence for a correlation between microsatellite variability and recombination rates. This suggests that the high average mutation rate of microsatellites in mammals counter balances the effects of long range hitchhiking in the mouse genome.

2.2 Introduction

Levels of polymorphism of neutral loci depend on mutation rate and effective population size. However, neutral loci can be linked to loci experiencing positive or negative selection (see Andolfatto 2001 and Schlötterer 2003 for recent reviews). In this case, parameters such as recombination rate and degree of selection on the linked locus can also influence the level of polymorphism at the neutral locus. For example, a local reduction in polymorphism can provide a signature of a selective sweep, i.e., the recent spread of beneficial mutation (Maynard Smith and Haigh 1974, Braverman et al. 1995, Slatkin 1995). Negative or background selection can also influence the

levels of polymorphism at linked loci (Hudson 1994, Charlesworth et al. 1995; Kim and Stephan 2000). Several cases of correlations between recombination rate and levels of nucleotide diversity are known, and this is widely interpreted as evidence that hitchhiking is common (Begun and Aquadro 1992, Nachman 1997, Stephan and Langley 1998, Nachman et al. 1998, Andolfatto and Prezeworski 2001, Betancourt and Presgraves 2002).

Knowledge of the correlations between recombination rate and levels of polymorphism in natural populations is essential for projects in which signatures of selective sweeps are used to identify genes experiencing positive selection (Schlötterer 2002). We have started a systematic screen for selective sweeps in house mouse using levels of polymorphism of mapped microsatellites (Ihle et al. in preparation). We are therefore particularly interested in whether differences in recombination rate are typically associated with differences in microsatellite polymorphism in mouse. Recombination rates are known to differ across the mouse genome (e.g. Nachman and Churchhill 1996; Froenicke et al. 2002; see Nachman 2002 for review) and detailed genomewide estimates of recombination rates are now available (Rowe et al. 2003, Jensen-Seaman et al. 2004). We have studied here the variability of 29 microstaellite loci in five wild populations and correlate these data with the associated recombination rates. We found no evidence that microsatellites located in regions with different recombination rates had significantly different levels of polymorphism. This result is consistent with findings in humans (Payseur and Nachman 2000).

	Total		*M. m. domesticus*		*M. m. musculus*		Germany		Kazakhstan		France		Cameroon		Czech Republic	
	expHet	VarRep	expHet	VarRep	expHet	VarRep	expHet	VarRep	expHet	VarRep	expHet	VarRep	expHet	VarRep	expHet	VarRep
D1Mit10	0.87	69.81	0.84	69.83	0.4	11.41	0.8	78.38	0.4	11.41	0.85	89.66	0.7	41.3	n.d.	n.d.
D1Mit64	0.84	26.56	0.81	30.3	0.77	11.28	0.78	24.13	0.77	11.28	0.62	17.04	0.38	18.83	n.d.	n.d.
D1Mit70	0.89	55.06	0.89	50.01	0.81	69.56	0.88	62.17	0.81	69.56	0.86	26.27	0.74	55.64	n.d.	n.d.
D1Mit136	0.87	37.7	0.82	16.02	0.89	98.09	0.88	21.9	0.89	98.09	0.83	25.54	0.23	1.43	n.d.	n.d.
D1Mit161	0.87	22.48	0.84	17.4	0.65	7.58	0.8	15	0.65	7.58	0.77	32.15	0.5	1.46	n.d.	n.d.
D1Mit187	0.88	60.34	0.83	17.53	0.76	11.86	0.9	20.29	0.76	11.86	0.79	20.29	0.65	11.84	n.d.	n.d.
D1Mit205	0.9	29.89	0.88	17.98	0.89	43.27	0.88	20.67	0.83	18.25	0.89	23.99	0.75	6.49	0.79	22.57
D1Mit404	0.83	22.56	0.7	6.7	0.9	25.73	0.75	7.63	0.85	23.55	0.6	3.04	0.58	7.8	0.88	11.22
D1Mit456	0.95	1914.94	0.94	377.82	0.79	127.61	0.92	90.6	0.76	16.81	0.96	852.18	0.73	31.98	0.78	241.47
D1Mit512	0.87	41.37	0.86	9.69	0.85	118.48	0.85	7.83	0.85	118.48	0.82	11.73	0.71	6.91	n.d.	n.d.
D2Mit252	0.94	289.71	0.9	154.72	0.91	95.14	0.92	113.31	0.84	82.47	0.78	151.19	0.76	116.93	0.87	14.97
D2Mit383	0.8	6.39	0.79	6.46	0.67	1.64	0.66	2.79	0.44	0.93	0.61	2.09	0.65	11.53	0.65	1.28
D5Mit149	0.91	31.8	0.89	23.05	0.87	26.04	0.86	32.45	0.8	16.87	0.86	26.38	0.63	8.65	0.87	26.22
D5Mit310	0.91	27.05	0.86	10.86	0.87	26.51	0.81	6.51	0.78	15.52	0.87	17.71	0.79	6.68	0.85	26.29
D6Mit180	0.78	51.14	0.52	59.99	0.82	32.63	0.49	63.46	0.73	48.32	0.7	60.51	0.25	41.11	0.71	3.15
D6Mit309	0.88	71.3	0.8	19.17	0.76	38.2	0.86	20.79	0.71	19.2	0.81	20.05	0.3	1.12	0.76	47.47
D7Mit158	0.91	72.47	0.84	47.85	0.87	35.98	0.89	72.25	0.84	20.08	0.68	12.26	0.71	28.22	0.89	50.74
D9Mit61	0.89	38.57	0.9	42.41	0.5	3.91	0.89	114.22	0.5	3.91	0.85	14.14	0.67	29.65	n.d.	n.d.
D9Mit165	0.86	109.18	0.77	10.65	0.85	134.7	0.79	18.17	0.85	134.7	0.73	6.83	0.61	3.41	n.d.	n.d.
D9Mit206	0.89	28.46	0.81	10.22	0.88	56.29	0.83	11.17	0.88	56.29	0.33	6.17	0.73	10.19	n.d.	n.d.
D9Mit223	0.8	10.4	0.74	9.43	0.79	10.04	0.69	7.21	0.82	15.55	0.74	14.19	0.58	1.75	0.7	3.97
D9Mit251	0.92	90.22	0.91	85.3	0.79	28.85	0.86	84.82	0.79	28.85	0.88	30.03	0.72	110.12	n.d.	n.d.
D9Mit330	0.88	66.52	0.79	63.71	0.88	17.36	0.73	70.07	0.82	16.68	0.78	46.98	0.68	75.2	0.84	13.96
D11Mit194	0.92	41.8	0.9	22.07	0.84	32.91	0.9	19.36	0.77	14.44	0.91	23.69	0.65	13.43	0.89	52.43
D11Mit319	0.92	122.62	0.87	126.17	0.89	42.36	0.88	119.18	0.85	34.04	0.86	111.41	0.76	128.88	0.8	50
D13Mit61	0.93	65.33	0.91	32.02	0.91	63.89	0.89	29.39	0.86	29.01	0.88	25.59	0.73	21.96	0.81	98.06
D14Mit67	0.81	99.48	0.56	0.82	0.84	22.42	0.32	0.52	0.75	18.7	0.48	0.51	0.46	0.58	0.86	26.56
D14Mit224	0.91	101.73	0.87	18.14	0.72	13.82	0.84	19.97	0.7	6.83	0.84	19.56	0.77	10.86	0.64	21.3
D19Mit39	0.9	80.94	0.79	37.99	0.95	123.33	0.86	42.39	0.9	72.27	0.79	34.89	0.55	25.42	0.9	26.2
Mean	0.88	127.1	0.82	48.08	0.8	45.89	0.81	41.26	0.77	35.23	0.77	59.52	0.62	28.6	0.8	40.99
standard error	0.01	64.61	0.02	13.54	0.02	7.56	0.02	6.84	0.02	6.57	0.03	28.99	0.03	6.66	0.02	13.04

Table 2.1 Two polymorphism parameters of interest: expected heterozygosity and variance in repeats, shown for all loci for all samples and populations, and summarized for the two subspecies *M. m. musculus* and *M. m. domesticus*.

2.3 Methods

Microstaellite analysis: The following loci were chosen from the mouse database of the Whitehead Institute (http://www-genome.wi.mit.edu/cgi-bin/mouse/index) on the basis of a visual inspection of the recombination maps provided in Rowe et al. (2003): D1Mit64, D1Mit70, D1Mit161, D1Mit10, D1Mit136, D1Mit187, D1Mit205, D1Mit456, D1Mit404, D1Mit512, D2Mit383, D2Mit525, D5Mit149, D5Mit310, D6Mit180, D6Mit309, D7Mit158, D9Mit251, D9Mit61, D9Mit223, D9Mit206, D9Mit330, D9Mit165, D11Mit319, D11Mit194, D13Mit61, D14Mit67, D14Mit224, D19Mit3. Using the recommended primers, the loci were amplified by PCR in 10 µl volumes applying 34 cycles of 45 sec. at 94 °C, 45 sec. at 45 °C, 1:30 min at 72 °C with 9 ng DNA-template, 4 µM primers, 1.5 µM $MgCl_2$, 0.6 µM dNTPs, and 0.4U of Taq-Polymerase. Annealing temperatures were estimated to be optimal between 53 and 62 °C found by applying gradient PCR. The PCR products were diluted in water (diluent volume ranged from 100 to 200 µl). 1 µl of each dilution was run on a 96 capillary sequencer (MegaBACE, Amersham Biosciences) for genotyping. For each run we pooled one FAM, one HEX, and one TET-labelled PCR-product, whereby we always took care that fragments labelled with HEX and TET had a different sizes to avoid problems of interference. We added 0.1 µl ROX 400-Size-Standard and 4.9 µl H_2O per sample as an internal lane standard.

Recombination rates: An estimate of local recombination rate for each locus was obtained by choosing a 10 Mb window centered around the marker. Each of these windows included at least five mapped markers. The genetic positions of these markers were taken from the Whitehead Institute Mouse Genome site (http://www.broad.mit.edu/cgi-bin/mouse/index) and the physical positions were taken from the ensemble genome server (http://www.ensembl.org/Mus_musculus/). The genetic and physical positions were plotted and the slope of the regression-line was taken as the local recombination rate. In addition to our own estimates, we used also the estimates from Jensen-Seaman et al. (2004) for 5Mb and 10Mb nonoverlapping windows.

Statistics: The program Genetic Profiler (Amersham Biosciences, Verson 2.0) was used for the analysis of genotyping data. The program MS-Analyzer (Dieringer and Schlötterer 2003) was used to estimate the expected heterozygosity and the variance of repeats. All parameters of interest were calculated for each population and for the total number of samples. To avoid any interspecific problems this analysis was also done separately for each subspecies. The data were tested by the Kolmogorov-Smirnov-test for normal distribution.

2.4 Results

To obtain polymorphism data from true wild type populations we have sampled mice in five different regions. *M. m. domesticus* populations came from Germany (near Cologne), from France (near Sévérac-le-Château, Massif Central) and from Cameroon (near Kumba). *M. m. musculus* populations came from the Czech Republic (near Námest nad Oslavou) and from Kazakhstan (near Almaty). The sampling scheme took care to avoid related animals which could potentially originate from a single nest, by using only one individual from a particular sampling site. Thus, we consider our samples as representative for the respective local populations.

29 microsatellite loci from ten autosomal chromosomes were chosen to represent different recombination classes. These range from 0.137-1.248 cM/Mb (Figure 2.1). The calculation of these recombination classes is based on using the chosen locus as a center of a 10 Mb window within which the local recombination rate was estimated (see methods). Our rate estimates are generally in agreement with those which were recently provided by Jensen-Seaman et al. (2004) for 10Mb and 5Mb nonoverlapping windows across the genome. Still, since differences exist between the different estimates (Figure 2.1), we have used all three measures for the further analysis.

The 29 loci were typed in the populations for 48 unrelated individuals each. The average values of expected heterozygosity and variance in allele size are depicted in Table 2.1. These measures of polymorphism are all very similar in the populations, with the exception of the Cameroon *M. m. domesticus* population, which showed significantly lower values, possibly due to a bottleneck effect. The large values of average variance in allele size are due mainly to locus D1Mit456, which had 2 classes of allele sizes separated by about 100 bp. We suggest that this bimodality is due to an

insertion in the flanking sequences of the larger group of alleles. A second locus (D14Mit67) showed signs of a selective sweep in the *M. m. domesticus* populations (further details will be presented in Ihle et al., in preparation). These latter two loci were therefore omitted in the further analysis (Table 2.2), although their inclusion would not change the inferences.

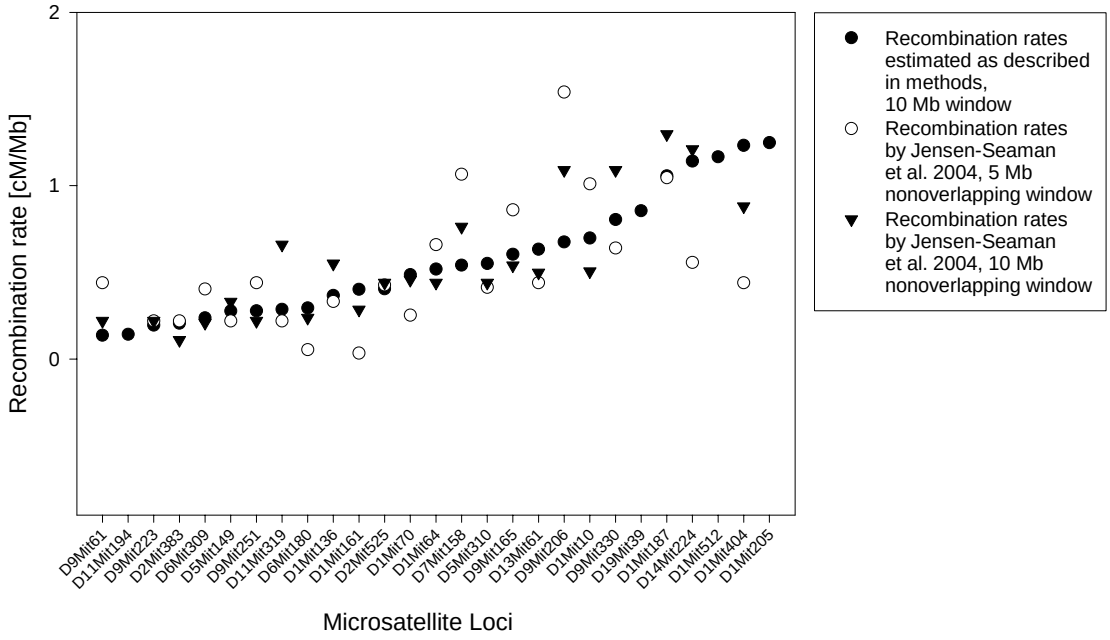


Figure 2.1 Depiction of the loci analysed in relation to their local recombination rate. Three such measures are plotted for each locus. The loci are sorted according to the measure that we have calculated on the basis of a 10Mb window centered around the locus.

Table 2.2 Descriptive statistics of two polymorphism parameters of interest: expected heterozygosity and variance in repeats, without the anomalous loci D1Mit456 and D14Mit67. These values were used for the correlation statistics in Table 3.

Parameter	Statistic	TOTAL	*M. m. domersticus*	*M. m. musculus*	Germany	France	Cameroon	Kazhak-stan	Czech Republic
Expected	N	27	27	27	27	27	27	27	16
Hetero-zygosity	Mean	0.88	0.83	0.80	0.82	0.78	0.62	0.77	0.80
	Standard error of mean	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.02
Variance in repeats	Mean	61.90	37.62	43.74	40.95	32.35	29.51	36.52	29.36
	Standard error of mean	10.50	7.06	7.46	6.95	6.59	7.08	7.00	6.29

Table 2.2 provides the correlation statistics for the two measures of polymorphism with the three measures of recombination rates. None of the correlations is significant. To correct for the fact that loci with different number of repeats can have different degrees of polymorphism (Goldstein and Clark 1995), we

have also calculated this statistic with normalized values, but find still no significant correlation (Table 2.3).

Table 2.3 Correlation between measure of polymorphism and recombination rate. Three different measures of recombination rate were used. RR1 represents our own calculation with the locus centered in a 10Mb window, RR2 and RR3 are the measures taken from Jensen-Seaman et al. (2004) for 5Mb and 10Mb windows. Note that four of our loci were not included in their estimates, i.e. they were omitted from the analysis.

Samples	Parameter	Statistical test	RR1 N=27	RR2 N=23	RR3 N=23
all populations	Expected Heterozygosity	Pearson's Correlation Coefficient	0.1	0.2	0.3
		p-value (2-tailed)	*0.74*	*0.32*	*0.26*
	normalized values	Pearson's Correlation Coefficient	0.1	-0.1	-0.2
		p-value (2-tailed)	*0.71*	*0.64*	*0.30*
	Variance in Repeats	Pearson's Correlation Coefficient	0.0	0.1	0.1
		p-value (2-tailed)	*0.89*	*0.81*	*0.65*
	normalized values	Pearson's Correlation Coefficient	0.1	0.2	0.1
		p-value (2-tailed)	*0.79*	*0.39*	*0.62*
M. m. domesticus	Expected Heterozygosity	Pearson's Correlation Coefficient	0.0	0.1	0.1
		p-value (2-tailed)	*0.94*	*0.57*	*0.83*
	normalized	Pearson's Correlation Coefficient	0.2	0.2	0.1
		p-value (2-tailed)	*0.42*	*0.32*	*0.79*
	Variance in Repeats	Pearson's Correlation Coefficient	-0.3	-0.1	-0.1
		p-value (2-tailed)	*0.20*	*0.60*	*0.71*
	normalized	Pearson's Correlation Coefficient	-0.1	0.2	0.0
		p-value (2-tailed)	*0.56*	*0.49*	*0.84*
M. m. musculus	Expected Heterozygosity	Pearson's Correlation Coefficient	0.2	0.0	0.3
		p-value (2-tailed)	*0.30*	*0.86*	*0.25*
	normalized	Pearson's Correlation Coefficient	-0.1	-0.2	-0.2
		p-value (2-tailed)	*0.69*	*0.47*	*0.38*
	Variance in Repeats	Pearson's Correlation Coefficient	0.2	0.1	0.0
		p-value (2-tailed)	*0.33*	*0.58*	*0.93*
	normalized	Pearson's Correlation Coefficient	0.3	0.1	0.0
		p-value (2-tailed)	*0.22*	*0.78*	*0.98*

2.5 Discussion

The population sample that was used for this study is unique with respect to representing true wildtype populations of the two known subspecies of the house mouse. Thus, the polymorphism estimates are likely to be very reliable and not influenced by population specific effects. The only exception might be the sample from Cameroon, which represents a population that has only relatively recently colonized this area. Still, even population specific analysis of correlation between polymorphism and recombination rate did not yield significant results in any of the

comparisons (not shown). On the other hand, our result depends on an accurate estimate of local recombination rates. The choice of relatively large windows within which the recombination rates are averaged can potentially confound the presence of recombination hotspots. A simple visual inspection of the recombination maps suggests that such hotspots exist (Figure 2.2). Thus, it is not surprising to see differences in rate estimates, depending on which windows are chosen (Figure 2.1). To compensate for this, we have alternatively tried to classify our markers into high, medium and low recombination groups, by comparing their direct neighborhoods in the map of Rowe et al. (2003). But even such a subjective classification scheme did not yield a correlation between recombination class and polymorphism (not shown). Hence, in spite of the remaining uncertainty of estimating exact local recombination rate, it would appear that our results are robust, at least when taking an average over many loci.

A lack of correlation between local recombination rate and microsatellite variability was also found by Payseur and Nachmann (2000) for humans, in agreement with our mouse data. In contrast to the mammalian studies, Schug et al. (1998) find a positive correlation between microsatellite variability and local recombination rate in *Drosophila melanogaster*. However, microsatellites of *D. melanogaster* are short (on average up to 12 repeats) and this influences their mutation rate. Schlötterer et al. (1998) estimate an average mutation rate of 6.3×10^{-6} per locus per generation whereas for humans, mice, rats, and pigs microsatellite mutation rates were estimated in a range of 10^{-2} to 10^{-5} (see Schug et al. 1997). Slatkin (1995) and Wiehe (1998) have suggested that higher mutation rates will counter balance the effects of selective sweeps in low recombining regions. Thus, high mutation rate potentially explains the lack of association between microsatellite polymorphism and recombination rate in mammals.

Nucleotide substitution rates are much lower than microsatellite mutation rates. The mutation rate per nucleotide in humans is estimated to be 2.5×10^{-8} (Nachman and Crowell 2000), and in mice 2.1×10^{-8} (Nachman 1997). Accordingly, a dependence of nucleotide diversity on recombination rate is evident for both mice and humans (Nachman 1997, Nachman et al. 1998), in line with background selection as well as hitchhiking models. On the other hand, Lercher and Hurst (2002) have suggested that the correlation between nucleotide polymorhism and recombination

rates in humans can be explained by the mutagenic effects of recombination and without invoking the action of selection.

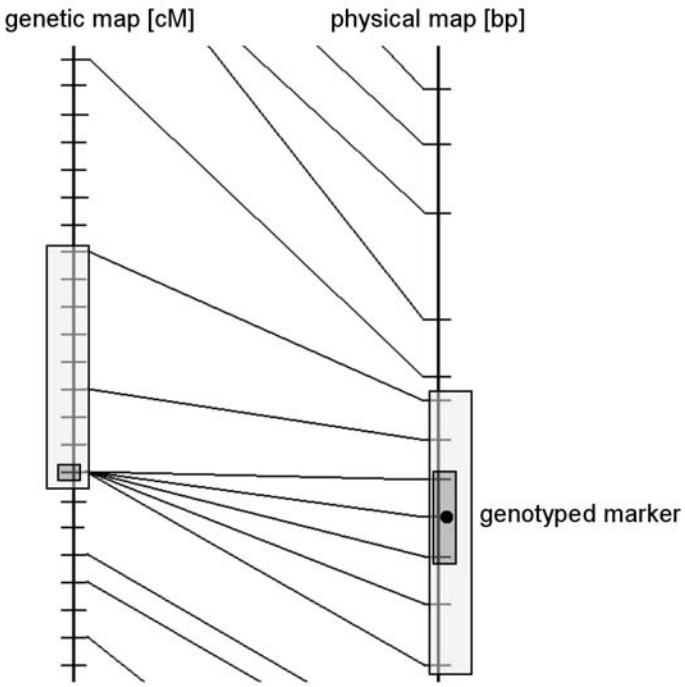


Figure 2.2 Comparison of genetic and physical map to show the problem with averaging recombination rates. Two different window sizes are centered around a marker (shown here for D14Mit224) to estimate recombination rates (light grey about 10 Mb window, dark grey about 5 Mb window). Recombination hotspots can lead to large differences in recombination rate estimates, in dependence of the positioning of the windows.

Furthermore, Hellman et al. (2003) found that not only nucleotide diversity but also divergence rates between human and primates are correlated with recombination rates and that the correlation between diversity and recombination was absent after correcting for divergence. However, recombination rates in humans are highly non-uniform across the genome (McVean et al. 2004) and findings in human may not easily apply to rodents (Hellman et al. 2003). In fact, Jensen-Seaman et al. (2004) reported a negative correlation between diversity and mutation rates in mouse, which may be explained by other genomic properties, such as a different GC-content variation in mouse than in human. It seems therefore possible that also microsatellite mutation rates are influenced by additional genomic factors and that our finding of a lack of correlation with recombination rate is confounded by these additional factors.

It will be the aim of future studies to further disentangle the evolutionary causes which shape microsatellite variability in mice. Still, the current results are compatible with the background selection or hitchking scenarios that take into account mutation rates at microsatellites. However, in mammals microsatellites may be a suitable marker only to detect exceptionally strong selective sweeps, such as the

spread of resistance mutations (Kohn et al. 2000). Theoretical models need to be refined in order to integrate the predictive power of different types of markers, for instance SNPs together with microsatellites, and different measures of variability, for instance heterozygosity (Schloetterer, 2003) together with linkage disequilibrium (Kohn et al. 2003), to get a clearer picture of the regions of the mouse genome under Darwinian selection.

2.6 Acknowledgements:

We are particularly greatful to Francois Bonhomme and Pierre Boursot for their support in initiating this study, as well as Barbara Dod, Louis Rouquet, Paul Pfander, Jaroslav Pialek, Barbora Bimova, Eva Bozikova, Solo Robespierre and Uli Schliewen for their help with obtaining the wild type samples. We thank also Tina Harr and John Dallas for valuable comments on the manuscript. This study is supported by a grant of the Volkswagen Stiftung.

2.7 Literature

Andolfatto, P. 2001. Adaptive hitchhiking effects on genome variability. Current Opinions in Genetics and Development **11**: 635-641.

Andolfatto, P. and Prezeworski, M. 2001. Regions of lower crossing over habor more rare variants in African populations of *Drosophila melanogaster*. Genetics **158**: 657-665.

Begun, D.J. and Aquadro, C.F. 1992. Levels of naturally occurring DNA polymorphism correlated with recombination rates in *D. melanogaster*. Nature **356**: 519-520.

Betancourt, A.J. and Presgraves, D.C. 2002. Linkage limits the power of natural selection in *Drosophila*. Proceedings of the National Academy of Sciences **99**(21): 13616-20.

Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langlaey, C.H. and Stephan, W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphism. Genetics **140**: 783-796.

Charlesworth, D., Charlesworth, B. and Morgan, M.T. 1995. The pattern of neutral molecular variation under the background selection model. Genetics **141**: 1619-1632.

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

Froenicke, L., Anderson, L.K., Wienberg, J. and Ashley, T. 2002. Male

mouse recombination maps for each autosome identified by chromosome painting. The American Journal of Human Genetics **71**: 1353-1368.

Goldstein, D.B. and Clark, A.G. 1995. Microsatellite variation in North American populations of Drosophila melanogaster. Nucleic Acids Research **23**: 3882-3006.

Hellmann, I., Ebersberger, I., Ptak, S.E., Pääbo, S. and Przeworski, M. 2003. A neutral explanation for the correlation of diversity with recombination rates in humans. The American Journal of Human Genetics **72**: 1527-1535.

Hudson, R.R. 1994. How can the low level of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? Proc. Natl. Acad. Sci. USA **91**: 6815-6818.

Jensen-Seaman, M.I., Furey, T.S., Payseur, B.A., Lu, Y., Roskin, K.M., Chen, C.F., Thomas, M.A., Haussler, D. and Jacob, H.J. 2004. Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. **14**: 528-538.

Kohn, M.H., Pelz, H.J. and Wayne, R. 2000. Natural selection mapping of the warfarin-resistance gene. Proc. Natl. Acad. Sci. USA 97: 7911-7915

Kim, Y. and Stephan, W. 2000. Joint Effects of genetic hitchhiking and background selection on neutral variation. Genetics **155**: 1415-1427.

Lercher, M.J. and Hurst, L.D. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. Trends Genet. **18**: 337-340.

Maynard-Smith, J. and Haigh, J. 1974. The hitch-hiking effect of a favorable

gene. Genetic Research **23**: 23-35.

McVean, G.A.T., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R. and Donnelly, P. 2004. The fine-structure of recombination rate variation in the human genome. Science **304**: 581-584.

Nachman, M.W. 1997. Patterns of DNA variability at X-linked loci in *Mus domesticus*. Genetics **147**: 1303-1316.

Nachman, M.W. 2002. Variation in recombination rate across the genome: evidence and implications. Current Opinions in Genetics and Development **12**: 657-663.

Nachman, M.W. and Churchhill, G.A. 1996. Heterogenity in rates of recombination across the mouse genome. Genetics **142**: 537-548.

Nachman, M.W., Bauer, V.L., Corwell, S.L. and Aquadro, C.F. 1998. DNA variability and recombination rates at X-linked loci in humans. Genetics **150**: 1133-1141.

Nachman, M.W. and Crowell, S.L. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics **156**: 297-304.

Payseur, B.A. and Nachman, M.W. 2000. Microsatellite variation and recombination rate in the human genome. Genetics **156**: 1285-1298.

Rowe, L.B., Barter, M.E., Kelmenson, J.A. and Eppig, J.T. 2003. The comprehensive mouse radiation hybrid map densely cross-referenced to the recombination man: a tool to support the sequence assemblies. Genome Research **13**: 122-133.

Schlötterer, C. 2002. A Microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics **160**: 753-763.

Schlötterer, C. 2003. Hitchhiking mapping—functional genomics from the population genetics perspective. Trends Genet. **19**: 32-38.

Schlötterer, C., Ritter, R., Harr, B. and Brem, G. 1998. High mutation rate for a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. Molecular Biology and Evolution **15**: 1269-1247.

Schug, M.D., Mackay, T.F.C. and Aquadro, C.F. 1997. Low mutation rates of microsatellite loci in *Drosophila melanogaster*. Nature Genetics **15**: 99-102.

Schug, M.D., Hutter, C.M., Noor, M.A.F. and Aquadro, C.F. 1998. Mutation and evolution of microsatellites in *Drosophila melanogaster*. Genetica **102/103**: 359-367.

Slatkin, M. 1995. Hitchhiking and associative overdominance at a microsatellite locus. Molecular Biology and Evolution **12**: 473-480.

Stephan, W. and Langley, C.H. 1998. DNA polymorphism in lycopersicon and crossing-over per physical length. Genetics **150**: 1585-1593.

Wiehe, T. 1998. The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. Theoretical Population Biology **53**: 272-283.

3 A high-throughput routine to detect signatures of selective sweeps in genome screens using microsatellites

Meike Thomas, Friedrich Möller, Thomas Wiehe and Diethard Tautz

Abstract

We have evaluated a pooling approach that can reduce the number of genotypes in a screen for selective sweeps by more than an order of magnitude. We show that the complex peak pattern that results from pooling of all samples from a given population is a faithful reflection of the composite pattern of the individual alleles, although with an under-representation of the larger alleles. Candidates for selective sweeps can be easily identified by visual inspection of the pool patterns. We have also implemented a software pipeline, which can automatically find suitable microsatellite loci in the vicinity of annotated genes.

3.1 Introduction

Identification of selective sweeps in natural populations has been proposed as an approach to identify genes involved in local adaptations (Schlötterer 2003). Positive selection can lead to the fixation of a favourable mutation in a population. This leads also to a loss of variability in the flanking region, due to hitchhiking (Maynard Smith and Haigh 1974). Because of the high density of microsatellite loci in the eukaryotic genomes it should be possible to trace selective sweeps by systematically scanning for population specific loss of variability at individual loci (Schlötterer 2003). However, since higher eukaryotic genomes may contain about 40,000 selectable loci and the detection of polymorphic variants requires testing of multiple individuals (at least 20) for several populations, a complete genome scan would require millions of genotypes to be determined. We present here a pooling strategy that allows to reduce the number of genotyping reactions significantly and which has been shown to be possible in principle (Pacek et al. 1993).

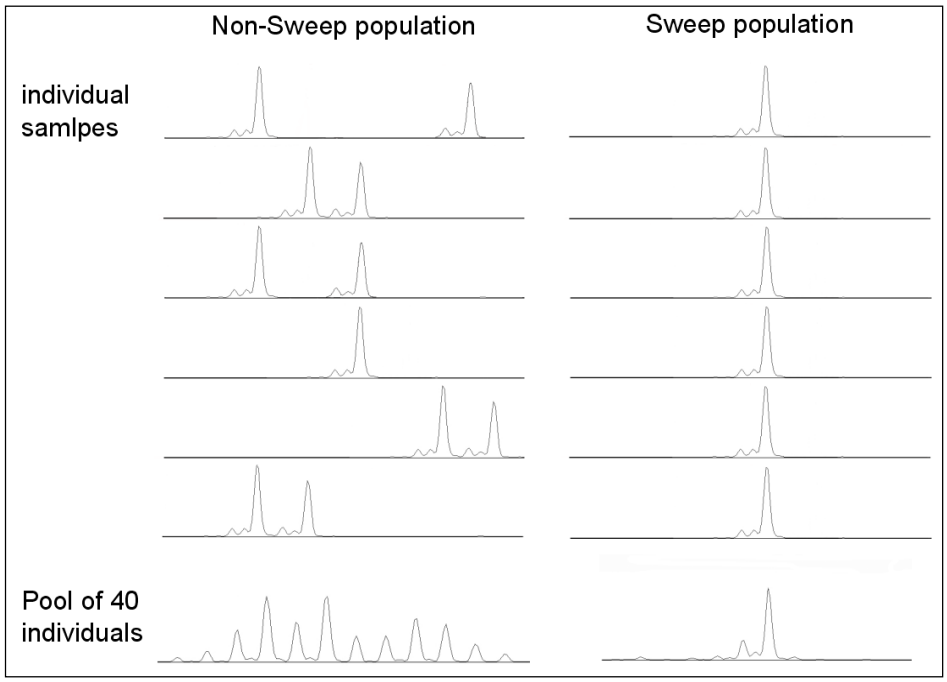


Figure 3.1 In non-sweep populations the pools of individual samples, which are polymorphic, exhibit a complex peak pattern, whereas a sweep locus, which shows the same allele in almost all individuals will show a simple pool pattern.

The basic idea is that DNA samples from all individuals of a given population can be pooled in equal amounts and only one PCR reaction is performed with primers flanking a microsatellite locus. This will result in a complex pattern of peaks in those cases where the locus is polymorphic, but in a relatively simple peak pattern in cases where polymorphism was lost, i.e. in the regions that are candidates for selective sweeps. Such loci can then be re-typed from individuals to confirm the sweep signature and to calculate exact allele frequencies to be used in appropriate statistics. In the following, we describe this approach for a genome screen based on approximately 1,000 microsatellite loci in five different populations of the house mouse (*Mus musculus*).

3.2 The Method

Sample collection schemes and DNA extraction were followed the scheme described in Ihle et al. (2006). DNA of single individuals was normalized to 10 ng/µl and combined in population specific pools, each consisting of 40 samples. Primers were purchased in 96 well plate format (primer sequences see supplementary data)

from Sigma Aldrich. Forward primers were labeled with FAM dye at the 5' end. PCR was performed using a multiplex kit (Quiagen Cat.No. 206143). All reactions were carried out in 10 µl volumes using 30 ng of pooled DNA template and following the protocol of the supplier of the kit. Amplification was started with a 15' melting step at 95°C, followed by 40 cycles of 30'' melting at 94°C, standard annealing temperature for all primers at 60 °C for 1'30'', and elongation at 72°C for 1', followed by a final elongation step at 72 °C for 30'. PCR products were diluted 1:20 in water and 1 µl of this dilution was added to 0.1 µl GeneScan™ 500 LIZ™ Size Standard (Applied Biosystems, Part Number 4322682) in 10 µl HIDI Formamide (Applied Biosystems, Part Number 4311320). Fragments were run on a 48-capillary 3730 DNA sequencer (Applied Biosystems).

The fragment patterns from the pools were displayed with the GeneMapper® Software v3.5 (Applied Bioscience, Part Number 4346647) and then analysed by eye. All output files were inspected by pairwise comparison between populations. Candidate loci were defined as those showing a rather simple pattern of peaks in one population but a complex one in the others. Figure 3.2 shows an example from three loci typed in the five populations, two belonging to the subspecies *Mus musculus musculus* (sampled in Kazakhstan and in the Czech Republic) and three to *M. m. domesticus* (sampled in Germany, France and Cameroon) (Ihle et al. 2006). Because each locus shows characteristic slippage patterns, we have also always included a single individual from each subspecies for comparison.

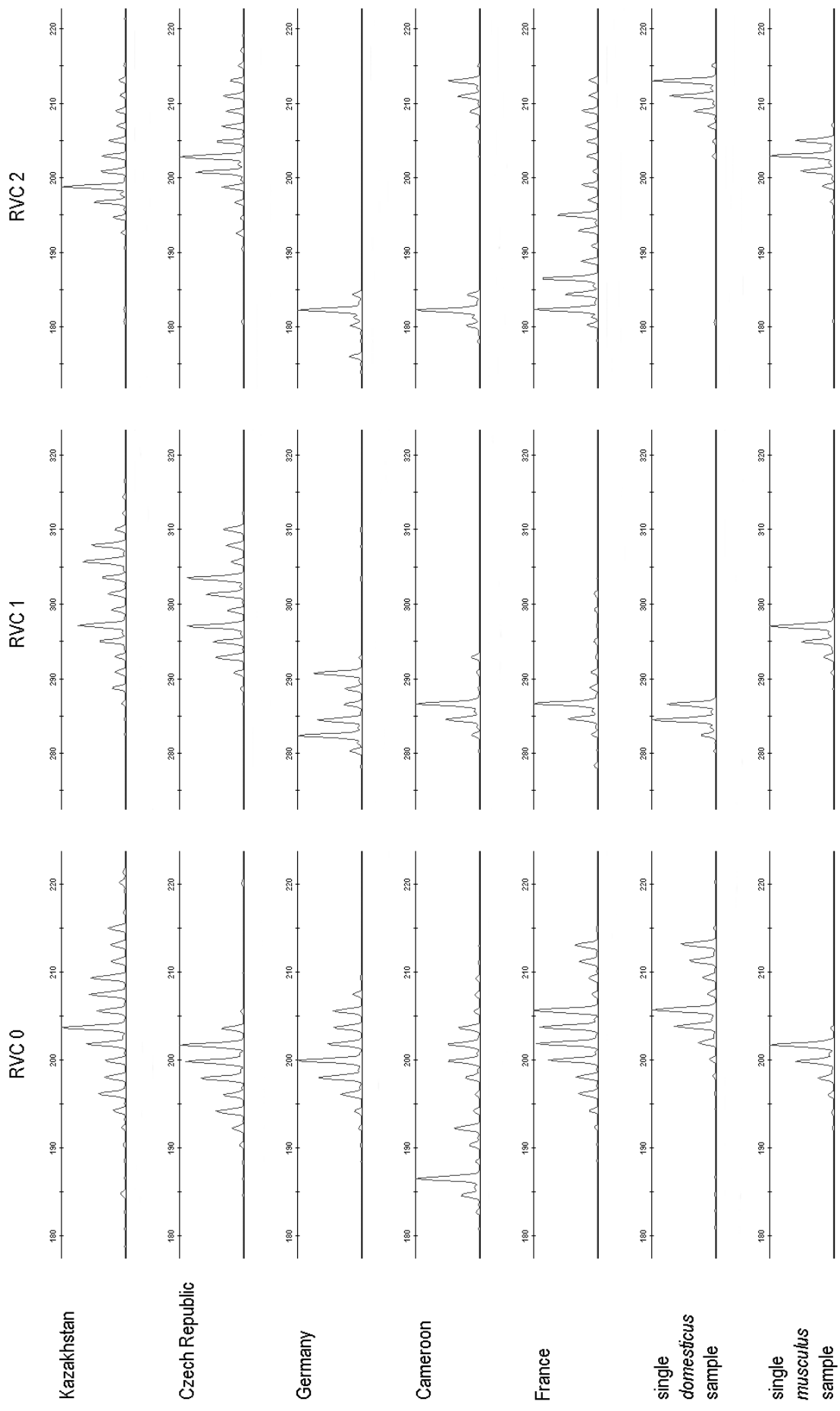

RVC 0
RVC 1
RVC 2
Kazakhstan
Czech Republic
Germany
Cameroon
France
single *domesticus* sample
single *musculus* sample

Figure 3.2 Examples of three output files form the pooling approach each representing one of the three different Reduced Variability Classes. The loci were amplified from all five mouse populations plus a single animal from each subspecies. The classification of the loci in RVC1 and 2 are based on a comparison of the German and the French population. In the example for RVC1 a weak reduction in variability is present in France, but also in Cameroon. The locus classified in RVC2, exhibits almost no variability in Germany compared to a variable French population. With respect to the Cameron population, it would have been classified as RVC1.

3.3 Data Analysis

To classify the differences in reduction in variability, we used a ranking scheme of 'Reduced Variability Classes' (RVC) 0 to 2, where RVC0 contains all loci that show no obvious difference in variability between the populations, RVC1 represents possible cases of reduction in variability and RVC2 represents clear cases. This classification was done by eye, since the complexity of the patterns encountered for the different loci did not allow us to find a simple solution for an automatic quantification, although this may become possible in the future. Figure 3.2 includes examples from each of these ranking classes.

Candidate loci from RVC2 were then amplified from individuals. Figure 3.3 shows three such examples. It is evident that the allele frequencies determined from the individual typings reflect rather well the patterns seen in the pools. However, in several cases we found that larger alleles tend to be under-represented in the pools, when compared to the individual typings. Comparison of the patterns in Figure 3.3 E and F shows that the large alleles in the German population appear only as minor peaks in the pool, but with significant frequencies in the individual typings. This effect has therefore to be kept in mind when evaluating the pool patterns. However, we found no case where an allele appeared with a significant frequency in the individual typings, which was not also at least present as a small peak in the pool pattern. Thus, the pool patterns are fairly reliable indicators of the allele spectrum found in a population.

The results from the individual typings can then be directly used for statistical analysis, for example Schlötterer´s lnRΘ statistic (Schlötterer 2002, Kauer et al. 2003), provided a reference set of randomly chosen loci is available. Using the reference set described in Ihle et al. (2006), we found that the loci in Figure 3.3 A, B and Figure 3.3 C, D show highly significant signatures of selective sweeps, while the locus in Figure 3.3 E, F is not significant (details on this see Chapter 4).

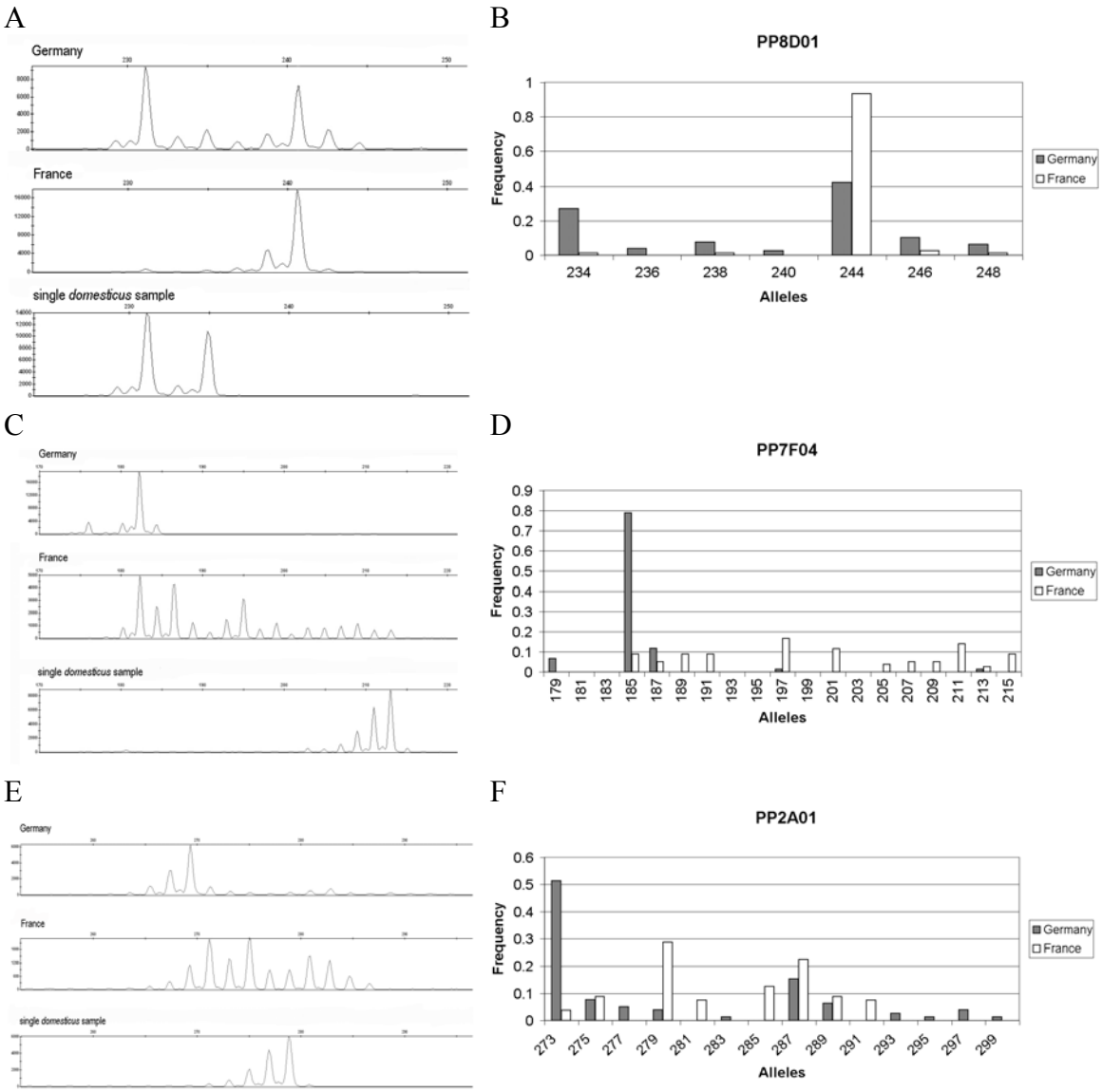


Figure 3.3 Pattern of the pooled samples from three candidate sweep loci (A, C, E) and the corresponding allele frequencies estimated form genotyping single individuals for the respective loci (B, D, F).

3.4 Software Pipeline

Given that the pooling approach allows an efficient screening of large numbers of loci, it is also important to have an automatic routine for selecting the microsatellite loci. Because sweep regions tend to be relatively small (Beisswanger et al. 2005; see Chapter 4) it is advisable to use microsatellite loci from the vicinity of genes for a systematic screen.

To find such loci, we implemented a software pipeline to identify all occurrences of dinucleotide repeats in the upstream region of annotated genes and to

list the results in a comprehensive tabular format. First, the pipeline extracts all upstream regions of annotated genes out of a genbank flatfile. These are then checked against the avialable EST data, to confirm the annotation. The upstream regions of these annotated regions are then extracted, by default 20kb upstream of the putative translation start. In a seed and extend strategy, our pipeline then calls the program ClustDB (Kleffe 2004: http://www.charite.de/molbiol/bioinf/bioinf/Computer programme/ClustDB/clustdb.html), an implementation of a suffix array algorithm. ClustDB finds all repeats of given minimum length in a set of sequences or between two sets of sequences. In our application, we compare the extracted upstream regions with a set of sequences composed of hexamers of all possible dinucleotide combinations. Applying ClustDB to compare these two sequence sets reports all left-maximal occurrences of all dinucleotide repeats on the genomic sequence. In a consecutive step we then (right-)extend the found seeds to their maximum length, and reject those repeats which remain below a given minimum length. Using this positional information, and the genome annotation, which is available from Genbank, the upstream and downstream flanking genes are extracted and returned a tabular output file. The result file contains size range and type of every microsatellite and its physical distance to adjacent upstream and downstream genes. To simplify primer design and further analysis not only the microsatellite sequence itself but also the flanking 600 bp are reported in the output file. This program was used to identify 1,000 loci, distributed across all chromosomes for which primers were designed.

3.5 Conclusions

By applying the pooling strategy we were able to reduce the number of required PCRs from approximately 200,000 to approximately 7,000 for a genome screen involving the previously selected 1,000 loci in five populations, each represented by 40 individuals. In particular, the time consuming part in microsatellite genotyping, namely the allele calling from the raw data, was also reduced to a minimum, because the preselection of candidate loci for signatures of positive selection is based on a simple visual inspection of pool patterns. With our approach it seems therefore feasible to eventually screen all annotated genes of the mouse genome for signatures of selective sweeps.

3.6 Acknowledgements

We thank Sonja Ihle and Susanne Kipp for providing the DNA samples used in this study. The work was funded by the Volkswagen Stiftung.

3.7 Literature

Beisswanger, S., Stephan, W. and De Lorenzo, D. 2006. Evidence for a selective sweep in the wapl region of *Drosophila melanogaster*. Genetics **172**: 265-274.

Ihle, S., Ravaoarimanana, I., Thomas, M. and Tautz, D. 2006. Tracing signatures of selective sweeps in natural populations of the house mouse. Molecular Biology and Evolution Advance Access published on January 18.

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

Maynard Smith, J. and Haigh, J. 1974. The hitch-hiking effect of a favourable gene. Genetic Research, Camb. **23**: 23-35.

Pacek, P., Sajantila, A. and Syvänen, A.-C. 1993 Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. PCR Methods and Applications **2**: 313-317.

Schlötterer, C. 2002. A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics **160**: 753-763.

Schlötterer, C. 2003. Hitchhiking mapping: functional genomics from the population genetics perspective. Trends in Genetics. **19**: 32-38

4 A Genome Screen gives first Insights into the Basic Parameters and Frequency of Signatures of Selective Sweeps in natural Populations

Meike Thomas and Diethard Tautz

4.1 Summary

One way of identifying the genetic basis of adaptations is to screen genomes for signatures of selection. Positive selection in a genomic region leaves a characteristic footprint behind, namely reduced neutral variation around the selected site (selective sweep). This can be employed to systematically screen for positively selected (adaptive) mutations and allows to estimate their frequency in a given population. However, the key to get a reliable measure of positive selection lies therefore in the choice of populations, for which historical information needs to be available. We have used here a comparison of two populations of the house mouse (*M. m. domesticus*) which have split upon arrival in Middle Europe about 3,000 years ago and for which subsequent major bottlenecks can be excluded. To obtain data on a large number of loci, we have employed a high throughput routine. In a first step, we have pooled the individual samples from natural populations and typed these pools for almost 1,000 microsatellite loci. Among the composite patterns that are obtained in this way, it is possible to visually select those with population-specifically reduced variability. In a second step, these candidates were then re-typed in individuals of a carefully chosen population background and statistically tested for significance. To narrow down the regions potentially affected by positive selection, additional loci in the flanking sequences of the significant candidates were analyzed. All candidate regions clearly indicate valley formations. Based on the size of the regions exhibiting a reduction in variability and the allele frequencies of the candidate loci, we postulate that the majority of positive selection events originates from standing variation and goes along with low selection coefficients as a continuous background effect of natural populations with a minimum frequency of one sweep every 100 generations.

4.2 Introduction

Evolutionary changes are caused by neutral drift and natural selection (Lewontin 1974, Kimura 1983). Despite decades of discussion, the relative importance of these factors is still not clear. One focus of current evolutionary research is to measure the frequency of adaptive changes due to positive selection in natural populations. While the current thinking is dominated by the assumption that neutral evolution shapes the major molecular divergence of genomes (Kimura 1983), Maynard Smith and Haigh (1974) proposed that the frequency of positive selection events may be several orders of magnitude higher than assumed. Until now, most attempts to estimate the average frequency of positive selection in different organisms are based exclusively on comparative sequence analysis between species (Smith and Eyre-Walker 2002, Birne and Eyre-Walker 2004, Bazykin et al. 2004). These are mainly investigations of coding regions to estimate the frequency of adaptive amino acid substitutions. Such comparative sequence analyses are able to measure protein evolution of genes which are under ongoing constraint (e.g. Fay and Wu 2001, Fay et al. 2002) but they miss recent selective events and all adaptive events that go along with changes in cis regulatory systems.

In two recently published studies new methods were presented for revealing functional noncoding DNA from comparative genomic approaches (Chin et al. 2005, Keightley et al 2005)). These projects use sequences of several closely related species and evolutionary theory to estimate the amount of functional regulatory DNA in noncoding sequences. It appears that in the larger genomes of mammals, functional elements are more diffuse than in yeast, but are clustered mostly within 2 kb surrounding protein coding sequences (Castillo-Davis 2005). Gene expression divergence seem to evolve much faster in murids than in humans, and, in comparison to the nucleotide divergence between these two organisms, it is implied that the proportion of gene expression changes that are under natural selection varies between humans and murids. Many of the mutations that affect gene expression in murids may in fact be under positive selection (Keightley et al. 2005). While several authors postulate that regulatory changes play an important role in the evolution of organisms (Zuckerkandl and Pauling 1965, Tautz 2000, Andolfatto 2005, Castillo-Davis 2005), it is obvious that positively adaptive changes in cis regulatory regions can only be measured in the early stages during the presence of selective sweep formations. These footprints are lost over time and distinguishing between single mutations in neutral and regulatory elements in the upstream regions is

difficult.

Previous approaches for identification of selective sweep signatures using microsatellite typing have been applied to situations where positive selection was predicted *a priori*. For example due to colonization of new habitats (Harr et al. 2002, Kauer et al. 2003, Kayser et al. 2003, Storz et al. 2004), resistance evolution (Wootton et al. 2002) or artificial selection (Vigouroux et al. 2005).

Many approaches screening for signatures of positive selection have focused primarily on the invasion of new habitats by humans and *Drosophila*. For both organisms the majority of results confirm the 'out-of-Africa' hypothesis, which predicts more signatures of selection in the derived versus ancestral due to adaptive changes to new environments. However, Haddrill et al. (2005) recently demonstrated that a reasonable number of signatures of positive selection within the derived populations of *Drosophila* are caused by a bottleneck connected to the 'out-of-Africa' event. Thus, the detected difference in the number of selective sweeps in prior studies may be partly due to such artifacts.

For frequency estimations of positive selection events, the major disadvantage of studies on *Drosophila* and humans is the lack of an accurate knowledge of the divergence time. Neither the splitting time of the different *Drosophila* species, nor the dispersal time of *D. melanogaster* populations is exactly known (David and Capy 1988, Lachaise et al. 1988). For *Drosophila,* dates on the 'out-of-Africa' event are anecdotic, and do not provide an accurate time frame (Haddrill et al. 2005). Furthermore, a massive bottleneck followed by an extreme population expansion makes it difficult to distinguish footprints of positive selection from the numerous artifacts (Eyre-Walker 2002).

We have chosen murid populations, which are independent of such *a priori* complications and are sufficiently young to assess the relevant historical parameters. We focus on two populations of the European house mouse *Mus musculus domesticus.* These populations harbour two major advantages: first, a well documented history based on fossil records (Cucchi et al. 2005), and second, the complete genome sequence is available for the mosaic laboratory mouse *Mus musculus* (Mouse Genome Sequencing Consortium 2002). The three major subspecies of the house mouse *Mus musculus musculus*, *M. m. domesticus*, and *M. m. castanaeus* split about 0.5 Mya on the Indian subcontient and from there spread over the world (see Boursot et al. 1993, and Guenet and Bonhomme 2003 for review). Careful analysis of the palaeontological record of

mouse fossils from excavations in human settlements of various ages have shown that the house mouse, *Mus musculus domesticus* invaded Western Europe only about 3,000 years ago from the Near East via the Mediterranean sea (Cucchi et al. 2005). The divergence of populations of the house mouse within western Europe has occurred very recently and therefore footprints of positive selection due to adaptive changes are expected to be still present. Additionally, complex influences of demographic events can be excluded because of the short timeframe within which the populations evolved.

The selection of suitable polymorphic markers and the establishment of a high throughput routine for detecting selective sweeps are necessary for the genomic analysis of several populations (Schlötterer 2002a). Since higher eukaryotic genomes contain roughly 40,000 genes and the empirical detection of polymorphic variants requires tests in multiple individuals (20-50) for diverse populations (2-10), theoretically millions of data points are necessary. Numerous studies have shown microsatellites to be excellent tools for identifying regions influenced by positive selection in a range of organisms (Schlötterer et al. 1997, Huttley et al. 1999, Kohn et al. 2000, Harr et al. 2002, Kauer et al. 2003, Kayser et al. 2003, Payseur et al. 2002, Schlötterer 2002b, Vingouroux et al. 2002, Wootton et al. 2002, Storz et al. 2004). Microsatellites occur at high numbers in most species, and above a certain repeat number they are expected to be polymorphic due to their high mutation rate. Thus, their multiallelic nature makes them informative markers that are particularly well-suited for the characterization of very recent sweeps (Schlötterer 2003, Schlötterer and Wiehe 1999). The complete genome sequence of the house mouse enables us to select and work with a large amount of microsatellites distributed throughout the genome and pinpoint their location with respect to annotated genes.

We sampled five different house mouse populations, three of the subspecies *M. m. domesticus* (one from the Cologne-Bonn area, one from the Massif Central and one from Cameroon), and for resons of comparison two of the subspecies *M. m. musculus* (one from the Czech Republic and one sampled in Kazakhstan). Whereas the major part of the analysis in this chapter is based on the comparison of the two Western Eurpean populations, the German and the Fremch one. To avoid influences of the natural inbreeding behaviour of mice, just one animal from any given sampling site was included in the analysis (Ihle et al. 2006). Genes influenced by positive selection can be identified by the characteristic footprint of reduced variability at linked neutral loci due to a

hitchhiking effect (Maynard Smith and Haigh 1974, Ohta and Kimura 1975, Slatkin 1995). Positive selection increases the frequency of a favourable mutation together with a certain linked neutral region in the population. The size of the region affected by a selective sweep depends on the local recombination rate and the selection coefficient (Maynard Smith and Haigh 1974). After the sweep event, variability is eventually recovered and the footprint is gradually lost by new mutations and random drift. To detect such regions on a genome scale we investigated almost 1,000 microsatellite loci throughout the mouse genome and determined variability levels at these loci in four natural populations. Microsatellites were identified applying a new software tool, which detects dinucleotide stretches within the first twenty kb upstream of start codons of all annotated genes in the genome (see Chapter 3). Investigated microstallite loci were selected randomly, independent of the ontology of the closely positioned gene. To analyze this amount of loci within a reasonable time, we established a novel pooling approach. In this high throughput screening approach we combine the DNA of 40 individuals from each population and amplified all samples in a pooled reaction. This provides a complex pattern of peaks when the locus is polymorphic, but a relatively simple pattern when polymorphism is lost, i.e. when a sweep candidate is detected. Genotyping single individuals shows that the peak-patterns of the pools represent the real allele frequencies reliably (see Chapter 3). The analysis of the pool pattern was done by eye. All output files were inspected by pairwise comparison between populations of the two subspecies. Loci identified by reduced polymorphism in one population but not in others are marked as candidates for selective sweeps.

The results of our investigation allow us to address four basic parameters of positive selection events in natural populations:

- The frequency of positive selection events in the house mouse genome, i.e. how often do such events occur?
- The presence of selective sweeps. Does adaptation always go along with strong environmental changes (like the colonization of new habitats) or is it a continuous background effect in all populations?
- The strength of selection that drives adaptive events. Are selective sweeps driven by strong or by weak selection? That is, do advantageous mutations

in general induce large or small beneficial effects?

- The origin of beneficial variants. Are the beneficial variants taken from standing variation or do adaptive events always go along with new mutations?

4.3 Material and Methods

4.3.1 DNA pooling approach

For the genome screen of 960 microsatellite loci (primer sequences are provided in supplement 1) we used pooled samples as templates. DNA samples of single individuals were normalized to 10 ng/µl and equal volumes of each sample were combined into population-specific pools of 40 individuals each. Primers were ordered from Sigma-Aldrich. Forward primers were labeled with FAM-dye at the 5' end. PCR was performed using a multiplex kit (Qiagen Cat.No. 206143). All reactions were done in 10 µl volumes using 30 ng of pooled DNA template and strictly following the kit protocol. Denaturation at 95°C for 15' was followed by 40 cycles of 30'' melting at 94°C, annealing (for all 960 primers) at 60 °C for 1'30'', and amplification at 72°C for 1', followed by a final elongation step at 72 °C for 30'. PCR products were diluted 1:20 in water and 1 µl of this dilution was added to 0.1 µl 500 (-250) LIZ Size Standard in 10 µl HIDI Formamid. Fragments were run on a ABI 3730 capillary sequencer.

The analysis of the pool pattern was done using Genemapper V.3.5 (Applied Bioscience) to visualize the data. All output files (see digital supplement 1) were inspected by pairwise comparison between the populations of two subspecies. Loci showing a complex pattern of peaks (indicating a certain degree of polymorphism) in one population and a relatively simple pattern (indicating extreme reduction in variability) in the second were marked as candidates for selective sweep (for further details about this pooling approach see chapter 3).

4.3.2 Genotyping single samples

Candidate loci were preselected - with a focus on the comparison of the German and the French population - based on their pool pattern and re-typed for about 40 single individuals of the respective populations. Amplification was done following the

previously described protocol for pooled samples (Chapter 3).

Flanking sequences of 200 Kb surrounding the significant candidate loci were downloaded from Genbank and screened for microsatellite loci applying the program "tandem repeats finder" (Benson 1999). Primers to amplify suitable microsatellites in these flanking sequences were designed with the software "FastPCR" (http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htm). All detected loci in these flanking regions were genotyped for the single individuals of the two investigated populations and nucleotide diversity, allele frequencies and lnRH values were calculated.

Furthermore, we downloaded eight regions of 200 Kb randomly taken from 'gene deserts' (visually identified) regions as a neutral control. Each control region came from a different chromosome, with no overlap with the chromosomes of the candidate regions. We screened these 'neutral' regions for microsatellites and amplified all possible loci for single individuals of the German and the French population. LnRH variability along these fragments were calculated.

4.3.3 Statistics

Estimation of significance

Gene diversity estimates (expected heterozygosity, corrected for sample size) for all individually genotyped loci were calculated using the program MS Analyser 3.15 (Dieringer and Schlötterer 2003).

$$\ln RH = \ln \frac{\left(\frac{1}{1 - H(loc1, pop1)}\right)^2 - 1}{\left(\frac{1}{1 - H(loc1, pop2)}\right)^2 - 1}$$

H = heterozygosity
Based on the estimator:
$H = 1 - (1 / (1+2q)^{1/2})$
(Ohta & Kimura 1973)

lnRH statistics (Kauer et al. 2003) was applied to compare variability levels between the populations. This statistic estimates the ratio of variability of a single locus between two populations. Measurement for variability is the expected heterozygosity. The significance of lnRH values is estimated by normalizing the measured values with a dataset of lnRH values from microsatellite loci throughout the genome. lnRH values of this reference data set for normalization should follow a normal distribution. Following a

z-transfomation (z = (x – mean) / standard deviation), a standard normal distribution is approximated and p-values of the investigated candidate loci can be taken from this distribution. In our approach, we used an independent set of 64 'neutrally evolving' microsatellites (collected by Ihle et al. 2006) as a reference to normalize the data. Estimations of mean and standard deviation are based on these 64 loci (mean 0.0875, standard deviation 0.8584); the data does not significantly deviate from a normal distribution (Kolmogorov-Smirnov test: lnRH p=0.724,).

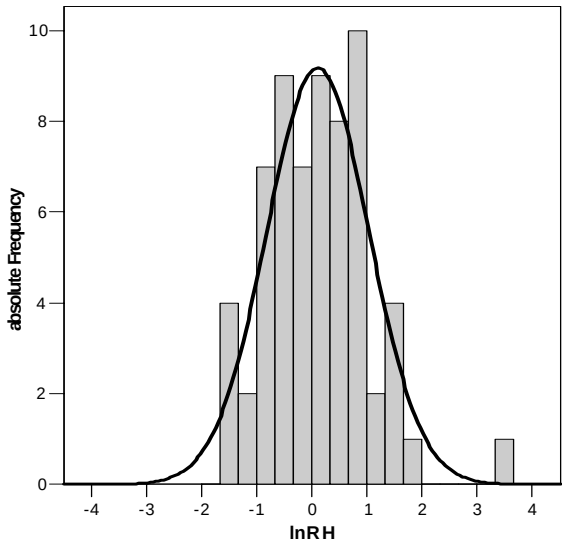


Figure 4.1 Reference data sets. A lnRH reference data set based on 64 'neutrally evolving' microsatellite loci genotyped for individual samples of the two investigated populations, Germany and France.

Significant outliers within the preselected candidate loci were detected by comparing them individually against this data set. Because of the accumulation of false positives due to multiple testing, we followed a stringent Bonferroni-adjustment. Therefore, the significance interval of z-transformed lnRH values ranged from -4.0556 to 4.0556. Significant outliers have p-values smaller than $5 \cdot 10^{-5}$.

Test for deviation of candidate regions from the genome

Deviation of the candidate regions from the neutral regions was studied by comparing the variance of lnRH values between the candidate and the neutral regions. Variance of lnRH was calculated over each of the investigated regions and applying a t-test, these variances were tested for significant differences.

Since demographic events will leave genome wide patterns behind, we included an additional data set of 118 randomly selected microasatellties distributed throughout the

genome (collected by Ihle et al. 2006) taken from the database of the Whitehead Institute/MIT Center for Genome Research and tested the homogeneity of variances in lnRH over all loci of (1) the candidate regions versus the neutral regions (=neutral state), (2) the candidate regions versus the 118 randomly picked loci (=genome variation), and (3) the candidate regions versus the microsatellite loci from the neutral regions together with the 118 randomly picked loci. A Levene test was applied to check for significant differences between the variances within the candidate regions and the variance of the neutral state/the genome variance.

In addition, we applied F-statistics (Weir and Coccerham 1984) to compare differentiation between populations in candidate regions and neutral regions. Population differentiation should be much higher in chromosomal regions that evolved adaptively, compared to neutrally divergent regions. Thus, we calculated global Fst-values over each of the valleys and over each of the investigated neutral regions. These regional global values were tested for significant differences by applying a t-test.

Under a bottleneck scenario one would expect a genome wide influence in the variance of heterozygosity. Thus, we compared the variance in heterozygosity of the 118 randomly selected microasatellties from diverse chromosomes (collected by Ihle et al. 2006) and the microsatellite loci form the neutral regions against the overall variance in heterozygosity of all genotyped loci within the candidate regions applying a Levene test.

In Figure 4.5 and Figure 4.6 z-transformed lnRH values of all investigated loci in the eight candidate and the eight neutral regions are plotted. The shape of the valleys (solid line) is displayed using a sliding window, which includes two to four loci (dependent on the marker density) for each region.

4.3.4 Selection coefficients

The estimation of selection coefficients is based on the heterozygosity values of flanking microsatellites to the most reduced locus and the recombination distance [in Morgans] between them (Stephan et al. 1992).

$$s = \frac{2c \cdot \ln(\frac{1}{2} N_e)}{\ln(1 - y)}$$

s selection coefficient

c recombination distance [recombination rate per bp distance to the sweep locus [bp]]

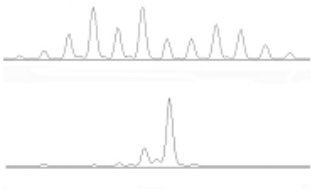
$y = H_1$(flanking locus in the non-reduced population) / H_0 (flanking locus in the reduced population)

Local recombination rates were taken from Jensen-Seaman et al. (2004). Values were calculated independently for each flanking locus. The selection coefficient of each region is an average over all estimated s-values of loci within a valley, i.e. all loci between the sweep locus and the first locus with a lnRH value larger than zero are included.

Flowchart of the approach

GENOME SCREEN

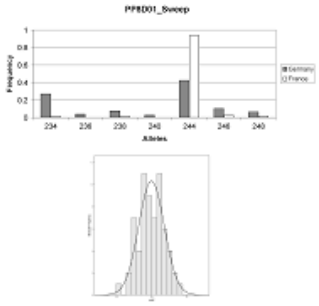
960 pooled loci



Visual inspection of the pool patterns and preselection of candidate loci for further analysis

OUTLIER IDENTIFICATION

40 candidate loci from the comparison of Germany and France



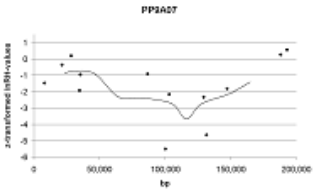
Re-typing of pre-selected candidate loci for single individuals and statistical test for singificance

Independent dataset:

64 loci as reference data set for the lnRθ statistic

VALLEY ANALYSIS

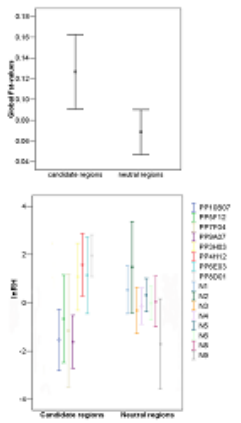
8 significant outlier
4 French sweeps : 4 German sweeps



Selection of microsatellites in 200 kb flanking sequences arround the significant outlier

Selective sweep valleys along the candidate regions

VERIFICATION OF CANDIDATE REGIONS



Statistical analysis applying variance in lnRH and F-statistics for tests of for deviation

Independent datasets:

Neutral regions

118 randomly picked loci throughout the genome

Significant deviation of candidate regions from the neutral state and the genome wide variation

Figure 4.2 Flowchart of the described approach. The selection of candidate loci from the genome screen is based on pairwise comparisons of pool patterns between the German and the French populations. All pre-selected candidate loci are in a second step re-typed for single individuals of the two focal populations. lnRH values of the candidate loci are compared to an independent reference data set of 64 microsatellites to test them for significance. Along 200 kb flanking the eight detected significant outliers, additional microsatellites were genotyped to analyze valley formations of the selective sweeps in these regions. The verification of significant deviations from the neutral state and the genome-wide variation is done by using two additional datasets: One consisting of 118 randomly picked microsatellite loci distributed throughout the genome, which should represent the genome wide variation in lnRH values, and a second one, consisting of microsatellites genotyped in eight neutral regions, which should represent the fluctuations of lnRH values from microsatellites along chromosomal regions of similar size. Applying F-statistics and comparing the variance in lnRH values led to significant results, which verified the deviation and therefore locus-specific evolution of these regions.

4.4 Results

4.4.1 Genome screen applying a DNA-pooling approach

By screening all chromosomes (each represented by 15 to 100 microsatellite loci apart from the Y-Chromosome) in a carefully chosen population background, we were able to detect candidates for signatures of positive selection events on a genome wide scale. Screened microsatellite loci are distributed randomly throughout the genome (Table 4.1). The distance to the start codon of neighboring genes is never larger than 5,000 bp, but in more than 90% of the loci it is less than 2,000 bp.

Table 4.1 Screened microsatellite loci and their distribution over the genome. Preselected loci that were marked for further analysis, and the number of loci that turned out to be significant by lnRH, thus showing a significant signature ($p \leq 5 \cdot 10^{-5}$) of selective sweeps either in the French (Fra) or the German (Ger) populations.

Chromo-some	No. of screened loci	No. of loci classified within RVC2 between Gernamy and France	No. of significant lnRH loci between Germany and France
1	31	4	1 [Fra]
2	66	3	
3	100	3	
4	42	2	
5	49	2	1 [Ger]
6	35	2	
7	83	3	
8	40	1	
9	31	1	1 [Ger]
10	74	1	1 [Fra]
11	83	1	
12	14	1	
13	46	1	1 [Fra]
14	30	2	1 [Ger]
15	37	3	1 [Fra]
16	29	1	
17	35	2	
18	15	1	
19	24	2	1 [Ger]
X	96	5	
	Σ 960	Σ 41	Σ 8 [4 Fra, 4 Ger]

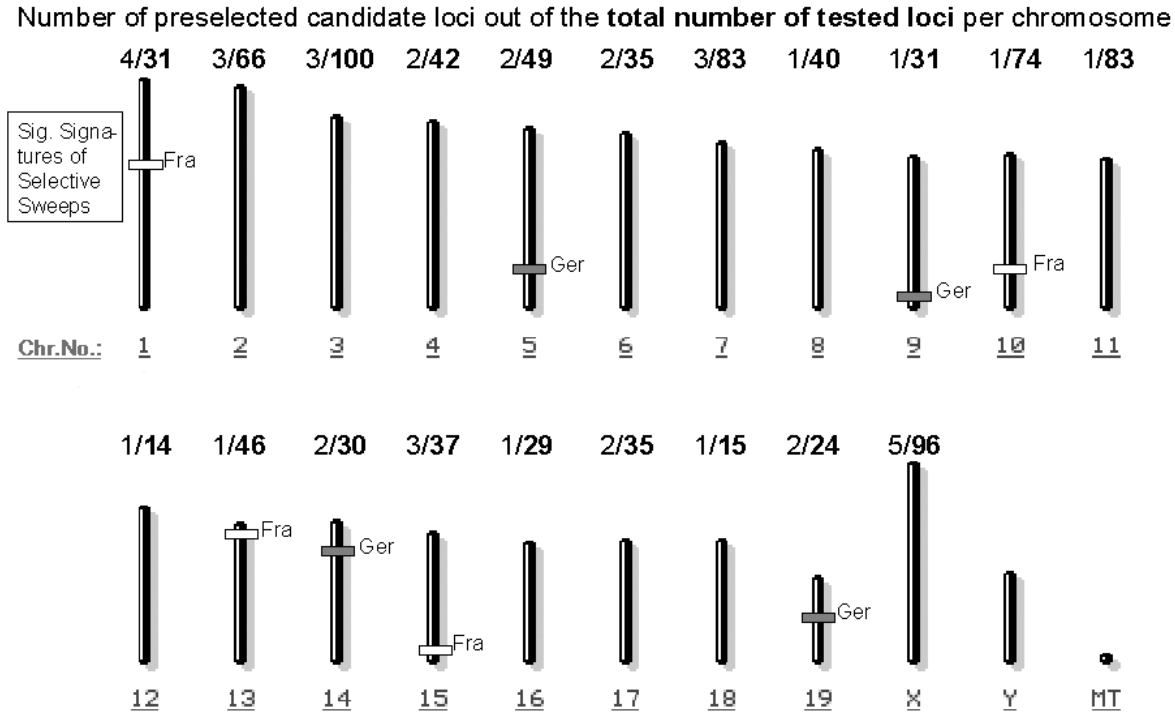


Figure 4.3 Distribution of screened and preselected loci throughout the genome. The location of significant signatures of selective sweeps in the French population are depicted with white and significant signatures in the German population with grey bars.

Almost all of the 960 investigated microsatellites amplified successfully, in only 52 cases were the results inconclusive. By comparing pool patterns of all investigated microsatellites, we found numerous loci of interest for further analysis (for results of the comparison of the German and the French population see Table 4.1).

We marked 41 candidates which showed an extreme difference in variability between the German and the French population. These loci are distributed all over the genome, they are postioned on the autosomes as well as on the X-chromosome (see Figure 4.3). Within these 41 loci, 19 were selected because of an extreme reduction in variability in the German population compared to France, and 22 were selected because of an extreme reduction in variability in the French population compared to Germany. Genotyping single individuals demonstrated that the pool patterns accurately represented the real allele frequencies. Individual typing of loci showing a polymorphic pool pattern in all populations as a control was not required (see Chapter 3).

4.4.2 Verification of candidate loci

To systematically assess the frequency of positive selection events, we genotyped all preselected loci for single individuals of the two focal populations and tested them for

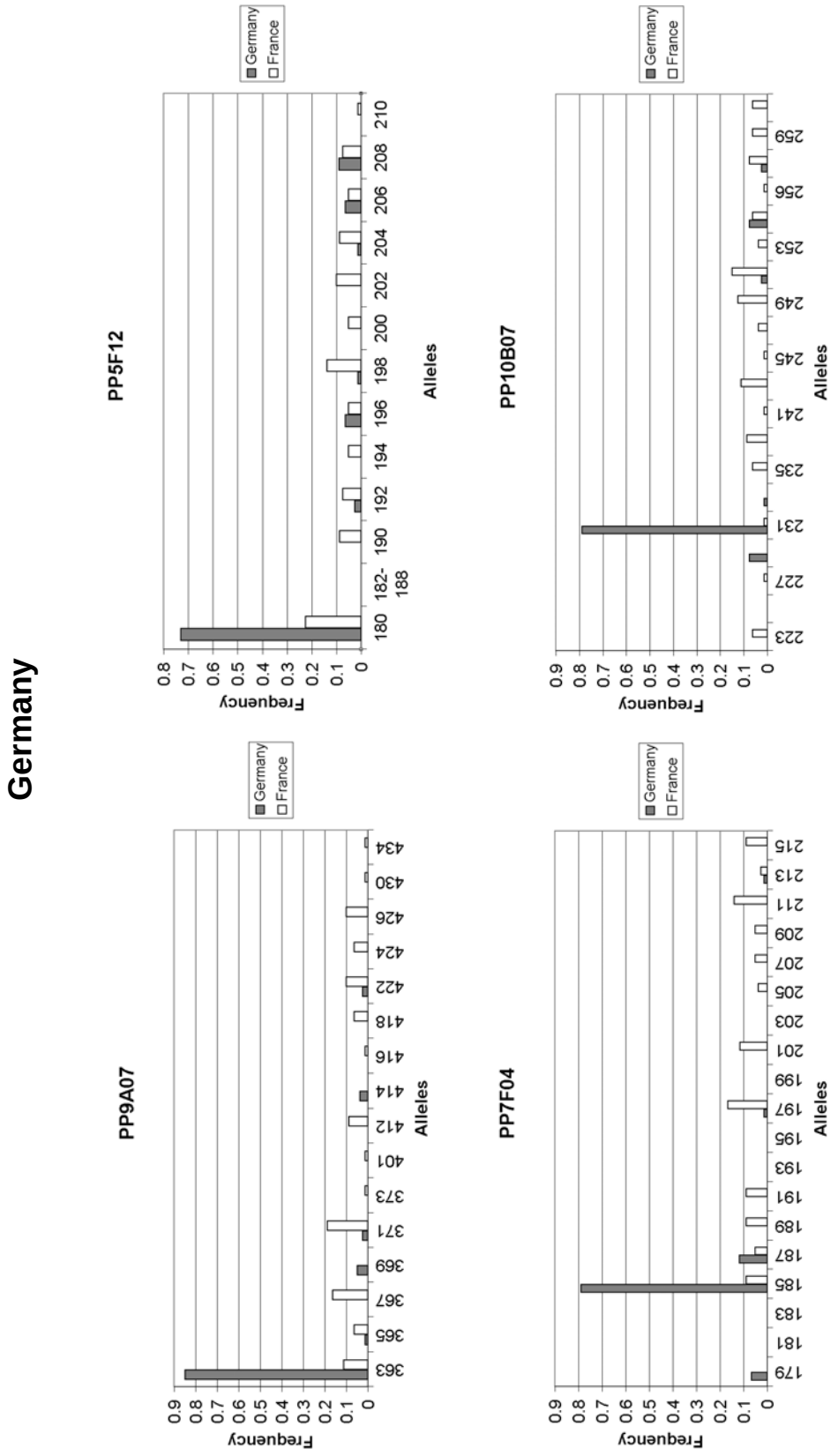
significance. The lnRH statistic (Schlötterer 2002b, Kauer et al. 2003) was applied to identify significant outlier in the comparison of the German and the French population. Candidate loci were individually compared to fixed reference data sets (see Chapter 4.3.3, Ihle et al. 2006). After following a stringent Bonferroni-adjustment, significant outliers have p-values $\leq 5 \cdot 10^{-5}$. According to lnRH, eight of the almost 1,000 screened microsatellite loci showed a significant sweep pattern between the two focal *M. m. domesticus* populations. Significant outliers are positioned on different autosomal regions of the genome and equally distributed between both populations. Candidate loci of the French population were found on Chromosome 1, 10, 13 and 15, whereas loci for the German population were located on Chromosome 5, 9, 14 and 19 (Figure 4.3). Each locus is positioned on a different chromosome, which excludes physical linkage. Recombination rates in these candidate areas vary between 0.22 to 2.06 cM/Mb (Table 4.2 taken from Jensen-Seamen et al. 2004). None of the preselected candidates on the X chromosome showed up as a significant outlier in the applied statistic.

Table 4.2 Loci that showed a significant difference in variability according to the lnRH test statistics, expected Heterozygosities, physical position and recombination rate (taken from Jensen-Seaman et al. 2003)

Marker name	Chromo-some	Physical Position [kb]	Recombination rate [cM/Mb] (5Mb window)	Exp.Het. Ger.	Exp.Het. Fra.	LnRH z-value (p<0.00005)	Sweep in
PP3H03	10	80,316	0.44	0.7092	0.1187	4.1250	Fra
PP4H12	13	3,690	0.22	0.5389	0.0250	4.8689	Fra
PP5F12	19	29,304	1.022	0.4545	0.8949	-4.3378	Ger
PP6E03	1	88,575	0.51	0.7247	0.0737	4.9066	Fra
PP7F04	5	104,783	0.81	0.3628	0.9084	-5.2187	Ger
PP8D01	15	98,140	2.06	0.7353	0.1277	4.2583	Fra
PP9A07	14	43,925	0.44	0.2756	0.8968	-5.4968	Ger
PP10B07	9	109,088	0.342	0.3718	0.9244	-5.6120	Ger

In almost all cases one of the shortest alleles is fixed. Only for locus PP8D01, the second longest allele is the major one in the sweep population. All French sweep loci showed a sweep allele frequency of more than 90%, whereas in three of the German candidates the sweep allele frequency was below 90%. In two of these more than 80% of the chromosomes carried the major allele and at the third locus the allele is present in more than 70% of the samples. In all cases the sweep allele is also present in the non-sweep population (see Figure 4.4).

Based on the pool patterns we found two candidates, for which we can be reasonably shure that lnRH statistic would yield significant results for a population specific sweep in Kazakhstan compared to the samples collected in the Czech Republic (both belong to the subspecies *M. m. musculus*). The samples from Kazakhstan represent an ancestral population (data not shown).


Germany
PP5F12
PP10B07
PP9A07
PP7F04
Germany
France
Frequency
Alleles

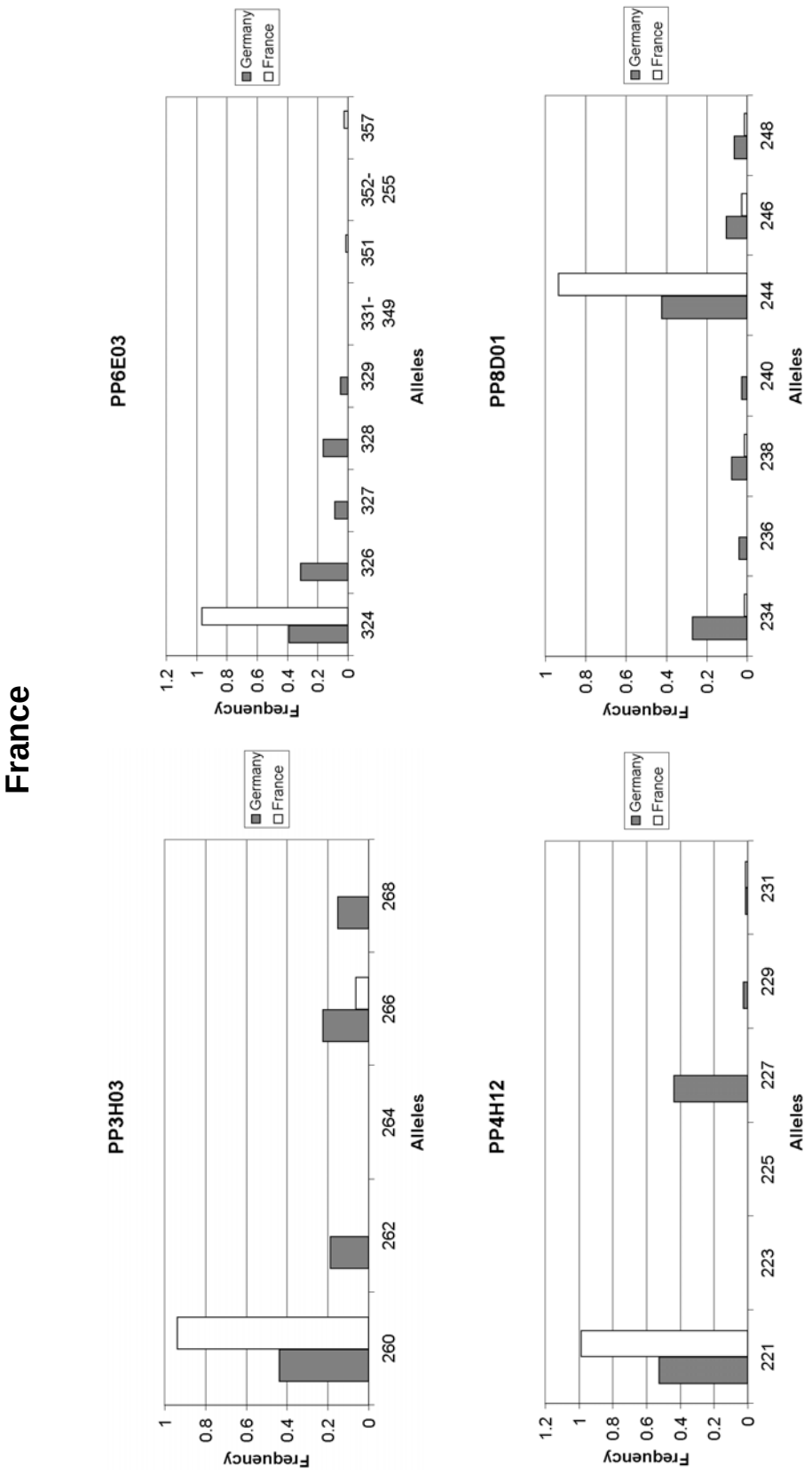


Figure 4.4 Allele frequencies of the eight significant outlier loci. In all cases at least four alleles were present in the variable population, and beside the one major allele in the sweep population, always one or more low frequency alleles were detected which appear distal to the high frequency allele.

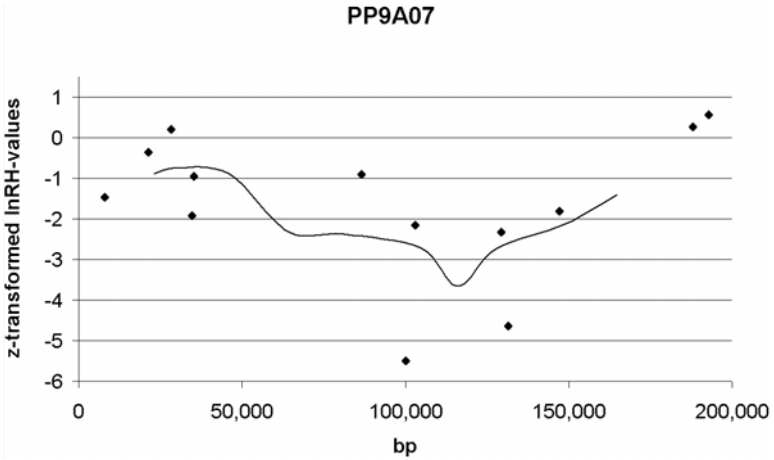
4.4.3 Sweep valley analysis

The size of the sweep window depends on the strength of selection and the local recombination rate. If the former is high and the latter is low, we would expect relatively large regions with reduced polymorphism. To assess the size of the sweep windows, we have typed flanking microsatellites for each locus with a significant sweep signature (primer sequences of these flanking microsatellites see supplement 3). We covered areas between 60 and 180 Kb surrounding the preselected outlier with five to twelve neighbouring microsatellite loci (see Figure 4.5). To cover the candidate regions as densely as possible we selected not only dinucleotide repeats, but also tri- and tetranucleotide repeats.

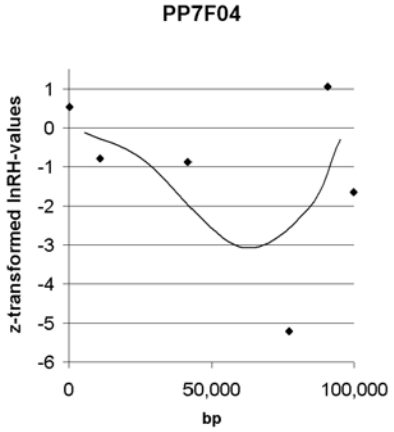
Testing whether the fluctuation of microsatellite lnRH values within these candidate regions deviates from the fluctuations of neutrally evolving regions requires a comparison to such regions. Thus, we picked eight areas on different chromosomes (Chromosome 3, 4, 6, 7, 8, 11, 12, and 17). We took two factors into account when selecting these areas: First, we picked them in 'gene-deserts' regions (visually identified) to reduce potential influences of selection, and second, chose loci that do not reside in regions with extreme recombination rates (Jensen-Seaman et al. 2004). Microsatellite loci along these eight regions were genotyped. Areas of 80 to 180 Kb were covered with numerous microsatellites (three to nine per region) (Figure 4.6).

Compared to lnRH fluctuations of microsatellites in neutral regions, four (PP8D01, PP10B07, PP4H12, PP9A07) of the eight candidate regions clearly indicate valley formations around the preselected microsatellites. These regions are positioned on Chromosome 9, 13, 14, and 15, two exhibiting a reduction in variability in the German population and two in the French population. Three of the four regions contain at least one closely positioned microsatellite flanking the preselected locus significantly reduced at the 5% level. The fourth locus (PP10B07) is flanked by two mirosatellites without a significant lnRH value. However, in this case, in both directions the second flanking microsatellites are significantly reduced at the 5% level.

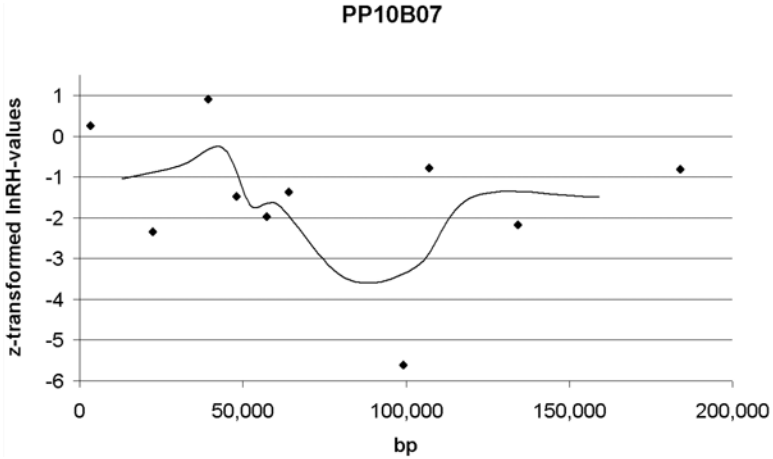
Germany



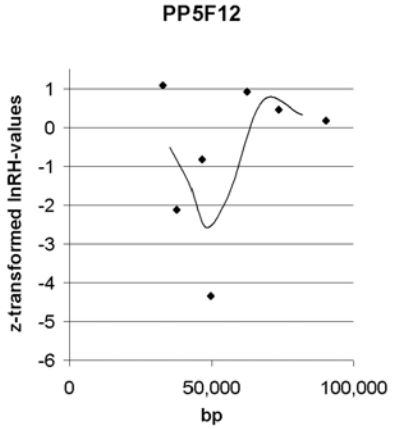
Chr. 14 / pp. 43,925 kb / r = 0. 44 / s = 0.0018



Chr. 5 / pp. 104,140 kb / r =0.81 / s = 0.0029

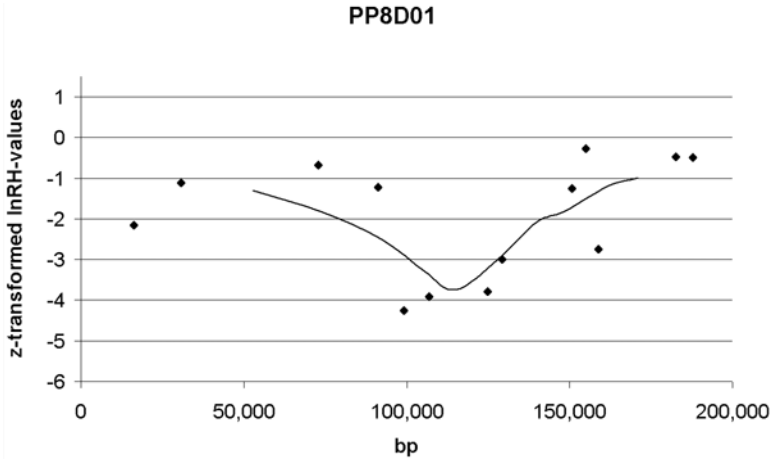


Chr. 9 / pp. 109,130 kb / r = 0.34 / s = 0.0015

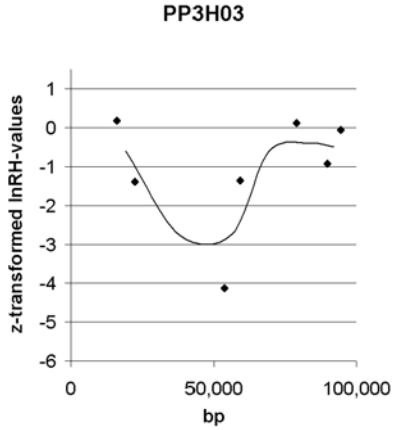


Chr. 19 / pp. 29,304 kb / r = 1.0 / s = 0.0011

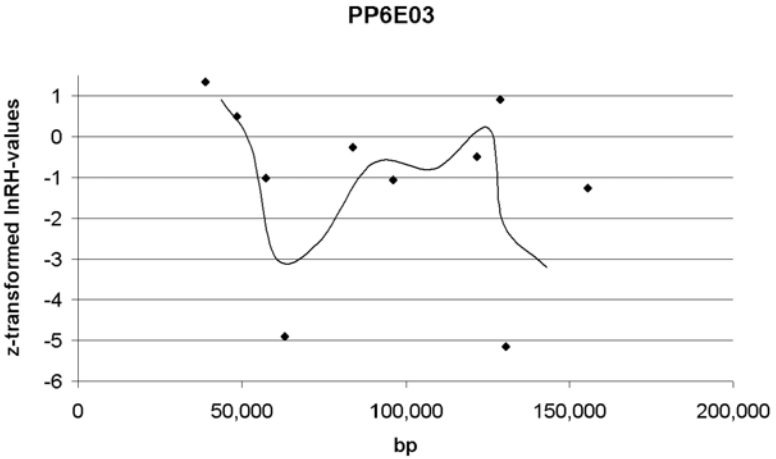
France



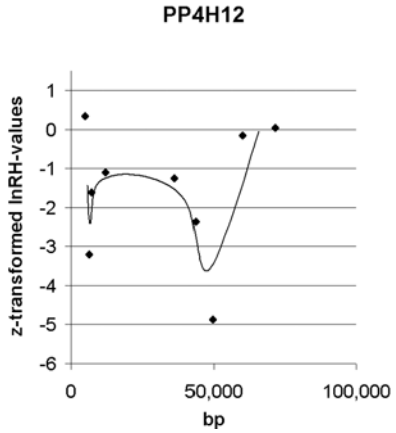
Chr. 15 / pp. 98,140 kb / r = 2.06 / s = 0.021



Chr. 10 / pp. 80,540 kb / r = 0.44 / s = 0.0008



Chr. 1 / pp. 88,575 kb / r = 0.51 / s = 0.0013

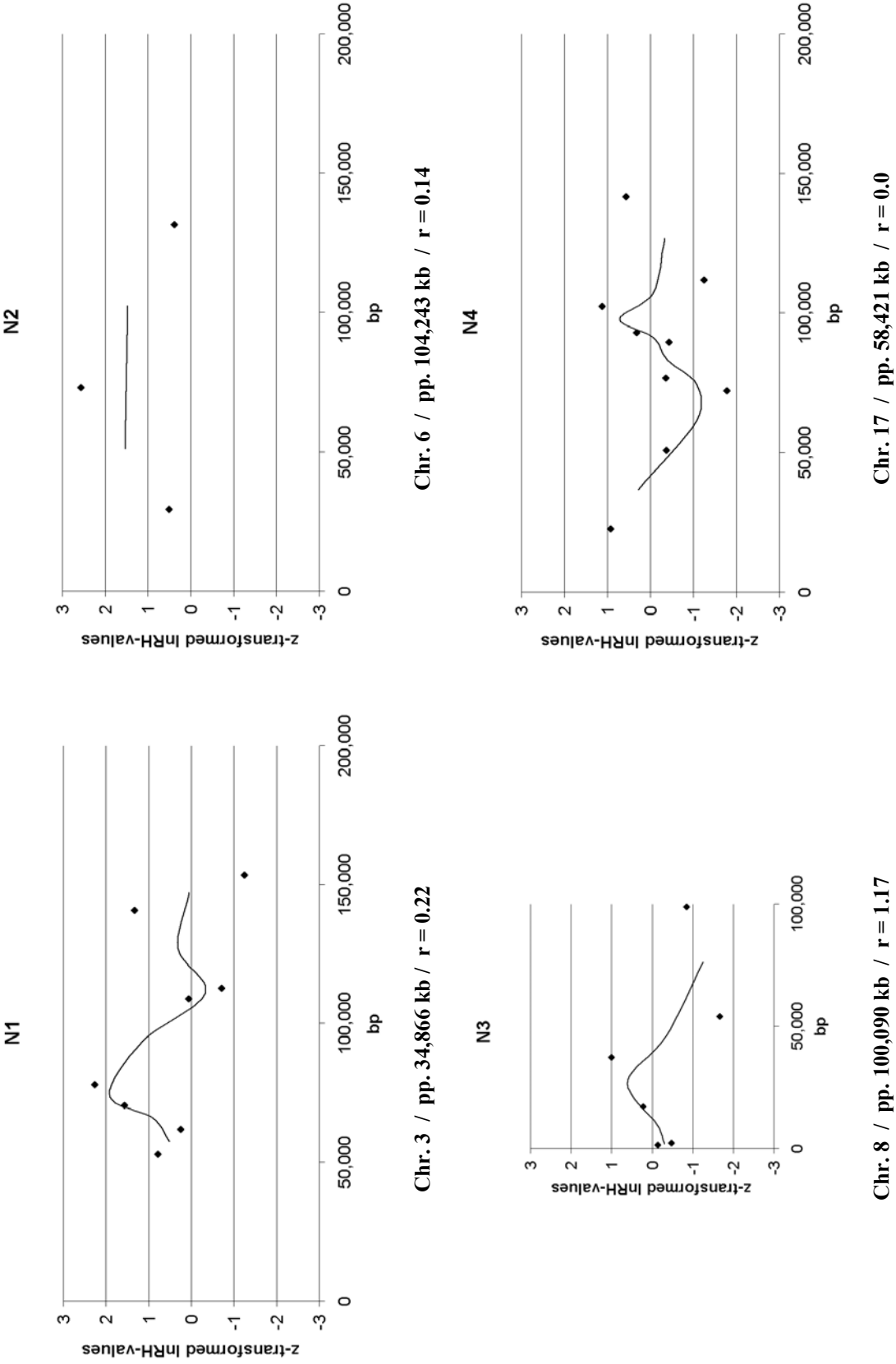


Chr. 13 / pp. 3,690 kb / r = 0. 22/ s = 0.0007

Figure 4.5 Depiction of sweep valleys around the loci with highly significant signatures of selective sweeps. The four loci in the left column represent the loci found in the German population, the ones in the right column the French population. Normalized (z-transformed) lnRH values are plotted against chromosomal position. All graphs have the same scales. lnRH values below 4.05 are highly significant ($p < 5 \cdot 10^{-5}$). The top four regions show at least two highly significant sweep loci within the valley. The lines represent the averages of the lnRH values, as determined by a sliding window technique (see Methods). "pp." represents the position in the genome sequence of the respective chromosome, "r" represents the local recombination rate, "s" the estimated selection coefficient (see Methods). Note that the lnRH values are not always consistently low in a given valley. We ascribe this to different mutation rates at different loci, or for different alleles, i.e. although all loci in a valley are expected to have gone through the sweep, some would recover more quickly than others because of a high primary mutation rate (see also supplement material).

The other four investigated regions (PP6E03, PP5F12, PP3H03, PP7F04), do not contain flanking loci exhibiting significant reduced variability. Here, the lnRH-values fluctuate diffusely. The lack of significant reduced variability in the flanking regions might be due to the absence of closer positioned microsatellites. These results are consistent with the observations of Harr et al. (2002), Kauer et al. (2003) and Kayser et al. (2003), who observed several regions for which the reduction in variability could be confirmed by flanking loci, but others lack significance in the flanking regions.

In Figure 4.7 the allele frequencies of microsatellites genotyped within two candidate regions are illustrated. The number of alleles varies between the investigated microsatellite loci. Given that the same number of individuals is always typed, a high number of alleles is an indication of a large number of repeats (since the number of detected alleles strongly correlates with the sample size). The more repeats a locus has, the faster it mutates. Thus, microsatellites with high mutation rates already might have recovered variability, whereas others with lower mutation rates, might still show the extreme reduction in variability due to the sweep event. This might be also the reason for the often observed pattern of the shortest allele being the sweep allele. Loci in which another than the shortest allele went to fixation due to a selective sweep event recover variability faster, because longer alleles have higher mutation rates than shorter alleles. If only one very short allele is present in a population, it might take a longer time to accumulate new alleles via mutations than in situations where a long allele went to fixation by positive selection. This gives us a possible explanation why we observe this pattern in almost all candidates. The sweep pattern persist longer, and is therefore more likely to be detected, in short fixed alleles, and is faster blurred in longer ones. The regions influenced by significantly reduced variability ranged from 20 kb to more

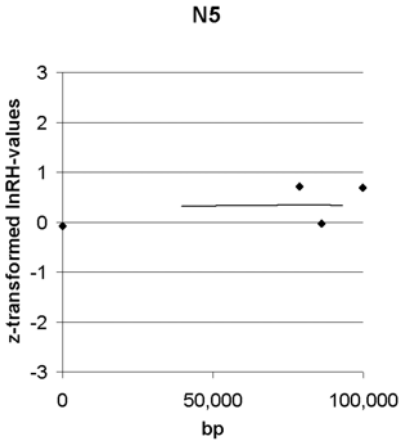


Chr. 3 / pp. 34,866 kb / r = 0.22

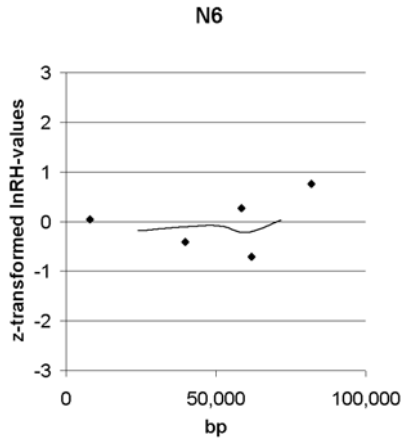
Chr. 6 / pp. 104,243 kb / r = 0.14

Chr. 8 / pp. 100,090 kb / r = 1.17

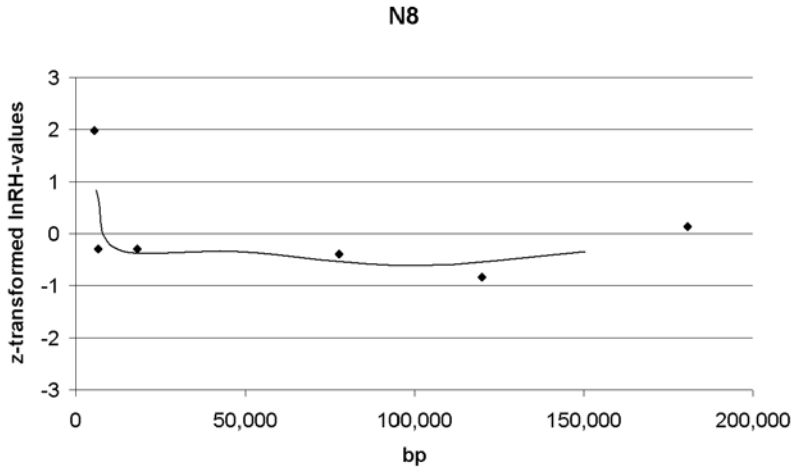
Chr. 17 / pp. 58,421 kb / r = 0.0



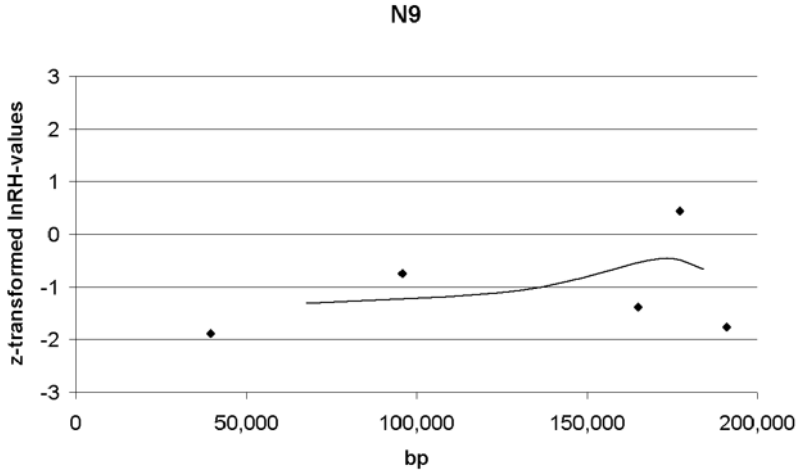
Chr. 4 / pp. 27,402 kb / r = 0.0



Chr. 12 / pp. 92,724 kb / r = 0.40

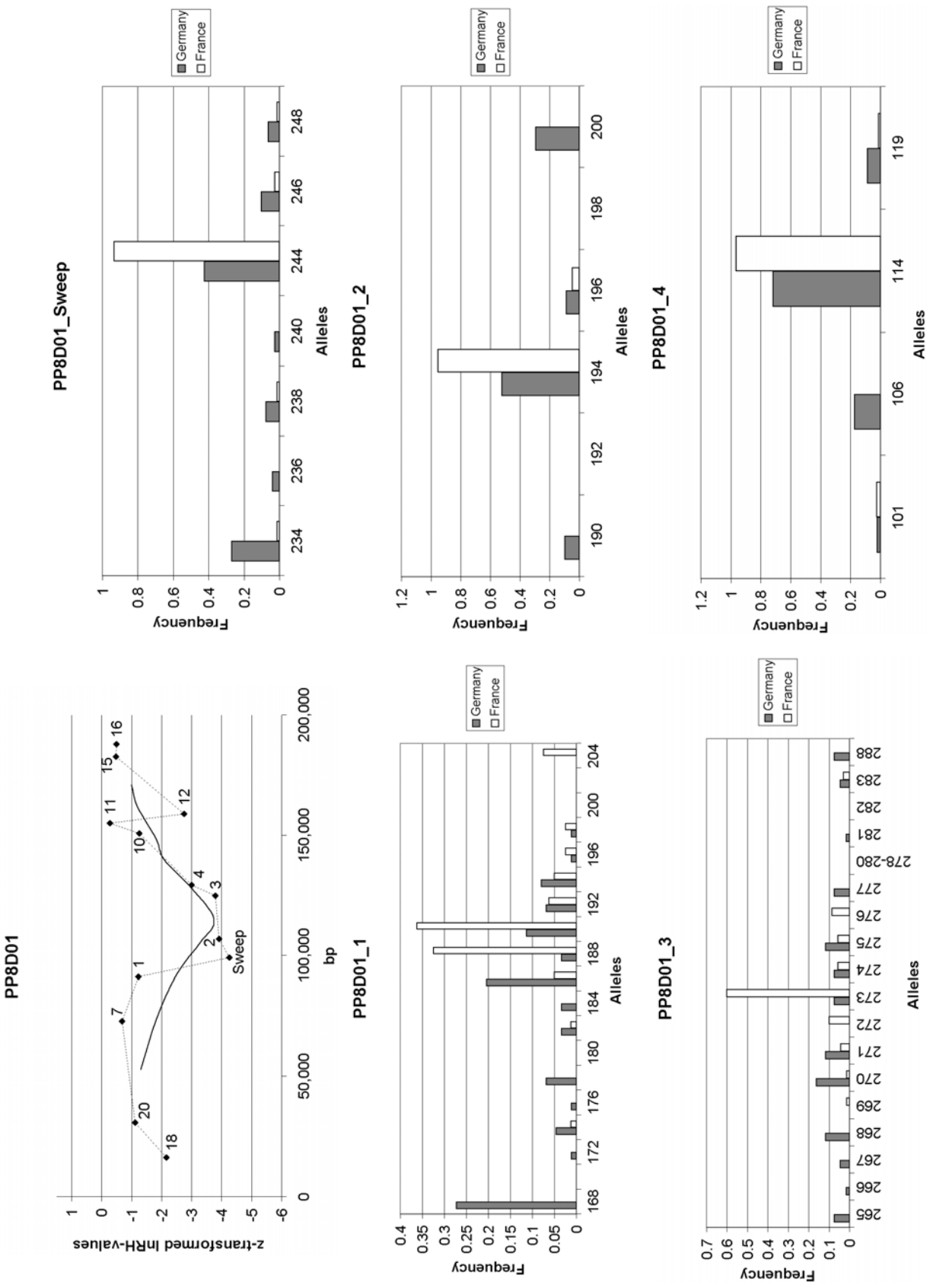


Chr. 7 / pp. 89,410 kb / r = 0.45



Chr. 11 / 38,981 kb / r = 0.66

Figure 4.6 Microsatellite loci genotyped for the two focal populations along 'Neutral regions'. The labeling of the Neutral Regions is from 1 to 9, Neutral Region 7 amplified in too few microsatellite loci and is therefore excluded.


PP8D01
z-transformed lnRH-values
bp
Sweep
PP8D01_Sweep
PP8D01_1
PP8D01_2
PP8D01_3
PP8D01_4
Germany
France
Frequency
Alleles

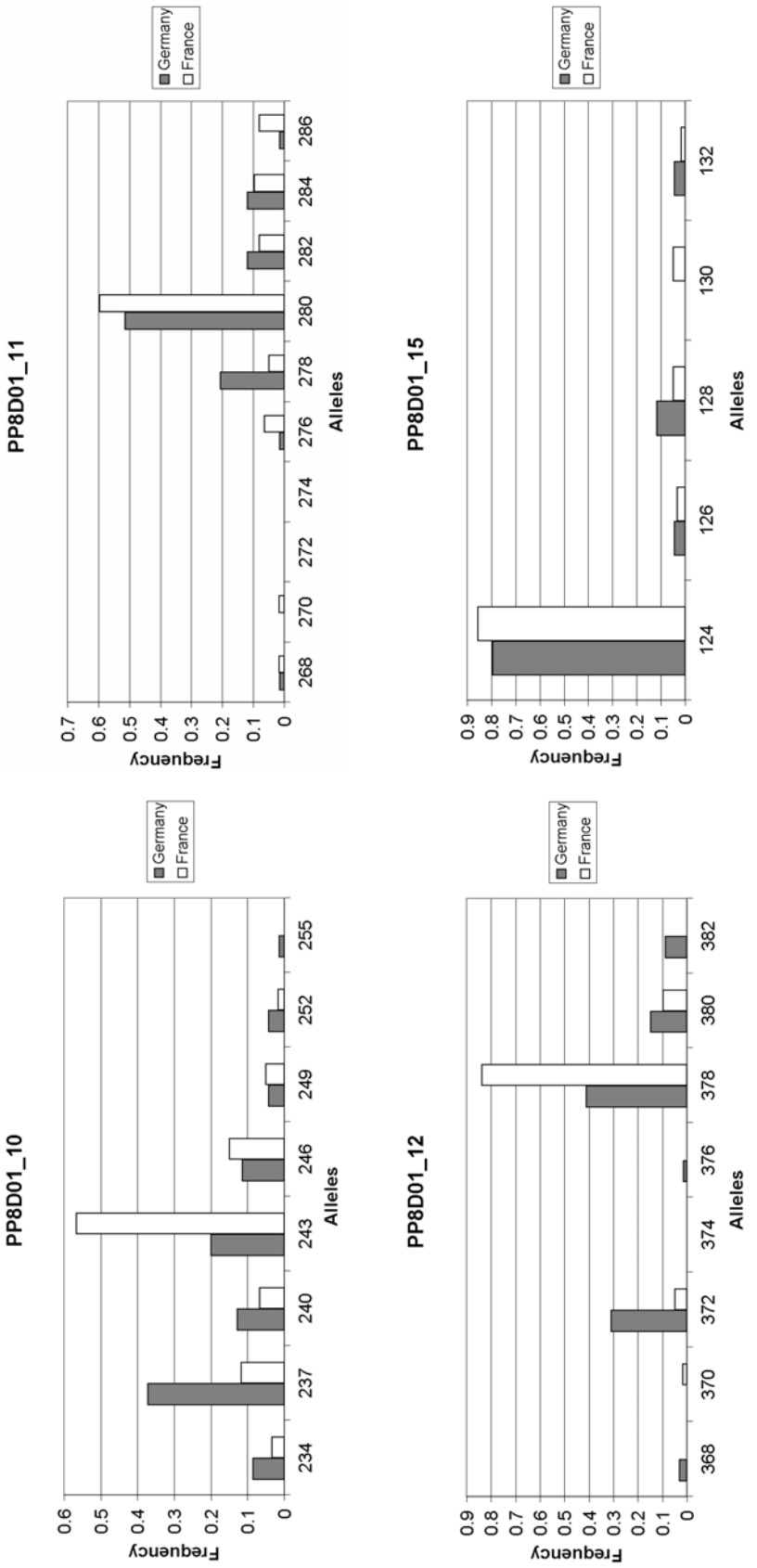
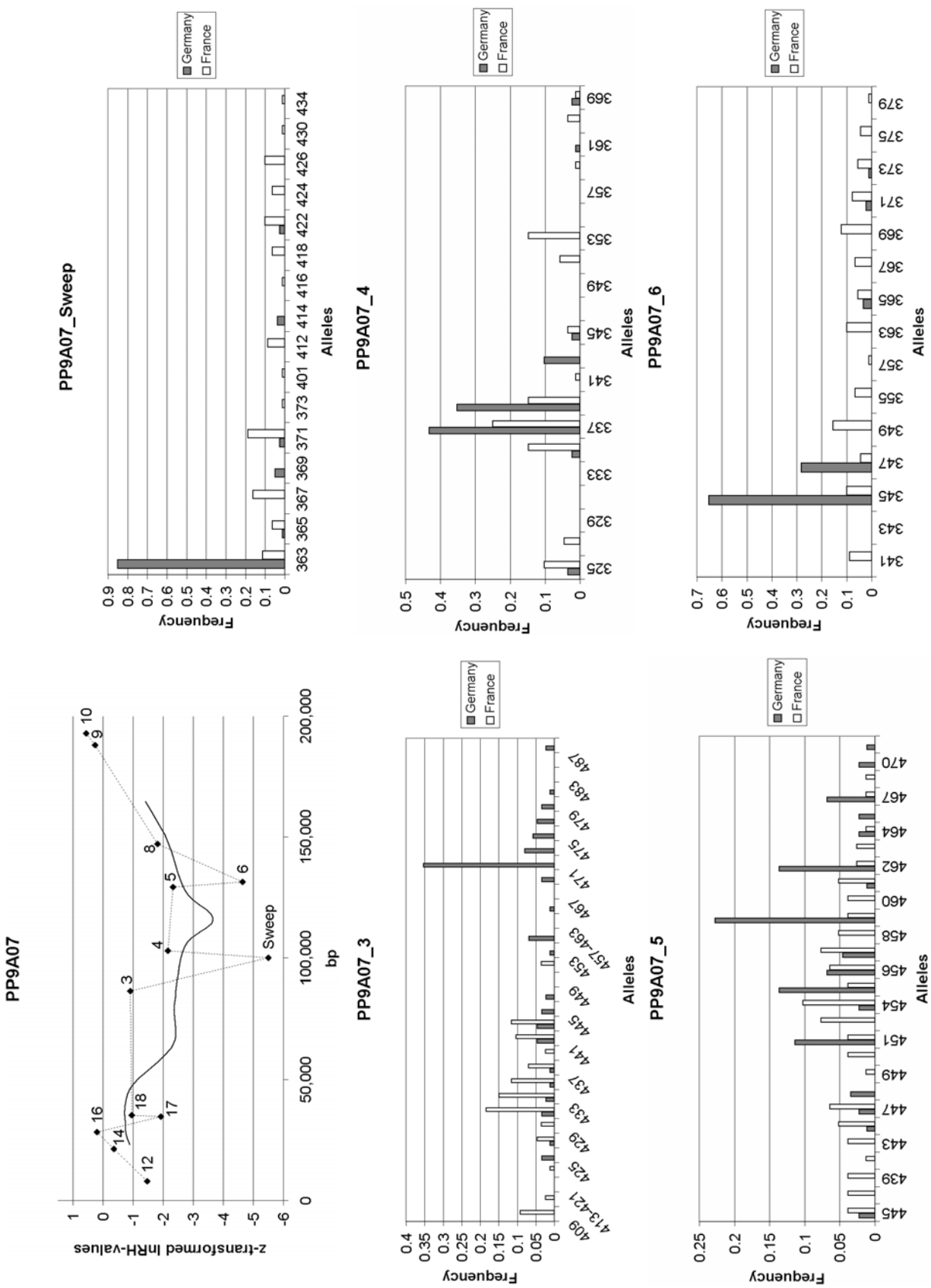


Figure 4.7 Allele frequency distribution for loci in two of the candidate regions

A Example 1 Locus PP8D01: The region shown below shows an extended sweep valley. The originally identified locus is labeled as "sweep" - the flanking loci are labeled with numbers. The allele distributions for each of these is shown below. Loci 10 an 11 appear to have recovered their variability after the sweep. Within the valley microsatellites with few alleles as well as microsatellites with numberous alleles show an extreme reduction in variability in the French population.


PP9A07
z-transformed lnRH-values
bp
Sweep
PP9A07_Sweep
PP9A07_4
PP9A07_6
PP9A07_3
PP9A07_5
Germany
France
Frequency
Alleles

B Example 2 Locus PP9A01: The region shows an extended sweep valley. The originally identified locus is labelled as "sweep" - the flanking loci are labeled with numbers. The allele distributions for each of these is shown below. Loci 4 and 5 appear to have recovered their variability after the sweep. The two major alleles in locus 4 are only one mutational step away, suggesting that one of them could have arisen by de novo mutation after the sweep. Locus 5 shows a large number of alleles, which is indicative of a generally high mutation rate, which could have led to a recovery of the variability after the sweep (note that the apparent new alleles are still grouped around the major allele, although not in the typical step-wise pattern).

than 160 kb. These numbers fit to earlier studies in *Drosophila*, where valleys of up to 100 kb were detected (Harr et al. 2002, Kauer et al. 2003, Beisswanger et al 2005).

4.4.4 Verification of sweep valleys

A comparison of variances in lnRH values (Schöfl and Schlötterer 2004) gave a significant difference between candidate and neutral regions (t-test p=0.0006, Figure 4.8). This analysis of variances represents an independent comparison. It is not influenced by differences in mutation rates because this parameter is eliminated in the lnRH statistic. Thus, the results validate the hypothesis that the contrasting pattern of candidate regions versus neutral regions is most probably not shaped by neutrality.

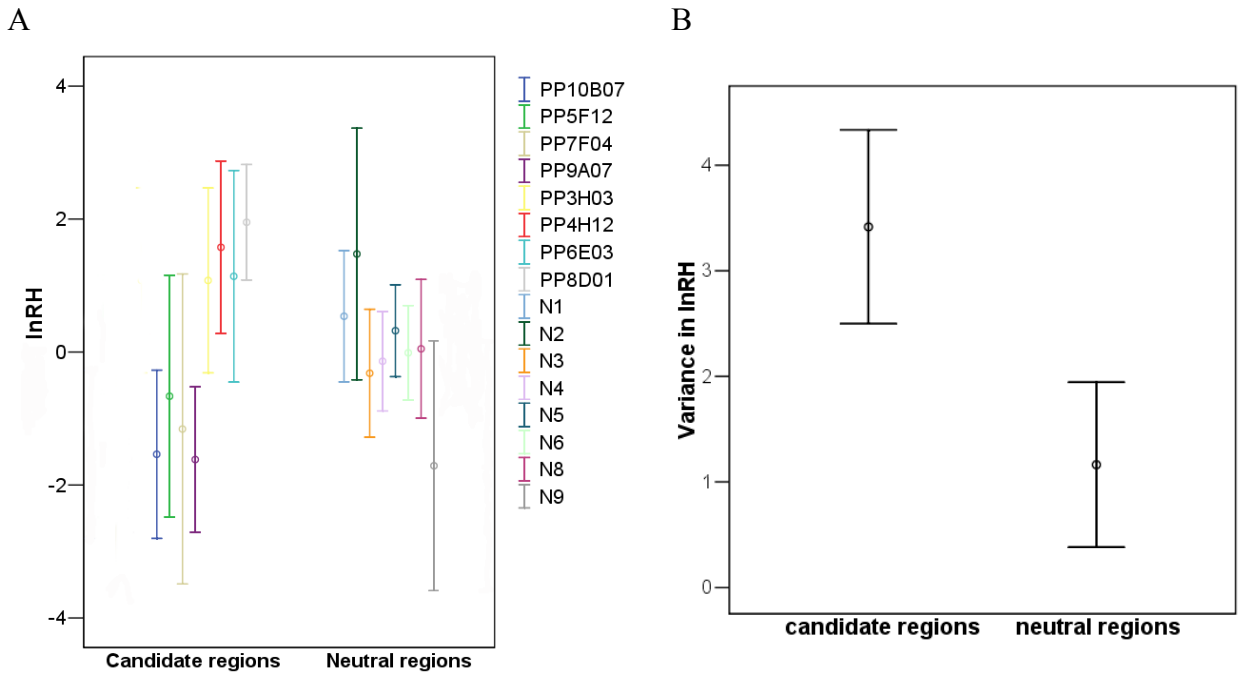


Figure 4.8 Comparison of variance in lnRH (error bars indicate the 95% confidence interval). (A) Distribution of lnRH values within the different regions of the candidate and the neutral areas. (B) Range of variance in lnRH values over the investigated candidate and neutral regions.

Table 4.3 Global Fst-values over the candidate regions and the eight neutral regions. The values significantly differ from eachother (t-test p=0.005; Kolmogorov-Smirnov p= 0.9).

Region	Size of the region [Kb]	Number of microsatellites in the region	global F_{st}
PP3H03	78	8	0.072957
PP4H12	77	9	0.176084
PP5F12	57	7	0.091722
PP6E03	117	10	0.109144
PP7F04	100	6	0.200342
PP8D01	172	13	0.120216
PP9A07	185	13	0.106477
PP10B07	180	10	0.13669
N1	100	8	0,061611
N2	100	3	0,096599
N3	100	6	0,064182
N4	120	9	0,050171
N5	100	4	0,060804
N6	80	5	0,035089
N8	180	6	0,064457
N9	150	6	0,116193

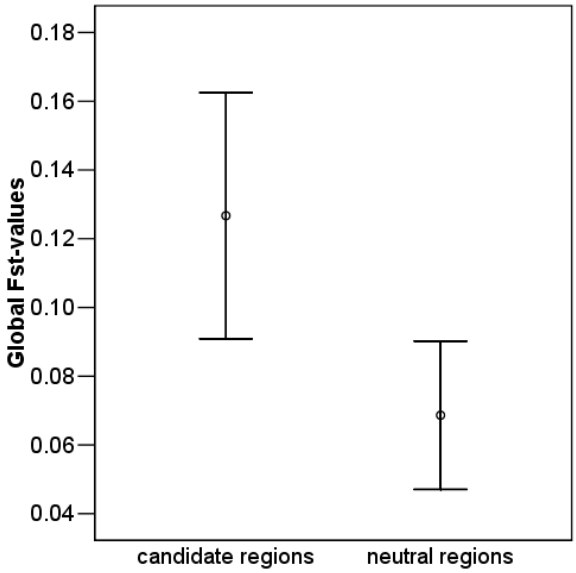


Figure 4.9 Comparison global F_{st}-values over candidate versus neutral regions (error bars indicate the 95% confidence interval). T-test confirmed a significant deviation of the two compared regional sets (p-value = 0.005).

Furthermore, we estimated global F_{ST} values (Weir and Cockerham 1984) over the candidate as well as the eight neutral regions using the software MSA (Dieringer and Schlötterer 2003). We tested whether F-statistic supports the deviation of candidate loci from the neutral expectation of differentiation. The t-test gave a significant result (Figure 4.9), a further indication for deviation from neutrality.

4.4.5 Bottleneck scenario

We included an independent dataset of 118 microsatellite loci distributed thoughout the genome (collected in Ihle et al. 2006) taken from the database of the Whitehead Institute/MIT Center for Genome Research and estimated the variance in heterozygosity for the German, the French, and the Kazakhstan population. The two *M. m. domesticus* populations are the 'potential bottleneck' populations, their variances in heterozygosity are 0.02061 (France) and 0.01656 (Germany), whereas an older, ancestral and therefore assumed to be 'non-bottlenecked' Kazakhstan population shows a higher value (0.02471), though this difference is not significant. Anyways, the ancestral population shows a higher variance than the candidate populations, an indication for a low probability of artificial valley formations due to a bottleneck. Consistent with the results of Schöfl and Schlötterer (2004) we do not detect a difference in mean heterozygosity between the ancestral (Kazakhstan 0.7859±0.1572) and the two derived popuations (Germany 0.7956±0.1287; France 0.7735±0.1436), which would be expected after a recent bottleneck event.

Table 4.4 Average Heterozygosity estimates from randomly picked loci (SD = standard deviation) (Ihle et al. 2006)

Parameter	Germany	France	Kazakhstan
N	118	118	118
Mean Heterozygosity	0.7956	0.7735	0.7859
SD Heterozygosity	0.1287	0.1436	0.1572
Variance	0.02061	0.01656	0.02471

Demographic events like bottlenecks have genome wide consequences, whereas selection is expected to have locus specific effects (Stajich and Hahn 2005). Thus, a deviation of candidate regions from the rest of the genome would validates the outstanding structure of these regions and clearly indicate that these regions have undergone a locus specific evolution. Additionally, comparisons of the variance in lnRH values over loci genotyped within the candidate regions with (1) the variance in lnRH over all loci genotyped in the eight neutral regions (for the comparison with the German candidates Levene test p=0.0325, for the comparison with the French candidates Levene test p=0.0331), (2) the variance in lnRH over 118 randomly

selected loci (for the comparison with the German candidates Levene test $p=2.91 \cdot 10^{-5}$, for the comparison with the French candidates Levene test $p=2.18 \cdot 10^{-5}$), and (3) the variance in lnRH over the loci in neutral regions together with the 118 randomly chosen loci (for the comparison with the German candidates Levene test $p=8.29 \cdot 10^{-5}$, for the comparison with the French candidates Levene test $p=9.13 \cdot 10^{-5}$) revealed significant results. Whereas the variances in lnRH values are homogenious between the candidate regions of the two investigated populations (Levene test p=0.823). This confirms the significant deviation of the candidate regions from the rest of the genome - a result that indicates locus specific evolution.

4.4.6 Selection coefficients

Based on the measured heterozygosity values of the flanking microsatellites within a valley and the local recombination rate, we estimated the selection coefficients. Values vary between 0.0007 and 0.0210. These numbers are averages over the locus wise estimation of two to twelve flanking microsatellites (Supplement A and B). For loci that have lnRH values larger than zero, the selection coefficient cannot be estimated. Thus, all loci that are positioned between the preselected candidate and the first locus exhibiting an lnRH value larger than zero are included in the regional selection coefficients.

Region	average selection coefficient
PP8D01	0.0210
PP4H12	0.0007
PP3H03	0.0008
PP6E03	0.0013
PP9A07	0.0018
PP10B07	0.0015
PP7F04	0.0029
PP5F12	0.0011

Table 4.5 Selection coefficients estimated for the eight candidate regions. The calculation is based on the heterozygosity values of flanking microsatellites and the recombination distance to the most reduced locus.

4.5 Discussion

Given the time frame within which the investigated populations were established, we can make a straightforward estimate of the frequency of positive selection events in natural populations of the house mouse. If we assume that the average window size of a sweep is not larger than 50kb, one should expect a total of 40,000 possible selectable loci (assuming a genome size of $2 \cdot 10^9$bp). From fossil records it is known that the house mouse invaded Western Europe no longer than 3,000 years ago (Cucchi et al. 2004). With an assumption of three generations per year (Karn et al. 2002), our lineages are separated by a total of 18,000 generations. With 0.42 - 0.83% sweep loci identified (four to eight out of 960), we arrive at an estimate of 0.93 - 1.85 selective sweeps per 100 generations, i.e. higher than the commonly stated maximum of 1 in 250 that was initially proposed by Haldane (1957).

Doubts about the authenticity of the identified sweep patterns because of possible bottleneck or drift events can be rejected. The lnRH statistic is a priori relatively refractory to the effects of bottlenecks (Kauer et al. 2003, Schlötterer 2002b, Schöfl and Schlötterer 2004). However, the consequences of bottlenecks comparable to that in *Drosophila* with the emigration out of Africa have yet to be sufficiently studied (Wiehe and Schlötterer, pers. comm.). Such Bottlenecks lead to a decrease in average heterozygosity as well as an initial increase in the variances of heterozygosities. Both factors would potentially inflate the number of signatures of selective sweeps in our statistics, although only under limited conditions (Wiehe and Schlötterer. pers. comm.). When we compare our populations with a more ancestral population from Kazakhstan, we detect no significant differences in the averages and variances of heterozygosities for a sample of 118 randomly picked microsatellite loci distributed throughout the genome (Table 4.4). This would suggest that there was no major bottleneck involved in the colonization of Middle Europe, which is in line with the proposal of Cucchi et al. (2005) that colonization became possible only after regular trading had started across the Mediterranean. Most importantly, however, the two populations that we compare in the lnRH statistic have very similar population parameters, i.e. we do not compare an ancestral and a derived population, but two that have recently split from each other. Thus, even if loci became randomly fixed during the colonization process, they would not show up in our analysis, because they would

be present in both populations and would therefore be cancled out in the lnRH statistic. Furthermore, if a subsequent bottleneck after the colonization of Middle Europe would have affected one of the populations, this would be visible in the average heterozygosity estimate, since the time was too short for many new mutations to occur, which could have reconstituted genome wide variability. Accordingly, we can exclude that demographic or historical population parameters influence our conclusions.

The shape of the sweep signature window is an additional important indicator for whether the signature reflects a true selective sweep. If local selection was the reason for a reduction in variability, one would expect that a closely flanking locus should also show some reduced variability (Maynard Smith and Haigh 1974, Kaplan et al. 1989, Kim and Stephan 2003). Taking this reasoning, we infer that at least the four loci where flanking variability was also lost (Figure 4.5) reflect true selective sweeps. On the other hand, we would expect for older sweeps that the signature is blurred and detected candidate loci are more isolated from flanking regions, i.e. the above criterion would be less likely to apply.

Fixation of alleles by neutral drift can be excluded as an explanation for variability reduction because of two major parameters: the age of the populations and the effective population size in mice. New alleles go to fixation by random drift after $4N_e$ generations. Therefore, with a given N_e of about 10,000, it would take about 40,000 generations, i.e. 13,000 years to fix new alleles. Fixation of existing alleles with a starting frequency of e.g. 0.5 roughly takes 12,000 generations, 4,000 years (Kimura and Ohta 1969). The age of the two investigated populations is not more than 9,000 generations (Karn et al. 2002, Cucchi et al. 2005), probably much younger, thus drift as a reason for the detected sweep candidates is fairly improbable.

lnRH values fluctuate within the candidate regions. These fluctuations seem to be due to differences in the mutation rates of the different investigated microsatellite loci. The mutation rates in microsatellites of humans and mice vary between 10^{-5} - 10^{-2} (see Schug et al. 1998). In the candidate region PP9A07 (Figure 4.5) loci PP9A07_4 and PP9A07_5, show much more variability than the two outer most extreme reduced loci (PP9A07_Sweep and PP9A07_6). By analysing the allele frequencies of these loci, we can conclude that the two inner microsatellites have a higher mutation rate than the two outer ones, and therefore recovered variability

within a shorter time (see Results 4.4.3). The extreme reduction in variability is therefore still present in the two loci with a lower mutation rate, whereas the two high mutating microsatellites already accumulated new alleles after the sweep event. Especially locus PP9A07_4 clearly depicts an accumulation of new alleles by stepwise mutation from the most frequent one. Applying different statistical tests (Variance in lnRH and F-statistic), we were able to show that the candidate regions significantly differ from neutrally evolving areas of about the same size. Thus, we reject neutrality as an explanation for the detected signatures.

Importantly, we have good reason to believe that we have in fact underestimated the true rate. Specifically, we want to point out that this number states a minimum of adaptive events for the following reasons: (1) with stringent statistics, we would detect only the most extreme values in a given window; we would not have detected most "soft sweeps" [i.e. positive selection events on alleles from the standing variation that were linked to more than one major allele of a flanking microsatellite (Hermisson and Pennings 2005)], and we would have missed relatively recent selection events because they would not have had time to build up a significant sweep signature, at least under the assumption of generally low selection coefficients; (2) the estimated number of generations per year (some authors postulate two generations per year in mice), and (3) the assumed divergence time (maximum of 3,000 years, but probably less).

Taken together, these factors lead to an underestimate of the true rate and are likely to change the above result, possibly by up to an order of magnitude. Previous estimates of the frequency of selection have been made by comparison of genome data in *Drosophila* (Smith and Eyre-Walker 2002, Andolfatto 2005). While the estimated rate of amino acid substitution driven by natural selection is about 1 in 450 generations from a comparison of *D. simulans* and *D. yakuba* (Smith and Eyre Walker 2002), or even one amino acid substitution every 800 ± 350 generations (Birne and Eyre-Walker 2004), the estimate that includes positive selection on non-coding DNA is up to 10 in 100 generations (Andolfatto 2005). Bazykin et al. (2004) have studied genome data from mouse and rat in this respect and also suggest that positive selection must be higher than previously anticipated. In addition to the adaptive events that go along with amino acid substitutions, we are able to detect adaptive changes that result from *cis* regulatory modifications. Such adaptive events cannot be

identified in comparative sequence analysis of old, diverged species. The footprints of reduced variability which go along with both types of evolutionary changes are still present in the young investigated populations, and can be identified in our genome screen for signatures of selective sweeps. Therefore we expect our result to reveal a higher frequency. Thus, although we compare different populations rather than different species and use a very different approach, we arrive at similar conclusions. In fact, it has been shown by Maynard-Smith (1968) that selection frequencies of 1 in 10 generations may be feasible, if one takes epistasis effects into account, although this issue is still not resolved (Nei 2005). It has been noted that such a high rate of positive selection in natural populations would have consequences for the estimation of population genetical parameters under the neutral model (Maynard Smith and Haigh 1974, Gillespie 2000).

By analysing the pool pattern of the other than the two Western European populations, we are able to detect about the same number of sweep candidates in an old ancestral population, Kazakhstan. This leads us to the conclusion that positive selection events do not necessarily occur exclusively with strong changes in the environment, e.g. with invasions into new habitats, but that selective sweeps seem to be a continuous background process that can be observed in all populations. By comparing the pool patterns of the two *M. m. musculus* populations (from Kazakhstan and the Czech Republic), we identified two unambiguous candidates for Kazakhstan specific signatures of selective sweeps. Thus, positive selection events are not exclusively present in populations that have recently undergone environmental changes, but can also be observed in old ancestral ones. Similar to these results, Schlötterer (2002b) found the same number of significant signatures of selective sweeps in the African and the non-African human population, which is in contrast to the model of out-of-Africa-associated adaptive mutations, but indicating that also in humans selection is not exclusively connected to environmental changes. Similar results were found in a recently published paper about a SNP based analysis in humans, were strong selection was detected in a Sub-Saharan African population (Voight et al. 2006). In contrast to this, Kayser et al. (2003) investigated 332 microsatellite loci and identified ten selective sweep loci in a European population but only one in an African population. In the recently published paper by Haddrill et al. (2005) the authors postulate that their results imply that highly reduced variation

observed at some loci in genome wide scans, e.g. Harr et al 2002, Glinka et al 2003, Kauer et al 2003, might be more easily explained by a bottleneck in the history of European populations than by recurrent selective sweeps. Thus, the excess of sweep signatures in derived populations in comparison to ancient ones might be an artefact. We therefore postulate that adaptive events can be observed in all populations more or less equally and that adaptation is not restricted to changes in the environment, it is a continuous background process.

Based on the size of the identified candidate regions and the age of the populations, we can conclude that the selective pressure acting on these loci is small. None of the windows is very broad, implying that the selection coefficients are roughly in balance with the recombination rates, effectively isolating positive selection on one locus from neighbouring loci. The calibration scale for values of selection coefficient 's' is the effective population size. In small populations selection has to be stronger to act efficiently due to the larger impact of drift (Maynard Smith and Haigh 1974). Positive selection is acting if $4N_e s > 1$. Given N_e equals 10,000, a minimal s of $2.5 \cdot 10^{-5}$ is required for the presence of positive selection. This number describes the lowest possible selection coefficient that can act efficiently in a house mouse population. The minimal required selection coefficient to fix an allele within the known maximal divergence time of 9,000 generations is $2.2 \cdot 10^{-3}$, estimated from $t=(2/s) \cdot \ln(2N_e)$ (Stephan et al. 1992). Using a formula proposed by Stephan et al. (1992) one can estimate selection coefficients for the respective sweep valleys based on the measures of heterozygosity and local recombination rate (see Methods). The results for our candidate regions range from 0.0007 to 0.021 (Table 4.5). Given the absence of large chromosomal regions showing reduced variability and the order of magnitude of our calculated s-values, we hypothesize that adaptive events are generally associated with small selection coefficients. There is no evidence for very strong selection. This confirms long standing speculations that strong selection on single loci is likely to be rare because of pleiotropic effects and possibility of linkage to maladaptive alleles in flanking loci (Barton and Partridge 2000).

Artificial selection experiments show invariably that any polymorphic population harbours enough standing variation to allow the selection of traits in almost every conceivable direction (Barton and Partridge 2000). If most of the selection in natural populations would also be derived from the standing variation

rather than new mutations, we would expect that signatures of selective sweeps are blurred because a beneficial mutation could be linked to more than one microsatellite allele. In fact, we find that none of the eight preselected candidate loci that we identified shows a single fixed allele. There are always one or more low frequency alleles present (see Figure 4.4). These alleles are usually distant from the major allele suggesting that they have not arisen via new mutations after fixation of one allele, since new mutations at microsatellite loci are expected to occur predominantly in a step-wise fashion, i.e. new alleles would flank the major one.

There is the potential that such alleles appear due to recombination events during the fixation phase, although this is unlikely because of their low frequency. Recombination takes place in the early phase of a selective sweep. Thus, alleles that become linked to the beneficial variant in this phase should be at a higher frequency at the end of a selective sweep. Additionally, since we detect the distal alleles consistently in all eight significant outliers, recombination becomes unlikely because this would mean all detected loci would have undergone similar recombination events leading to low frequency of additional alleles. It is much more likely that there is a general mechanism behind these consistent results. Thus, our data are compatible with the assumption that beneficial alleles come from the standing variation, although we can not exclude new mutations either.

We postulate that positive selection events occur with a minimal frequency of one sweep every 100 generations, and are generally associated with small selection coefficients. Strong selection is rare, which is in line with previously assumed consequences of pleiotropic effects and an accumulation of negative mutations that might hitchhike on large chromosomal regions which are fixed during a sweep. Furthermore, we conclude that beneficial variants are generally taken from the standing variation and that positive selection events are not exclusively linked to changes in the environment, but represent a continously acting background process in all populations. Populations are never perfectly adapted, adaptation is always on its way to optimization.

4.6 Acknowledgements

We thank Sonja Ihle and Susanne Kipp for providing the population samples as well as Thomas Wiehe and Bettina Harr for discussions and suggestions on the manuscript. Daniel Zivkovic helped us with the interpretation of our data regarding the bottleneck scenario. This work was supported by a grant of the Volkswagen Stiftung.

4.7 References

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature **437**(20): 1149-1152.

Bazykin, G.A., Kondrashov, F.A., Ogurtsov, A.Y., Sunyaev, S. and Kondrashov, A.S. 2004. Positive selection at sites of multiple amino acid replacements since rat-mouse divergence. Nature **429**: 558-562.

Barton, N. and Partridge, L. 2000. Limits to natural selection. BioEssays **22**, 1075-1084

Beisswanger, S., Stephan, W. and De Lorenzo, D. 2005. Evidence for a selective sweep in the *wapl* region of *Drosophila melanogaster*. Genetics Oct 3; [Epub ahead of print].

Begun, D.J. and Aquadro, C.F. 1993. African and North American populations of Drosophila melanogaster are very different at the DNA level. Nature **365**: 548-550.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research **27**(2): 573-580.

Birne, N. and Eyre-Walker, A. 2004. The genomic rate of adaptive amino acid substitutions in Drosophila. Molecular Biology and Evolution **21**(7): 1350-1360.

Boursot, P., Auffray, J.C., Britton-Davidian, J. and Bonhomme, F. 1993. The Evolution of House Mice. Trends in Ecology and Evolution **24**: 119-152.

Castillo-Davis, C.I. 2005 The evolution of noncoding DNA: how much junk, how much finc? Trends in Genetics **21**(10): 533-536.

Chin, C.-S., Chuang, J.H. and Li, H. 2005. Genome wide regulatory complexity in yeast promoters: separation of functionally conserved and neutral sequence. Genome Research **15**: 205-213

Cucchi, T., Vigne, J.-D. and Auffray, J.-C. 2005. First occurrence of the house mouse (Mus musculus domesticus Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. Biological Journal of the Linnean Society **84**: 429-446**.**

David, J.R. and Capy, P. 1988. Genetic variation of Drosophila melanogaster natural populations. Trends in Genetics **4**: 106-111.

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

Eyre-Walker, A. 2002. Changing the effective population size and the McDonald-Kreitman Test. Genetics **162**: 2017-2024.

Fay, J.F. and Wu, C.-I. 2001 The neutral theory in the genomic era. Current Opinion in Genetics and Developement **11**: 642-646

Fay, J.C., Wyckoff, G.J. and Wu, C.-I. 2002. Testing the neutral theory of molecular evolution with genomic data from Drosophila. Nature **415**: 1024-1026.

Gillespie, J. H. 2000. Genetic drift in an infinite population. The pseudo-hitchhiking model. Genetics **155:** 909-919

Glinka, S., Ometto, L., Mousset, S., Stephan, W. and de Lorenzo, D. 2003. Demography and natural selection have shaped genetic variation in Drosophila melanogaster: A multi-locus approach. Genetics **165**: 1269-1278.

Guénet, F. and Bonhomme, F. 2003 Wild mice: an ever increasing contribution to a popular mammalian model. Trends in Genetics **19**: 24-31.

Haddrill, P.R., Thornton, K.R., Charlesworth, B. and Andolfatto, P. 2005 Multilocus patterns of nucleotide variability and the demographic and selection history of Drosophila melanogaster populations. Genome Research **15**: 790-799.

Haldane, J.B.S. 1957. The Cost of natural selection. Genetics **55**: 511-524.

Harr, B., Kauer, M. and Schlötterer, C. 2002. Hitchhiking mapping – a population based fine scale mapping strategy for adaptive mutations in Drosophila melanogaster. Proceedings of the National Academy of Sciences, USA **99**: 12949-12954.

Hermisson, J. and Pennings, P. S. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics **169**: 2335-2352.

Huttley, G.A., Smith, M.W., Carrington, M. and O'Brien S.J. 1999. A scan for linkage disequilibrium across the human genome. Genetics **152**:1711-1722.

Ihle, S., Ravaoarimanana, I., Thomas, M. and Tautz, D. 2006 Tracing signatures of selective sweeps in natural populations of the house mouse. Molecular Biology and Evolution **23**(4):790-797.

Jensen-Seaman, M.I., Furey, T.S., Payseur, B.A., Lu, Y., Roskin, K.M., Chen, C.F., Thomas, M.A., Haussler, D. and Jacob, H.J. 2004. Comparative recombination rates in the rat, mouse, and human genomes. Genome Research **14**(4):528-38.

Kaplan, M.L., Hudson, R.R. and Langley, C.H. 1989. The "hitchhiking effect" revisited. Genetics **123**: 87-899.

Karn, R.C, Orth, A., Bonhomme, F. and Boursot, P. 2002. The Complex History of a Gene Proposed to Participate in a Sexual Isolation Mechanism in House Mice. Molecular Biology and Evolution **19**:462-471

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

Kayser, M., Bauer, S. and Stoneking, M. 2003. A genome scan to detect candidate regions influenced by local natural selection in human populations. Molecular Biology and Evolution **20**(6): 893-900.

Keightley, P.D., Lerchen, M.J. and Eyre-Walker, A. 2005. Evidence for widespread degradation of gene control regions in hominid genomes. PLOS **3**(2): e42.

Kim, Y. and Stephan, W. 2003. Selective Sweeps in the presence of interference among partially linked loci. Genetics **164**: 389-398.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge UK.

Kimura, M. and Otha, T. 1969. The average number of generations until fixation of a mutant gene in a finite population. Genetics **61**:763-771.

Kohn, M.H., Pelz, H.-J. and Wayne, R.K. 2000. Natural selection mapping of the warfarin-resistance gene. Proceedings of the National Academy of Sciences, USA **97**(14): 7911-7915.

Lachaise, D., Cariou, M.L., David, J.R., Lemeunier, F., Tsacas, L. and Ashburner, M. 1988. Historical biogeography of the Drosophila melanogaster specied subgroup. Evolutionary Biology **22**: 159-225.

Lewontin, R.C. 1974. The Genetic Basis of Evolutionary Change. Columbia University Press, New York.

Maynard Smith, J. and Haigh, J. 1974. The hitch-hiking effect of a favourable gene. Genetic Research, Camb. **23**: 23-35.

Maynard Smith, J. 1968. "Haldane´s dilemma" and the rate of evolution. Nature **219**: 1114-1116.

Mouse Genome Sequencing Consortium 2002. Initial sequencing and comparative analysis of the mouse genome. Nature **420**(6915): 520-62.

Nei, M. 2005. Seletionism and neutralism in molecular evolution. Molecuar Biology and Evolution **22**: 2318-2342.

Ohta, T. and Kimura, M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetic Research **22**: 201-204.

Ohta, T. and Kimura, M. 1975. The effect of a selected linked locus on heterozygosity of neutral alleles (the hitchhicking effect). Genetic Research **25**: 313-325.

Payseur, B.A., Cutter, A.D. and Nachman, M.W. 2002. Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. Molecular Ecology and Evolution **19**: 1143-1153.

Przeworski, M. 2002. The signature of positive selection at randomly chosen loci. Genetics **160**: 1179-1189.

Schlötterer, C. 2002. Towards a molecular characterization of adaption in local populaitons. Current Opinion in Genetics and Developement **12**: 683-687.

Schlötterer, C. 2002b. A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics **160**: 753-763.

Schlötterer, C. 2003. Hitchhiking mapping functional genomics from the population genetics perspective. Trends in Genetics **19**: 32-38.

Schlötterer, C., Vogl, C. and Tautz, D. 1997. Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural Drosophila melanogaster populations. Genetics **146**: 309-320.

Schlötterer, C. and Wiehe, T. 1999. Microsatellites, a neutral marker to infer selective sweeps. In: Goldstein, D.B., and Schlötterer, C. (eds.) Microsatellites Evolution and Applications. Oxford. pp. 238-247.

Schöfl, G. and Schlötterer, C. 2004. Patterns of microsatellite variability among X chromosomes and autosomes indicated a high frequency of beneficial mutations in non-African D. simulans. Molecular Biology and Evolution **21**(7): 1384-1390.

Schug, M.D., Hutter, C.M., Noor, M.A.F. and Aquadro, C.F. 1998. Mutation and evolution of microsatellites in Drosophila melanogaster. Genetica **102/103**: 359-367.

Slatkin, M. 1995. Hitchhiking and associative overdominance at a microsatellite locus. Molecular Biology and Evolution 12: 473-480.

Smith, N.G. and Eyre-Walker, A. 2002. Adaptive protein evolution in Drosophila . Nature **415**: 1022-1024.

Stajich, J.E. and Hahn, M.W. 2005. Disentangling the effects of demography and selection in human history. Molecular Biology and Evolution **22**(1): 63-73.

Stephan, W., Wiehe, T.H.E. and Lenz, M. 1992. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. Sci. **41**: 237-254.

Storz, J.F., Payseur, B.A. and Nachman, M.W. 2004. Genome scan for DNA variability in humans reveal evidence for selective sweeps outside of Africa. Molecular. Biology and Evolution **21**(9): 1800-1811.

Tautz, D. 2000. Evolution of transcriptional regulation. Curr Opin Genet Dev. **10**(5):575-9.

Vigouroux, Y., Mullen, M., Hittinger, C.T., Houchings, L. and Schlulz, L. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proceedings of the National Academy of Sciences, USA **99**: 9650-9655.

Vigouroux, Y., Mitchell, S., Matsuoka, Y., Hamblin, M., Kresovich, S., Smith, J.S., Jaqueth, J., Smith, O.S. and Doebley, J. 2005. An analysis of genetic diversity across the maize genome using microsatellites. Genetics **169**: 1617-1630.

Voigtht, B.F., Kudaravalli, S., Wen, X. and Pritchard, J.K. 2006 A map of recent positive selection in the human genome. Plos Biology **3**(72): 0446-0458.

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the Analysis of Population Structure, Evolution **38**(6): 1358-1370

Wittkopp, P.J., Härum, B.K. and Clark, A.G. 2004. Evolutionary changes in cis and trans gene regulation. Nature **430**: 85-88

Wootton, J.C., Feng, X., Ferdig, M.T., Cooper, R.A, Mu, J., Baruch, D.I., Magill, A.J. and Su, X.Z. 2002. Genetic diverstiy and chloroquine selective sweeps in Plasmodium falsiparum. Nature **418**: 320-323.

Zuckerkandl, E. and Pauling, L. 1965. Molecules as documents of evolutionary history. Journal of Theoretical Biology **8**(2): 357-3

5 Supplement

- Supplement 1 Calculation of selection coefficiens
- Supplement 2 Primer sequences of the 960 screened microsatellite loci.
- Supplement 3 Primer sequences of the microsatellite loci genotyped along the candidate regions.
- Supplement 4 Primer sequences of the loci genotyped along the 'neutral regions'.
- Supplement 5 Allele frequencies of the loci genotyped in candidate region PP10B07 (German sweep).
- Supplement 6 Allele frequencies of the loci genotyped in candidate region PP7F04 (German sweep).
- Supplement 7 Allele frequencies of the loci genotyped in candidate region PP5F12 (German sweep).
- Supplement 8 Allele frequencies of the loci genotyped in candidate region PP4H12 (French sweep).
- Supplement 9 Allele frequencies of the loci genotyped in candidate region PP3H03 (French sweep).
- Supplement 10 Allele frequencies of the loci genotyped in candidate region PP6E03 (French sweep).

Supplement 1 Calculation of selection coefficiens

A For the regions that show a signature of a selective sweep in the German population.

Locus	Het.Ger	Het.Fra	lnRH	Distance to sweep locus [bp]	RR [5 Mb]	recombination distance [M]	HGer/HFra	Selection coefficient per locus	average selection coefficient
PP9A07_12	0.7147	0.8367	-1.4700	-92277	0.4400	0.0004	0.8541		
PP9A07_14	0.6248	0.6594	-0.3604	-78814	0.4400	0.0003	0.9475		
PP9A07_16	0.7959	0.7684	0.2071	-71873	0.4400	0.0003	1.0358		
PP9A07_17	0.8005	0.9073	-1.9265	-65489	0.4400	0.0003	0.8822	0.0027	
PP9A07_18	0.9056	0.9343	-0.9513	-65029	0.4400	0.0003	0.9693	0.0016	
PP9A07_3	0.8571	0.8982	-0.9028	-13653	0.4400	0.0001	0.9543	0.0004	
PP9A07_S	0.2756	0.8968	-5.4968	0	0.4400	0.0000	0.3073		
PP9A07_4	0.6839	0.8634	-2.1569	2799	0.4400	0.0000	0.7921	0.0002	
PP9A07_5	0.8924	0.9584	-2.3268	29056	0.4400	0.0001	0.9311	0.0009	
PP9A07_6	0.4986	0.9179	-4.6460	31187	0.4400	0.0001	0.5432	0.0035	
PP9A07_8	0.5772	0.7815	-1.8129	47037	0.4400	0.0002	0.7385	0.0031	
PP9A07_9	0.8559	0.8316	0.2697	87731	0.4400	0.0004	1.0292		
PP9A07_10	0.4495	0.3411	0.5599	92601	0.4400	0.0004	1.3180		0.0018
PP10B07_14	0.8745	0.8542	0.2540	-95949	0.3424	0.0003	1.0238		
PP10B07_15	0.3474	0.6864	-2.3355	-76769	0.3424	0.0003	0.5061		
PP10B07_17	0.8948	0.8390	0.9080	-59764	0.3424	0.0002	1.0666		
PP10B07_18	0.7470	0.8565	-1.4756	-51090	0.3424	0.0002	0.8722	0.0017	
PP10B07_4	0.4444	0.7134	-1.9749	-41997	0.3424	0.0001	0.6228	0.0029	
PP10B07_6	0.6173	0.7666	-1.3730	-35067	0.3424	0.0001	0.8053	0.0015	
PP10B07_S	0.3718	0.9244	-5.6120	0	0.3424	0.0000	0.4023		
PP10B07_3	0.5935	0.6842	-0.7778	7943	0.3424	0.0000	0.8675	0.0003	
PP10B07_2	0.8399	0.9334	-2.1706	35086	0.3424	0.0001	0.8998	0.0010	
PP10B07_9	0.8527	0.8910	-0.8146	84872	0.3424	0.0003	0.9570	0.0018	0.0015

Locus	Het.Ger	Het.Fra	lnRH	Distance to sweep locus [bp]	RR [5 Mb]	recombination distance [M]	HGer/HFra	Selection coefficient per locus	average selection coefficient
PP7F04_5	0.6940	0.6107	0.5357	-77064	0.8094	0.0006	1.1363		
PP7F04_7	0.7867	0.8390	-0.7801	-66322	0.8094	0.0005	0.9377	0.0038	
PP7F04_2	0.8197	0.8699	-0.8803	-35473	0.8094	0.0003	0.9423	0.0020	
PP7F04_S	0.3628	0.9084	-5.2187	0	0.8094	0.0000	0.3994		
PP7F04_3	0.8741	0.7965	1.0472	13444	0.8094	0.0001	1.0974		
PP7F04_4	0.5366	0.7399	-1.6474	22567	0.8094	0.0002	0.7252		0.0029
PP5F12_3	0.8571	0.7661	1.0880	-16790	1.0220	0.0002	1.1188		
PP5F12_4neu	0.5394	0.7871	-2.1240	-11869	1.0220	0.0001	0.6853	0.0021	
PP5F12_8	0.8707	0.9048	-0.8235	-3133	1.0220	0.0000	0.9623	0.0002	
PP5F12_S	0.4545	0.8949	-4.3378	0	1.0220	0.0000	0.5079		
PP5F12_5	0.9159	0.8706	0.9135	12688	1.0220	0.0001	1.0520		
PP5F12_6	0.8673	0.8320	0.4607	23867	1.0220	0.0002	1.0425		
PP5F12_7	0.9020	0.8899	0.1738	40554	1.0220	0.0004	1.0137		0.0011

B For the regions which show a signature of a selective sweep in the French population.

Locus	Het.Ger	Het.Fra	lnRH	Distance to sweep locus [bp]	RR [5 Mb]	recombination distance [M]	HFra/HGer	Selection coefficient per locus	average selection coefficient
PP8D01_18	0.8327	0.5910	-2.1614	-83063	2.0560	0.0017	0.7097	0.0274	
PP8D01_20	0.8588	0.7661	-1.1160	-68537	2.0560	0.0014	0.8921	0.0125	
PP8D01_7	0.6564	0.5450	-0.6763	-26312	2.0560	0.0005	0.8303	0.0060	
PP8D01_1	0.8589	0.7563	-1.2195	-7993	2.0560	0.0002	0.8805	0.0015	
PP8D01_S	0.7353	0.1277	-4.2583	0	2.0560	0.0000	0.1737		
PP8D01_2	0.6314	0.0878	-3.9182	7725	2.0560	0.0002	0.1390	0.0210	
PP8D01_3	0.9188	0.6032	-3.7878	25623	2.0560	0.0005	0.6564	0.0098	
PP8D01_4	0.4447	0.0703	-2.9970	30023	2.0560	0.0006	0.1580	0.0711	
PP8D01_10	0.7925	0.6452	-1.2539	51519	2.0560	0.0011	0.8141	0.0125	
PP8D01_11	0.6743	0.6245	-0.2755	55816	2.0560	0.0011	0.9262	0.0087	
PP8D01_12	0.7151	0.2893	-2.7489	59640	2.0560	0.0012	0.4045	0.0469	
PP8D01_15	0.3569	0.2676	-0.4749	83325	2.0560	0.0017	0.7498	0.0245	
PP8D01_16	0.9139	0.8893	-0.4893	88546	2.0560	0.0018	0.9731	0.0100	0.0210
PP4H12_8	0.5720	0.6068	0.3399	-44850	0.2200	0.0001	1.0609		
PP4H12_1	0.9321	0.7291	-3.2046	-43261	0.2200	0.0001	0.7822	0.0012	
PP4H12_10	0.8116	0.6277	-1.6168	-42515	0.2200	0.0001	0.7734	0.0012	
PP4H12_3	0.6371	0.4532	-1.1026	-37606	0.2200	0.0001	0.7114	0.0013	
PP4H12_12	0.8634	0.7604	-1.2529	-13468	0.2200	0.0000	0.8808	0.0003	
PP4H12_2	0.8973	0.7150	-2.3633	-6031	0.2200	0.0000	0.7968	0.0002	
PP4H12_S	0.5389	0.0250	-4.8689	0	0.2200	0.0000	0.0464		
PP4H12_5	0.5158	0.4747	-0.1526	10427	0.2200	0.0000	0.9204	0.0002	
PP4H12_4	0.7957	0.7908	0.0435	21814	0.2200	0.0000	0.9938	0.0002	0.0007

Locus	Het.Ger	Het.Fra	lnRH	Distance to sweep locus [bp]	RR [5 Mb]	recombination distance [M]	Hfra/HGer	Selection coefficient per locus	average selection coefficient
PP3H03_2	0.8454	0.8507	-0.1866	-37472	0.4400	0.0002	1.0064		
PP3H03_3	0.8864	0.7886	1.3824	-31256	0.4400	0.0001	0.8897	0.0012	
PP3H03_S	0.7092	0.1187	4.1250	0	0.4400	0.0000	0.1673		
D10Mit23	0.7427	0.5547	1.3532	5638	0.4400	0.0000	0.7469	0.0004	
PP3H03_5	0.8304	0.8315	-0.1178	25420	0.4400	0.0001	1.0013		
PP3H03_6	0.8804	0.8156	0.9294	36108	0.4400	0.0002	0.9264		
PP3H03_7	0.8055	0.7926	0.0547	40928	0.4400	0.0002	0.9839		0.0008
D1Mit305	0.8359	0.9030	1.3470	-24296	0.5094	0.0001	1.0802		
PP6E03_1	0.6688	0.7155	0.4930	-14634	0.5094	0.0001	1.0698		
PP6E03_3neu	0.7840	0.6643	-1.0091	-5921	0.5094	0.0000	0.8473	0.0003	
PP6E03_S	0.7247	0.0737	-4.9066	0	0.5094	0.0000	0.1017		
PP6E03_2	0.6251	0.5726	-0.2620	20642	0.5094	0.0001	0.9160	0.0008	
D1Mit440	0.7774	0.6487	-1.0559	32840	0.5094	0.0002	0.8344	0.0018	
PP6E03_5	0.8139	0.7620	-0.4974	58432	0.5094	0.0003	0.9363	0.0021	
PP6E03_6	0.7429	0.8153	0.9115	65638	0.5094	0.0003	1.0975		
PP6E03_7	0.7855	0.0977	-5.1503	67291	0.5094	0.0003	0.1244		
PP6E03_9	0.9317	0.8778	-1.2640	92255	0.5094	0.0005	0.9422		0.0013

Supplement 2 Primer sequences of the 960 screened microsatellite loci.

Label		Marker	F-Primer	Sequence	R-Primer	Sequence
PP1	A01	D6Mit115	D6Mit115F	ccatttaataagtgatccctctgg	D6Mit115R	tgtcacaccacaatgggc
PP1	B01	D6Mit364	D6Mit364F	tagaccttgtctcaaatgtatgtgtg	D6Mit364R	cccctgatgctgtaggtgtt
PP1	C01	D6Mit138	D6Mit138F	gctcttattaatgaagaagaaggagg	D6Mit138R	caaagaaagcatttcaagactgc
PP1	D01	D6Mit139	D6Mit139F	atagaaggcgagaactaacccc	D6Mit139R	tgtttctgcccctgtagttg
PP1	E01	D6Mit365	D6Mit365F	gtctggtgtatttgcatatatggg	D6Mit365R	gcaggcagacacacagacaa
PP1	F01	D6Mit391	D6Mit391F	ttctctcagtcttgtctgtgtaca	D6Mit391R	gtgaggctcaaagaaagggc
PP1	G01	D6Mit73	D6Mit73F	atacactttgacacaagagcaagg	D6Mit73R	agcacagaggtcagaaactgc
PP1	H01	D6Mit326	D6Mit326F	tgactggaggacagagattgg	D6Mit326R	atgtccatttaagtcttttctggg
PP1	A02	D6Mit38	D6Mit38F	cttagtgcgtgtaaggcaagg	D6Mit38R	gactgctgagctagtgccct
PP1	B02	D6Mit199	D6Mit199F	gcccttcctactcaaaataaatacc	D6Mit199R	gggtagttcaaattaaacctggg
PP1	C02	D6Mit175	D6Mit175F	gttagtgagatccaaagccacc	D6Mit175R	gccaccatctcaaccctg
PP1	D02	D6Mit210	D6Mit210F	ttagaggaagagaactgatagaatgtg	D6Mit210R	attaacttcaaggagaagcccc
PP1	E02	D6Mit315	D6Mit315F	agaaaatagagtcgatctagacacaca	D6Mit315R	atgaaagtcaggttgtcggg
PP1	F02	D6Mit333	D6Mit333F	tcctcactacaattcatctattactgc	D6Mit333R	tgcttctggtataggcagttagg
PP1	G02	D6Mit200	D6Mit200F	catcaggtgtcttcaggttctg	D6Mit200R	tcccctctatccttactgttgc
PP1	H02	D6Mit288	D6Mit288F	agcactggctagagaatcatcc	D6Mit288R	cattcgactcttcaggccat
PP1	A03	D6Mit302	D6Mit302F	aatgaccctggttagtgtcagg	D6Mit302R	gaattccattcgaggggc
PP1	B03	D6Mit254	D6Mit254F	agtgtccctagggggtgg	D6Mit254R	ggggccttagaggtagcaac
PP1	C03	D6Mit44	D6Mit44F	cccgtgtccagggtactg	D6Mit44R	gcatggtaccccagcttcta
PP1	D03	D6Mit193	D6Mit193F	tagtacagaaagttattctggatgtgg	D6Mit193R	taagccacactgattgatgtcc
PP1	E03	D6Mit362	D6Mit362F	gaaaactgtcctctgacatttataagc	D6Mit362R	ttgatcttgagggtttaaaatcg
PP1	F03	D6Mit253	D6Mit253F	gtcaacatctatgttccactcagg	D6Mit253R	ctcatgcacgtctatacacaagc
PP1	G03	D6Mit308	D6Mit308F	ttactagagaacttgggagaaccg	D6Mit308R	ctactgtcgccacctaacctg
PP1	H03	D6Mit119	D6Mit119F	gggctagtttctcatgaagtaagc	D6Mit119R	tacattttatcactaggtgaatgtgtg
PP1	A04	D6Mit29	D6Mit29F	cttctttacacctgtatggcacc	D6Mit29R	ggttggtcactgcaggagtt
PP1	B04	D6Mit323	D6Mit323F	gactgtaaatggatccttcctcc	D6Mit323R	acctaacaacctacttcaagaaaagc
PP1	C04	D6Mit46	D6Mit46F	ttagaactgtgaagagggtcagc	D6Mit46R	tggctgtttgttaattcgacc
PP1	D04	D6Mit141	D6Mit141F	cacctctcaccaccaccac	D6Mit141R	acctgcatttcccacacc
PP1	E04	D6Mit153	D6Mit153F	attctgagtatatgagcctctggg	D6Mit153R	gatgttactaagtaagatacttcgccg
PP1	F04	D6Mit21	D6Mit21F	ctgggattaaagactaccatgagc	D6Mit21R	cacctgactctaatccctgtcc
PP1	G04	D3Mit215	D3Mit215F	taaacatctagaagatgctgcagg	D3Mit215R	ctgcatggccaggactagtt
PP1	H04	D3Mit240	D3Mit240F	ccaattgggataagatactgttcc	D3Mit240R	gtaacgtaaccttactgctgtgtagg
PP1	A05	D3Mit286	D3Mit286F	tcctactggaatccaatacaagtg	D3Mit286R	gtggtgtgtcttgacattctatatcc
PP1	B05	D6Mit5	D6Mit5F	cacggagaggacctacatgc	D6Mit5R	agctgctcgtctccacactt
PP1	C05	D6Mit55	D6Mit55F	caagcagagcacctagggtc	D6Mit55R	tcactgcaggattggtcttg
PP1	D05	D6Mit280	D6Mit280F	acagcaggtagactctgtacaatca	D6Mit280R	gcataaagtgaggtttttatttgtaca
PP1	E05	D6Mit300	D6Mit300F	tgagatcctacctcattacacttcc	D6Mit300R	ttggttcatagcctttcatgg
PP1	F05	D6Mit372	D6Mit372F	ttaatacacttaggtgtggctctcc	D6Mit372R	gagaggcatatagaaaaggataatgc
PP1	G05	D3Mit160	D3Mit160F	caatgacacagcatgtctaacttg	D3Mit160R	ttaagaactactcaatgctgcacc
PP1	H05	D3Mit99	D3Mit99F	ggagactaaaatgcaagactgatg	D3Mit99R	cctcacatgtaattagcacatgc
PP1	A06	D3Mit293	D3Mit293F	acctgcacgtattttatgtctactc	D3Mit293R	gagtgtgcatgtatattttgtttgg
PP1	B06	D3Mit307	D3Mit307F	gaacttgtagcagaaagaaagatgg	D3Mit307R	tggtattttaagaatcgctcataca
PP1	C06	D3Mit155	D3Mit155F	aagcaacattatacagactgaacagg	D3Mit155R	aagttcacagcctccccc
PP1	D06	D3Mit305	D3Mit305F	gaatctaacttggggacctgc	D3Mit305R	tacagaaaggtgattggagttgg
PP1	E06	D3Mit323	D3Mit323F	tatacacacactcgagcacgc	D3Mit323R	gctaaagtaacagagtgacctttgg
PP1	F06	D3Mit42	D3Mit42F	tgacctccagagagtcttcca	D3Mit42R	taagcagctgagactcaagtgg
PP1	G06	D3Mit251	D3Mit251F	atccatacacacagagacatacaca	D3Mit251R	aacagggactagtgtggagtaagc

PP1	H06	D3Mit279	D3Mit279F	ttacccctctctttctttaagtgtg	D3Mit279R	atccctagggcttcttgctc
PP1	A07	D3Mit283	D3Mit283F	ctcagcattgtgttttgtatacagc	D3Mit283R	cattcattttcatcaaagtatatttcc
PP1	B07	D3Mit194	D3Mit194F	gtttttgtgatgttacatactggacc	D3Mit194R	acaactcatccccctcctct
PP1	C07	D3Mit338	D3Mit338F	gtccaataccaagaagtcagcc	D3Mit338R	caatcctaatctgaaacagaattcc
PP1	D07	D3Mit353	D3Mit353F	tccccagtagtgagaaatcagg	D3Mit353R	tatgttcatgttcttccatttctagg
PP1	E07	D3Mit32	D3Mit32F	caccctggttaactcagaaagg	D3Mit32R	gcacttgtgtttcatgtcactg
PP1	F07	D3Mit226	D3Mit226F	gcatagaggacttctttgagaataca	D3Mit226R	tctgaactttgctgcccc
PP1	G07	D3Mit278	D3Mit278F	aactaccatctaaaacatcctctgtg	D3Mit278R	agatccctagagaaacagaactgg
PP1	H07	D3Mit298	D3Mit298F	ttaacttctggcttctatacacatgc	D3Mit298R	acaggaacatgcacacgtgt
PP1	A08	D3Mit12	D3Mit12F	tagaccaatcttgggagtgtcc	D3Mit12R	ggaaaagcataagaaacaaccg
PP1	B08	D3Mit257	D3Mit257F	cctagcgcaggaatagttaacc	D3Mit257R	acaaacagaacaaaacaaaaagtcc
PP1	C08	D3Mit28	D3Mit28F	gatgagagattctgatgtggagg	D3Mit28R	ccagcctcagtatctcaaaacc
PP1	D08	D3Mit29	D3Mit29F	gatgagagattctgatgtggagg	D3Mit29R	ccagcctcagtatctcaaaacc
PP1	E08	D3Mit220	D3Mit220F	caggacactcagtgcccc	D3Mit220R	tctcaagcatcatgtgttcatg
PP1	F08	D3Mit25	D3Mit25F	gtctgggtcgtcagtggc	D3Mit25R	tggaggctaccatctccaag
PP1	G08	D3Mit269	D3Mit269F	cagtggtcttatctatctatttcacca	D3Mit269R	taacatgtgtatgtatgagtgtgtgtg
PP1	H08	D3Mit303	D3Mit303F	aggtcctaggttcttattttctacaca	D3Mit303R	atgcctgctctgtgtgtgtc
PP1	A09	D3Mit313	D3Mit313F	gaaatagtacagagaaacagacacaca	D3Mit313R	cattctaaaacatactgtgcaggg
PP1	B09	D3Mit331	D3Mit331F	atatcactgagtacacacatgattgtg	D3Mit331R	gagtacacttgatgtcaaccaacc
PP1	C09	D3Mit28	D3Mit28F	gatgagagattctgatgtggagg	D3Mit28R	ccagcctcagtatctcaaaacc
PP1	D09	D3Mit297	D3Mit297F	gtgctagcaaggaggatcttatg	D3Mit297R	ggtaaagatgaggaatatcagtctca
PP1	E09	D3Mit211	D3Mit211F	catgtactgtgtaagttcatgctcc	D3Mit211R	ctctacatatatgctgtgtcatgtgc
PP1	F09	D3Mit263	D3Mit263F	aacaggtggatgttctagtcatagc	D3Mit263R	caaccagggtctccctgata
PP1	G09	D3Mit342	D3Mit342F	ccagggatctacataactaatgtgc	D3Mit342R	ctgagattcctcaagggcag
PP1	H09	D3Mit351	D3Mit351F	attcacacagtgacaggtacacg	D3Mit351R	gacaaaaaggtgaaaactcatgg
PP1	A10	D3Mit28	D3Mit28F	gatgagagattctgatgtggagg	D3Mit28R	ccagcctcagtatctcaaaacc
PP1	B10	D3Mit103	D3Mit103F	ccaggggtggtggtcttac	D3Mit103R	tgtcaggtgcccaggtct
PP1	C10	D3Mit302	D3Mit302F	ccccttctctgtcctgctc	D3Mit302R	cagaagtctatgcatagatctcaagg
PP1	D10	D3Mit322	D3Mit322F	catgctgtggagtgtacatacttg	D3Mit322R	ccaagatgacccccctaacaa
PP1	E10	D3Mit332	D3Mit332F	agcctaagttagttgttcctgtgg	D3Mit332R	tgtaaaaaccagaggagacaagg
PP1	F10	D3Mit147	D3Mit147F	tctgcctctgttagatagatatccg	D3Mit147R	ttgttcatctatcctctgaagttcc
PP1	G10	D7Mit342	D7Mit342F	cagtgagaccctaacttaaaaaactg	D7Mit342R	gcaagtgctcttaactactagcaca
PP1	H10	D7Mit267	D7Mit267F	ctctttctgttacatggttagatttcc	D7Mit267R	aaagacagttgaagttgacttctgg
PP1	A11	D7Mit309	D7Mit309F	tgataaggaccctacaagcacc	D7Mit309R	cacagagatggacagatacagaca
PP1	B11	D7Mit279	D7Mit279F	agcaacatctcaccctataccc	D7Mit279R	tagtcctaccaacagcctgtca
PP1	C11	D7Mit347	D7Mit347F	ctgagttccctaaagtctgtttcc	D7Mit347R	aacccagtgtccaccct
PP1	D11	D7Mit163	D7Mit163F	ggacagacaccctcaccg	D7Mit163R	cggctgtgagagcatagtga
PP1	E11	D7Mit319	D7Mit319F	gatctgaagacacatgatctgagg	D7Mit319R	gttaaacttcagaagtgtgagactgc
PP1	F11	D7Mit16	D7Mit16F	ctggtctctgtccttggagc	D7Mit16R	aaagaaaatattcttgttgccagc
PP1	G11	D7Mit124	D7Mit124F	gtcagacattggcttaggatcc	D7Mit124R	ggtttgtgcgctctctctct
PP1	H11	D7Mit358	D7Mit358F	ttagtgtgcttaaggtctacacacg	D7Mit358R	aacaaatactagagacaaacgtgtgc
PP1	A12	D7Mit44	D7Mit44F	ttctggcctctgtgaagtagtg	D7Mit44R	gtgaaaccatggtgcagatg
PP1	B12	D7Mit166	D7Mit166F	tatataaatctatcaagacacaccccc	D7Mit166R	agtatctgattttgatctctggcc
PP1	C12	D7Mit361	D7Mit361F	tactaccaatgctaagcgctacc	D7Mit361R	accagagtttgtgcccattc
PP1	D12	D7Mit362	D7Mit362F	ggggttagaccagttgaccc	D7Mit362R	tttgcatatgtgtgtatacatgtgg
PP1	E12	D8Mit353	D8Mit353F	tctggcttgtagtatctgtattcaca	D8Mit353R	atgaaggtgtgtgtgtatatatgtgc
PP1	F12	D8Mit35	D8Mit35F	gagtcctacaagttgtgccttg	D8Mit35R	ggagataaggagacaagtgtccc
PP1	G12	D8Mit125	D8Mit125F	atcgctctatctactcatctattcaca	D8Mit125R	gaccctgactcttaatcctagtgc
PP1	H12	D8Mit341	D8Mit341F	ctggctgcatacatacagatatataca	D8Mit341R	tgattgcttgtaggtatacaggtaca
PP2	A01	X_2	X_2F	actctcaggaatgtgttgcgtc	X_2R	tggagagaccttgggctaccca
PP2	B01	X_3	X_3F	cttcgccaagcgggtctgcagc	X_3R	tccggtctccagtgcctgggtc

PP2	C01	X_4	X_4F	actaaagctgccgatccgcctg	X_4R	caaagacagcccaggcaccttc
PP2	D01	X_5	X_5F	ctcgcatactcaaatgacccac	X_5R	ctaataagcccgagtaggccac
PP2	E01	X_6	X_6F	gaatggcctctctcaggaggtg	X_6R	tctccttggccctgtgcttcag
PP2	F01	X_7	X_7F	tgcagggatactggctcaaacc	X_7R	agttccaggacccactagcctg
PP2	G01	X_8	X_8F	tcacatcttgaaccactagcag	X_8R	tgaatacttggtcctgagttgc
PP2	H01	X_9	X_9F	ttggagactgtaatcccaaacc	X_9R	aactagctgggtggtcacagtc
PP2	A02	X_12	X_12F	agaggccagtaaagctggaagg	X_12R	tggctggcagtgaactcgttgc
PP2	B02	X_13	X_13F	tccttgggaagactgaaccacc	X_13R	agctagtctgaagcagccacac
PP2	C02	X_14	X_14F	agcaagacaggatagcacaacg	X_14R	atctcagcacctgtcccaagtg
PP2	D02	X_15	X_15F	acctttggaagagcagttggtg	X_15R	tccttcttggattgcaggtaac
PP2	E02	X_16	X_16F	tgactgcctttacctcccaagg	X_16R	ggctagaagtatttagctcagtgg
PP2	F02	X_17	X_17F	tggctggaggtaacatctgagc	X_17R	ggaatcaccatgtggttgttgg
PP2	G02	X_18	X_18F	tggaatggtttacaggctttgg	X_18R	ttaaaggcgtgtgccaccactg
PP2	H02	X_19	X_19F	actgaggtagtcggagctgacc	X_19R	acggcacagtttactcctgagg
PP2	A03	X_20	X_20F	gtcctctgatgtcaacatgtgc	X_20R	acgctctgcttctgccttcagg
PP2	B03	X_21	X_21F	agtgagactctgcacccatctc	X_21R	agtggaagcaaacaggcaatcc
PP2	C03	X_23	X_23F	cacccacatcatggtagctcac	X_23R	agctggcgggagactggtttgg
PP2	D03	X_24	X_24F	tgatgccctcttctggagtgtc	X_24R	aggcaaacacacgatttgctgc
PP2	E03	X_26	X_26F	gggattggttctcgtgtcaggt	X_26R	acagagtgagttccaggacagc
PP2	F03	X_27	X_27F	tgaagcacctttctgactgagc	X_27R	tgagcaccataacaagccatcc
PP2	G03	X_29	X_29F	tgcctggggtgagagcaccctc	X_29R	gagagctttgtgcgtgtgtgtg
PP2	H03	X_30	X_30F	tctgaacctgtaagccagctcc	X_30R	tccaggactttcagggctacac
PP2	A04	X_32	X_32F	ttgtggctcacaaccagcaacc	X_32R	agacagcccttgagcccagcag
PP2	B04	X_34	X_34F	acaactggagtggcagctatcc	X_34R	actgagtgaaccaccaatgctg
PP2	C04	X_35	X_35F	tgtgcttaggaaaccttgtgtg	X_35R	acgctgattggcttcccagcag
PP2	D04	X_36	X_36F	tctttcctaaggtccagtggga	X_36R	gctcctaactcatcaacacctg
PP2	E04	X_37	X_37F	accaagacgtgaggcaatggag	X_37R	agagaggcattaccttagcctc
PP2	F04	X_38	X_38F	gagtcactgttaaggtcgccct	X_38R	tgggagtgttcaggatcaggtg
PP2	G04	X_39	X_39F	actggaggtcttcacagccttg	X_39R	cagtctctgagttctagggcag
PP2	H04	X_42	X_42F	tgggagttgtagttccagggct	X_42R	atcacagtgcaggcttcatgga
PP2	A05	X_43	X_43F	acgtctctgctcatcctgggag	X_43R	tctgagccatctcttcagccag
PP2	B05	X_44	X_44F	aaccaaaccaggatgtcccagc	X_44R	taagccttccgaactcaactgc
PP2	C05	X_45	X_45F	aaggacttggatattgctccac	X_45R	cctgaactcacagagatccgtc
PP2	D05	X_46	X_46F	tcccacggcaaaggcaccaagc	X_46R	agctctttgagatacagccagg
PP2	E05	X_47	X_47F	acactttgtccatcagcattgg	X_47R	tgtcattccagtttgggtgagg
PP2	F05	X_50	X_50F	ccttccaggattccactgagag	X_50R	ttccagaagacccaaagccctc
PP2	G05	X_51	X_51F	tagctgcccacagtggatgtcc	X_51R	aaaccactggccgggaatgctc
PP2	H05	X_52	X_52F	tgatgccagatggaacccttgg	X_52R	tgcggttcacagaccagtcagc
PP2	A06	X_55	X_55F	acactcccaatcctccaagtgc	X_55R	tgcttgcaattcctgaaccagg
PP2	B06	X_56	X_56F	agcagggaaaggtgcagacctc	X_56R	agaactgggtaacaggagccac
PP2	C06	X_57	X_57F	aggtttcctctgacatctgcac	X_57R	gttgcctacaagggcacatgac
PP2	D06	X_58	X_58F	tgcctctgactcccaagagctg	X_58R	agcttggtctggagggagttcc
PP2	E06	X_59	X_59F	atgtgttccctggtcagagctg	X_59R	gaggtgagtggacagatggctg
PP2	F06	X_60	X_60F	atgtacactttggtggtcatgg	X_60R	agctctttggttccccccagag
PP2	G06	X_61	X_61F	aatgcttgcccagcttgtgtgg	X_61R	tgaagttgctttggtcagggtg
PP2	H06	X_62	X_62F	agcactgggagtatgcttgctg	X_62R	tgctgtgtgtggtgatgcatgg
PP2	A07	X_63	X_63F	gcacactcatcaaaccacagac	X_63R	agagaaactggtgtcatcctgc
PP2	B07	X_66	X_66F	cgggtcgaggctgctcacagtg	X_66R	cactctgtcttgagcctagctc
PP2	C07	X_67	X_67F	acttgtgcctgaggacttgagc	X_67R	cttgggaactttgcaggtgctc
PP2	D07	X_69	X_69F	atcctggactgtcaagccaagc	X_69R	tgccagggccaggaagcaggag
PP2	E07	X_71	X_71F	ttagtgggttgaatggacttgc	X_71R	agagtcctacaagtcctcaagg

PP2	F07	X_72	X_72F	tgtatgtgccaggtgtcctcag	X_72R	tctgcaagtccattcaacccac
PP2	G07	X_75	X_75F	tctctgactgctacacacacac	X_75R	aggtaaaggcatttgcagccag
PP2	H07	X_76	X_76F	taagggacttaggggcaaactc	X_76R	tagggattgaacccaaggttcc
PP2	A08	X_78	X_78F	agattccactctacggactagac	X_78R	ctgagtgaagtatccgaggctg
PP2	B08	X_79	X_79F	aaggctgagcatgagagtccac	X_79R	agaagatggcctccactactgc
PP2	C08	X_80	X_80F	ggtcttgtgataactgcctgtg	X_80R	tttccagcaaacacacacacac
PP2	D08	X_82	X_82F	ttcagggttcaaggtcagtctg	X_82R	acacagtgagttccatgccagc
PP2	E08	X_83	X_83F	attgtggctcaaaacactggtc	X_83R	ccatcagtggaagtctgagtgg
PP2	F08	X_84	X_84F	ttaggttgccagcacccatagc	X_84R	gatttgcctctaagtgcgttgc
PP2	G08	X_85	X_85F	agctgtgttcccttggaccagc	X_85R	tgatggttgcccatcccaggag
PP2	H08	X_86	X_86F	agcgttagccattgttggacac	X_86R	agcttgattatatggagcccag
PP2	A09	X_87	X_87F	agttgcccagctattgtgctgc	X_87R	tccctccctgaagtcaccaagg
PP2	B09	X_88	X_88F	aactcaggtgaagcacatttgc	X_88R	aatcccagtccgtggaaactgg
PP2	C09	X_90	X_90F	gcatctgcacaggtacacgcac	X_90R	agtgattgcgtatctctgccag
PP2	D09	X_91	X_91F	tgctccctgattggtctcagtg	X_91R	actgcacacgtcttggtccagg
PP2	E09	X_92	X_92F	caacatgcacccatgcacacac	X_92R	ttacgcttgaatgtgtacgagc
PP2	F09	X_93	X_93F	aggtctgatccttagggctcac	X_93R	atatagcaagttctgaggcagc
PP2	G09	X_94	X_94F	tgagaattgaacccagggcttc	X_94R	ggaagctacatactgacaagcac
PP2	H09	X_96	X_96F	agcaagttcagtggagctttgg	X_96R	tgctcagatcagtgtcagccac
PP2	A10	X_97	X_97F	gtgcttcaacatgccaaattgc	X_97R	agacttggagcttccagggcca
PP2	B10	X_98	X_98F	tcctctcggaaggatgacccag	X_98R	aggattgctctagggtagtcac
PP2	C10	X_99	X_99F	tgggagttgtagttccagggct	X_99R	tgtcatttccgagtgtcactgc
PP2	D10	X_100	X_100F	acgtctctgctcatcctgggag	X_100R	tctgagccatctcttcagccag
PP2	E10	X_101	X_101F	gctttgttgcaatatgggcctg	X_101R	cccagaacttgagggtgtactc
PP2	F10	X_102	X_102F	aaaccactggccgggaatgctc	X_102R	tagctgcccacagtggatgtcc
PP2	G10	X_103	X_103F	aggtctgtctgcccggtaggtc	X_103R	cccaaggcctccatgtcagcag
PP2	H10	X_104	X_104F	tcccacggcaaaggcaccaagc	X_104R	agctctttgagatacagccagg
PP2	A11	X_105	X_105F	acactttgtccatcagcattgg	X_105R	tgtcattccagtttgggtgagg
PP2	B11	X_106	X_106F	tcctcagcagggccagtctgag	X_106R	gtgggaacgcgcacacacacac
PP2	C11	X_107	X_107F	cacctacagagaccacacacac	X_107R	tgagtgcccacccagcccagca
PP2	D11	X_109	X_109F	gcctctgcctcccaagtgctgg	X_109R	tccagagtcagtgagagaccct
PP2	E11	X_112	X_112F	tgaaagcggattacccatagca	X_112R	tcccttgctctacacacacac
PP2	F11	X_113	X_113F	tcggatgctcttatccgctgag	X_113R	gtctcacctggtggcaacttcc
PP2	G11	X_114	X_114F	cactagaacctgtgactgaagc	X_114R	cagggatcaagttacgtgtgtg
PP2	H11	X_115	X_115F	tgcagactctcggacctgtgtc	X_115R	tccaaagagcacacacacacac
PP2	A12	X_119	X_119F	cctatgtcccttccactcactc	X_119R	tcctgtctgaagtggtcagctc
PP2	B12	X_120	X_120F	agctctcaggcttgtgtgtgca	X_120R	tcggacaggataaactgcccac
PP2	C12	X_121	X_121F	ctgagactcactgaccaaccag	X_121R	cctcagatggctgcagtgacag
PP2	D12	X_123	X_123F	gagagttccccaagtagcctcc	X_123R	gggtctaggttacctcacctgg
PP2	E12	X_124	X_124F	agctgcaaatatctggctttgg	X_124R	tttcgctggcgcccacccaagc
PP2	F12	X_125	X_125F	aatcgccatagaaacacacctc	X_125R	cctccaagtgtacccacacagg
PP2	G12	X_126	X_126F	tcccatgcgtaatcacctctcc	X_126R	tcaggcctcaaagcgtgcacag
PP2	H12	X_127	X_127F	agggtgtctctgtgaccctgac	X_127R	tcctgtcctttatctggaagca
PP3	A1	Chr10_1	Chr10_1F	accaagtctgatgacctcagtg	Chr10_1R	ggacctgaactatgtagccctg
PP3	B1	Chr10_2	Chr10_2F	aggacctctatgcgtgagctgc	Chr10_2R	tggagctgggcctgcacaccag
PP3	C1	Chr10_3	Chr10_3F	acagtgacatcacctggagacc	Chr10_3R	acacacatgcagacacacgcac
PP3	D1	Chr10_4	Chr10_4F	acagtccaaagcagccctacac	Chr10_4R	tcctctggaactggagttctgg
PP3	E1	Chr10_5	Chr10_5F	gaactgggcttatgacaggcac	Chr10_5R	gggcagtgagctgcttgtgagc
PP3	F1	Chr10_6	Chr10_6F	agggctttagcttgtgccaagc	Chr10_6R	tgactccgggtcttcctggtgg
PP3	G1	Chr10_7	Chr10_7F	acaagtgctcttaccactgagc	Chr10_7R	tccagggcagccagggctacac
PP3	H1	Chr10_8	Chr10_8F	ttggtctccatcagcattctgg	Chr10_8R	actgtcaacaacagcttcggtg

PP3	A2	Chr10_9	Chr10_9F	gccgatagccttgctctggctg	Chr10_9R	ccacttggtaaggtgtccatgc
PP3	B2	Chr10_10	Chr10_10F	acactggatgtggatctgggtg	Chr10_10R	accctgaggtccgtgcctgctg
PP3	C2	Chr10_11	Chr10_11F	tggcatacagtcacacatggac	Chr10_11R	ggtttgaagggaggctgaacac
PP3	D2	Chr10_12	Chr10_12F	aggtgtggttcagcgttaacac	Chr10_12R	tgtgatgcttgcctctgagtgc
PP3	E2	Chr10_13	Chr10_13F	ccctaggctctattgcagtacg	Chr10_13R	agctagttggcttcagggtacc
PP3	F2	Chr10_14	Chr10_14F	cctgacacacatatgcaggcac	Chr10_14R	gccttcctctgtatgtgtgcag
PP3	G2	Chr10_15	Chr10_15F	caagctgatgctcatcctcagg	Chr10_15R	catcggaggcttaccaacatcc
PP3	H2	Chr10_16	Chr10_16F	accctatggaggatcagtgctg	Chr10_16R	agatgacaaactccaggttcac
PP3	A3	Chr10_17	Chr10_17F	gggtttcctaagatcacgggtg	Chr10_17R	caacattggaaagtcccaagtc
PP3	B3	Chr10_18	Chr10_18F	agttcccagtactcggtaacca	Chr10_18R	cctggtctacatggtgagactc
PP3	C3	Chr10_19	Chr10_19F	tgagcaggttcttccagatgca	Chr10_19R	agatctcgccatgcttcagagg
PP3	D3	Chr10_20	Chr10_20F	cccacaggacagacggtgtctg	Chr10_20R	agatttctgagttcgaggccag
PP3	E3	Chr10_21	Chr10_21F	tcctagggacaagggtatgtcc	Chr10_21R	agcacttgtgtctgtttgcagg
PP3	F3	Chr10_22	Chr10_22F	tttagatgctaccgggtgccag	Chr10_22R	aaaccctgactgagacacacac
PP3	G3	Chr10_23	Chr10_23F	cttaggggtctgggaatggctc	Chr10_23R	tcctcctgtttgtgggtgctgg
PP3	H3	Chr10_24	Chr10_24F	gtgaagtgggccgtcctcatgg	Chr10_24R	tccagtgcagacccaagaccag
PP3	A4	Chr10_25	Chr10_25F	gtgctgagattataggcatgtacc	Chr10_25R	tcaccctctgtcccagagtagc
PP3	B4	Chr10_26	Chr10_26F	tgggatgtctgcctgctgggag	Chr10_26R	tcctctgaaagagcagccagtg
PP3	C4	Chr10_27	Chr10_27F	agcctgaccatcacttgggctg	Chr10_27R	cctggcccagcaaacctcaagc
PP3	D4	Chr10_28	Chr10_28F	tcctagctcttggcggtggagg	Chr10_28R	cagggtgcgtgaatggctttcc
PP3	E4	Chr10_29	Chr10_29F	tgaacaccagtgatcccagtgc	Chr10_29R	cccagggcttcaggcttgtgac
PP3	F4	Chr10_30	Chr10_30F	ctcacaagcctgaaagcctcag	Chr10_30R	aagatgtttcccacaatgctgc
PP3	G4	Chr10_31	Chr10_31F	tcagtcctcagctccacaaagg	Chr10_31R	tgggtcttccattcatgagcac
PP3	H4	Chr10_32	Chr10_32F	agggttcctgtgctctgggcag	Chr10_32R	gcccaagcttctgaccttcagg
PP3	A5	Chr10_33	Chr10_33F	caccacggcctggctctgtgag	Chr10_33R	cacgcctttaatcccagcagag
PP3	B5	Chr10_34	Chr10_34F	agccacttgtaactccaggacc	Chr10_34R	tgctctgtagaccaggatgtcc
PP3	C5	Chr10_35	Chr10_35F	acctgctgaaccatcttgctgg	Chr10_35R	agcattctgtcaacccagagac
PP3	D5	Chr10_36	Chr10_36F	tgagacagggaagcagacacac	Chr10_36R	tgctttgtctagtctgtggacg
PP3	E5	Chr10_37	Chr10_37F	cctcagaaccatccagtcaacc	Chr10_37R	tggcctgtccaaccacaggtac
PP3	F5	Chr10_38	Chr10_38F	tccagctacagaccacacacac	Chr10_38R	ccagatagctcagccgcaagtg
PP3	G5	Chr10_39	Chr10_39F	acgcggacaccgtgagcctctg	Chr10_39R	tgttgttgggaagggctcacac
PP3	H5	Chr10_40	Chr10_40F	gcacctctaactagtcgaagaacc	Chr10_40R	ccaacagttcagatatggcattcc
PP3	A6	Chr10_41	Chr10_41F	aggatcccaagtttgaggccag	Chr10_41R	agcttactccagggatgccctc
PP3	B6	Chr10_42	Chr10_42F	tgcaggttgcaggttgggtcag	Chr10_42R	cagagtgctccagcatgcctgg
PP3	C6	Chr10_43	Chr10_43F	ttggcgcttcccgacctgttgg	Chr10_43R	tgggtgcagacctgccaagatc
PP3	D6	Chr10_44	Chr10_44F	attacagatggttgtgagccac	Chr10_44R	gagtaccacagaacccacttgg
PP3	E6	Chr10_45	Chr10_45F	gggacacatggtatgtctgcac	Chr10_45R	tacagtgcttctccgactaagg
PP3	F6	Chr10_46	Chr10_46F	acctaagcagtgtcccagcctg	Chr10_46R	tgccatgaaagtctggtgactg
PP3	G6	Chr10_47	Chr10_47F	tggagggagtttttgtcagtgc	Chr10_47R	gcttctcaacatggacatcagc
PP3	H6	Chr10_48	Chr10_48F	agtctccagggcacacatggtg	Chr10_48R	agcatcaacttccctggcaacg
PP3	A7	Chr10_49	Chr10_49F	tgaaggtgtcttcaagtggtgc	Chr10_49R	ccacagggtggaccacgaactc
PP3	B7	Chr10_50	Chr10_50F	cataggccaggtatgacgctgc	Chr10_50R	tgatagctggtctccctggtgg
PP3	C7	Chr10_51	Chr10_51F	agtccgctacctccgaagctcg	Chr10_51R	cttaccaggccggaggccatgc
PP3	D7	Chr10_52	Chr10_52F	tcttgggttggtacgatggctc	Chr10_52R	acgctgtactctgctgtcagac
PP3	E7	Chr10_53	Chr10_53F	agcacacaggatgtggcaactg	Chr10_53R	catgactgcttaccatcaagca
PP3	F7	Chr10_54	Chr10_54F	tccaacagtcagacacagacac	Chr10_54R	tgatgcccacatacacagcttg
PP3	G7	Chr10_55	Chr10_55F	tgcaatggccttcggagttctg	Chr10_55R	tcacctgtgactccagttccag
PP3	H7	Chr10_56	Chr10_56F	agttaggcaagacctctgccac	Chr10_56R	acagagtgagttccaggacagc
PP3	A8	Chr10_57	Chr10_57F	tttggagggtagagccactgag	Chr10_57R	tctctgactgctttgcctgaag
PP3	B8	Chr10_58	Chr10_58F	tgtccttggagttggcttcagg	Chr10_58R	tgtgcatccctgggcatcctgg
PP3	C8	Chr10_59	Chr10_59F	ttcagcgcaaatgcttgcttgg	Chr10_59R	cacaagctcccaggctctacac

PP3	D8	Chr10_60	Chr10_60F	tcctggtggcctgtttcccagg	Chr10_60R	aaagccaatgctgtctgcaagg
PP3	E8	Chr10_61	Chr10_61F	cccaggttaggtggttgctcag	Chr10_61R	agatggctcccaaacaccactc
PP3	F8	Chr10_62	Chr10_62F	gcgaatctctctagggtctcac	Chr10_62R	catgcccaccctgtccagatgg
PP3	G8	Chr10_63	Chr10_63F	acttccctgagcaaggtccagc	Chr10_63R	agcttcccgctatagtgcgctg
PP3	H8	Chr10_64	Chr10_64F	cccagacgcaagtaagccgtgc	Chr10_64R	accatctcccaagaatccggtg
PP3	A9	Chr10_65	Chr10_65F	taggccctatggcccaccggag	Chr10_65R	gggtatgtgtggagtgtgtgtg
PP3	B9	Chr10_66	Chr10_66F	tgggcacaccctgcagtgttgc	Chr10_66R	ggctacacagaggaaccctgtc
PP3	C9	Chr10_67	Chr10_67F	tggaaatcagtctggcggttcc	Chr10_67R	acgtcagcgcacacacacctgc
PP3	D9	Chr10_68	Chr10_68F	tgagagggcattaccttcccag	Chr10_68R	attggaactcaaggttcacctg
PP3	E9	Chr10_69	Chr10_69F	aaggctctactatctcccaagg	Chr10_69R	tcacatggtggctcacaaccac
PP3	F9	Chr10_70	Chr10_70F	acaacagcattacctgctagac	Chr10_70R	gtgttagcgtgtgtggcattgg
PP3	G9	Chr10_71	Chr10_71F	acagtccatgcggccaagccag	Chr10_71R	agctctgagttccaggccagct
PP3	H9	Chr10_72	Chr10_72F	tcatcttaggcacacacggtgc	Chr10_72R	catgcctgcacagcaagggctc
PP3	A10	Chr10_73	Chr10_73F	agctgcggtgtgatggctgtgg	Chr10_73R	acagatggctgctcacccagtg
PP3	B10	Chr10_74	Chr10_74F	tgacaatgcttgtgtgtgtgtg	Chr10_74R	aaagtcgctgggcctgggtagg
PP3	C10	Chr11_1	Chr11_1F	tggagctgccctgtgagtgctg	Chr11_1R	acctccacgcacccacgcactg
PP3	D10	Chr11_2	Chr11_2F	tgaggcagcccgaagtggcctc	Chr11_2R	tgcctgctgtctcgggtctctg
PP3	E10	Chr11_3	Chr11_3F	actgcagcttgtgacaggtctg	Chr11_3R	tgtgtcgattcagtgacacgtc
PP3	F10	Chr11_4	Chr11_4F	agcatcctgcttgctcagagtc	Chr11_4R	tggaggtggagcccaggacctc
PP3	G10	Chr11_5	Chr11_5F	tctgcccactgaggctgtttgc	Chr11_5R	tgctggccctttgtgatccctg
PP3	H10	Chr11_6	Chr11_6F	cccggattccaggcagtgctgg	Chr11_6R	ccagtgtgcgatccacgtagtg
PP3	A11	Chr11_7	Chr11_7F	tttatgggcagttaacggttgc	Chr11_7R	tttcctttaggaggcgggtctc
PP3	B11	Chr11_8	Chr11_8F	tctcctcgtgggaagcgtgtcg	Chr11_8R	tgggctacatgtacttgccagg
PP3	C11	Chr11_9	Chr11_9F	gctagttgtgagccaccatgtg	Chr11_9R	acactcacaaagacccatgtgc
PP3	D11	Chr11_10	Chr11_10F	ttgggaaacacaacacccttgg	Chr11_10R	gtgagccatcatgtggttgctg
PP3	E11	Chr11_11	Chr11_11F	cccgcccaccactgacacacac	Chr11_11R	ttgttggcggaaggagctcgtc
PP3	F11	Chr11_12	Chr11_12F	aggcttggcttctctgggatcg	Chr11_12R	gagatggctcctacctctggag
PP3	G11	Chr11_13	Chr11_13F	cctgtggagcgagtaggctcag	Chr11_13R	agctcccagccaatcgttgagc
PP3	H11	Chr11_14	Chr11_14F	gctgagactccagagcgttcac	Chr11_14R	cggtacatcaggcaagccatgc
PP3	A12	Chr11_15	Chr11_15F	ctagcctcagtctgtgcaggac	Chr11_15R	tgcaccaggcttcactagctgg
PP3	B12	Chr11_16	Chr11_16F	agttcagggaaggactgtgctc	Chr11_16R	ctgccacagtgctgagttgtgg
PP3	C12	Chr11_17	Chr11_17F	tgtgtagccctggctgtcctgg	Chr11_17R	aacacccgatggtggctcacag
PP3	D12	Chr11_18	Chr11_18F	tgcatgactggtgcctgtggag	Chr11_18R	tggatgcagggactgacacacc
PP3	E12	Chr11_19	Chr11_19F	tcatgtgaggtaggcactgtac	Chr11_19R	agagctgacatgatgcctcagc
PP3	F12	Chr11_20	Chr11_20F	agtctgaggttagcctggtctg	Chr11_20R	tccaggacagtcagggctacac
PP3	G12	Chr11_21	Chr11_21F	acctttggaagtgcagtcagtg	Chr11_21R	attccaaggcctttgcacagac
PP3	H12	Chr11_22	Chr11_22F	aacacatgctccacaactggtg	Chr11_22R	tgagttccagcaaagccagacc
PP4	A1	Chr11_23	Chr11_23F	ggctatgcagtcatggtggacc	Chr11_23R	gctaaaggctaaggcatctctg
PP4	B1	Chr11_24	Chr11_24F	agcaagtgccatgccttctgag	Chr11_24R	gcacctgtttgacctgacaacc
PP4	C1	Chr11_25	Chr11_25F	ctcgaactcagaaatccgcctg	Chr11_25R	gggatgtggctcagtcgacctg
PP4	D1	Chr11_26	Chr11_26F	agcttcaagaactgctgtcctg	Chr11_26R	gtggagcctcccattccagcag
PP4	E1	Chr11_27	Chr11_27F	tgcccatagagaagctgaaacc	Chr11_27R	tctctccagggatcagagccac
PP4	F1	Chr11_28	Chr11_28F	ttcagtgaccctccagtgtagg	Chr11_28R	agcaagcctgtaagcagtgttc
PP4	G1	Chr11_29	Chr11_29F	tgctactcgcgccagccacctg	Chr11_29R	agactcgttggaagactcgctg
PP4	H1	Chr11_30	Chr11_30F	gaaagctccaggaagcacgcac	Chr11_30R	aatgcagacctctcctgagacc
PP4	A2	Chr11_31	Chr11_31F	tccagtcccagttgccagacac	Chr11_31R	aacctggtctcagaggctgtcg
PP4	B2	Chr11_32	Chr11_32F	tggaaaccagctctgtaggtgc	Chr11_32R	gagatgaccttgaccctgactg
PP4	C2	Chr11_33	Chr11_33F	agggcgaactctccagcactgc	Chr11_33R	gcaacatgcaagacatctgtgc
PP4	D2	Chr11_34	Chr11_34F	ttaaaggcgtgtgccaccatgc	Chr11_34R	agacctgtgactccagttccag
PP4	E2	Chr11_35	Chr11_35F	tgactgctcttccagaggttcc	Chr11_35R	aagttgccatgtgagtgcttgg
PP4	F2	Chr11_36	Chr11_36F	caacatacggtcttcgggaacc	Chr11_36R	cgctagggcacagagcaagtcc

PP4	G2	Chr11_37	Chr11_37F	agggctgtgtcctgacgctcac	Chr11_37R	gtgaagcaccaggtcaacccac
PP4	H2	Chr11_38	Chr11_38F	actgagtgagttccagaacagc	Chr11_38R	agctgaggatttagagttccca
PP4	A3	Chr11_39	Chr11_39F	aggcacttactgcacaagcctg	Chr11_39R	ggctgtggttctggaagtcacc
PP4	B3	Chr11_40	Chr11_40F	tctacagagcaagttccaggac	Chr11_40R	tctgtaacccactgctctcagg
PP4	C3	Chr11_41	Chr11_41F	acgcctttaatcccagcactcg	Chr11_41R	ggctttcacgattcaacacctg
PP4	D3	Chr11_42	Chr11_42F	agcattccaaacccacctacag	Chr11_42R	cagcctgacacagcatacccag
PP4	E3	Chr11_43	Chr11_43F	tgagtctgtctagcttggactc	Chr11_43R	tgacccaaagttgatgctccac
PP4	F3	Chr11_44	Chr11_44F	tcgacttctgatggcttaggac	Chr11_44R	acaggcactgtcatgctcactg
PP4	G3	Chr11_45	Chr11_45F	gcattggtggatggcctacctg	Chr11_45R	tgcatgggcgtgttgtgcgtgc
PP4	H3	Chr11_46	Chr11_46F	gatggctcaggtcaggctttgc	Chr11_46R	gctgaccggctgatccctggag
PP4	A4	Chr11_47	Chr11_47F	tgtgcagttgtgcttgctggtg	Chr11_47R	tgagctccaggcttcttggtgg
PP4	B4	Chr11_48	Chr11_48F	acctggcagtgttggcacatgc	Chr11_48R	agctgaggatttagagttccca
PP4	C4	Chr11_49	Chr11_49F	gggtttgaacactcagatgcag	Chr11_49R	caagttgtcctctgacctgcac
PP4	D4	Chr11_50	Chr11_50F	tgattcccagcacctacactgg	Chr11_50R	tgttgctcatactgaaacccag
PP4	E4	Chr11_51	Chr11_51F	ggactggagacttgcagattgc	Chr11_51R	agggctacagagaaaccctgtc
PP4	F4	Chr11_52	Chr11_52F	acataaatcagcacagccctgg	Chr11_52R	aggctctgacagccatcaggtg
PP4	G4	Chr11_53	Chr11_53F	cctggcagagcctgtgtggacc	Chr11_53R	ggcaaaccgataaccctactgg
PP4	H4	Chr11_54	Chr11_54F	tctaggctccataacaaggctc	Chr11_54R	tgactgtaccacacacacgctg
PP4	A5	Chr11_55	Chr11_55F	tctaactccgaggtggtggctc	Chr11_55R	ggagactgccctcccagatagc
PP4	B5	Chr11_56	Chr11_56F	gaagctgactttgccatagggt	Chr11_56R	acatgatggggtaacagccctc
PP4	C5	Chr11_57	Chr11_57F	tgaaactcagtggagactcacc	Chr11_57R	accctgtatggtgtctgtggac
PP4	D5	Chr11_58	Chr11_58F	gttggtgccatctagtgtgggt	Chr11_58R	tgagaccttgaatgcaatgacc
PP4	E5	Chr11_59	Chr11_59F	tggcaatcagctttcaagcctg	Chr11_59R	actagcccaagtgcacacacac
PP4	F5	Chr11_60	Chr11_60F	tgagcaggcactttgcctcagc	Chr11_60R	gatggtggcctctctggtcagc
PP4	G5	Chr11_61	Chr11_61F	agaaaccagggaagtgacacac	Chr11_61R	ttaggtggcttctcagacgtgg
PP4	H5	Chr11_62	Chr11_62F	cgaaggctttatcaagctctgtg	Chr11_62R	ggctcagtggttcacagactgc
PP4	A6	Chr11_63	Chr11_63F	gacctgggttctgcctcgagtg	Chr11_63R	tcatgtgacaactggtagcctg
PP4	B6	Chr11_64	Chr11_64F	ttcagcagccatttggcaccag	Chr11_64R	agctacatagtgaggcactgtc
PP4	C6	Chr11_65	Chr11_65F	aggccaacctgctccagcaagg	Chr11_65R	ccagtaagctaaggctcccagg
PP4	D6	Chr11_66	Chr11_66F	tgctgacacgatgacaacatgc	Chr11_66R	ccagttccaagggagccagcac
PP4	E6	Chr11_67	Chr11_67F	gtaggctgtgacaccagcactg	Chr11_67R	aaagacaggctctctgacccag
PP4	F6	Chr11_68	Chr11_68F	agtcttctgcacacacacgcac	Chr11_68R	cctgtgtgccgagcaccgtgtc
PP4	G6	Chr11_69	Chr11_69F	acaggctttaacaaccaccacc	Chr11_69R	taggctccaagtccagtgagac
PP4	H6	Chr11_70	Chr11_70F	tgggaaccggcttgctgcacac	Chr11_70R	ttgatagccactggcatcacac
PP4	A7	Chr11_71	Chr11_71F	gctaaaggctaaggcatctctg	Chr11_71R	ttcccacagaggaagctccagg
PP4	B7	Chr11_72	Chr11_72F	agcaagtggaggccaactggag	Chr11_72R	tgggcttaaagttatgcgccac
PP4	C7	Chr11_73	Chr11_73F	agaagctgtctctgcagcacac	Chr11_73R	agctgttcttctgtgactggtc
PP4	D7	Chr11_74	Chr11_74F	acacacctaaaggcatctgcac	Chr11_74R	agttccagcggctctgatgcca
PP4	E7	Chr11_75	Chr11_75F	ttacggatggttgtgagccacc	Chr11_75R	tgggaatagcccggccattgag
PP4	F7	Chr11_76	Chr11_76F	attgtggtgatctcacctcagg	Chr11_76R	taggtagccgaaggcacttacc
PP4	G7	Chr11_77	Chr11_77F	acctggcaagcacgaagacctg	Chr11_77R	ctcgaactcagaaatccgcctg
PP4	H7	Chr11_78	Chr11_78F	gattgaacctcatgtagcccag	Chr11_78R	catgagcctaaacggtgctcag
PP4	A8	Chr11_79	Chr11_79F	tggaaaccaagcacccaggtcc	Chr11_79R	ccacatgcgattcacacggtgc
PP4	B8	Chr11_80	Chr11_80F	aggtctgtccagcagcctttgg	Chr11_80R	tgtgtcgattcagtgacacgtc
PP4	C8	Chr11_81	Chr11_81F	tcctctggctccaaggctccag	Chr11_81R	aggaagaggtagggagttcctg
PP4	D8	Chr11_82	Chr11_82F	tgagtttggttcccagcaccca	Chr11_82R	gtcctctggaagagtagccctg
PP4	E8	Chr11_83	Chr11_83F	gccctggtgctgtggctttgag	Chr11_83R	tctgtggggtgagttccaggac
PP4	F8	Chr13_01	Chr13_01F	agccctccatgaacaatggctg	Chr13_01R	agatgccgcacctgcatactgg
PP4	G8	Chr13_02	Chr13_02F	agggtctcattaagttgcccag	Chr13_02R	tctctctgctgcacacagaacc
PP4	H8	Chr13_03	Chr13_03F	agtttggagccacctgctatgg	Chr13_03R	gatctccagtttcccagtcgtg
PP4	A9	Chr13_04	Chr13_04F	gatctcaccttgctcccaggac	Chr13_04R	tgtccaggagacacagttgtgc

PP4	B9	Chr13_05	Chr13_05F	tgatgaggccagtcctgcagac	Chr13_05R	tgtcgtcgtagatgcttgtgag
PP4	C9	Chr13_06	Chr13_06F	acatctgcatgcacactgtgac	Chr13_06R	ccaacagcaagatccacagtgc
PP4	D9	Chr13_07	Chr13_07F	ctccatgttggtaccttgcagc	Chr13_07R	acacaccagagacagtgtgctc
PP4	E9	Chr13_08	Chr13_08F	tggcaaacatgacatctggagc	Chr13_08R	aggcatcttccacaaagcctgg
PP4	F9	Chr13_09	Chr13_09F	cgatcagaggtgcaaagccagc	Chr13_09R	cgggctgatgttggtcaaggtc
PP4	G9	Chr13_10	Chr13_10F	acatgcctcgtgcctatgcagg	Chr13_10R	ctggttgagaggctaggcccac
PP4	H9	Chr13_11	Chr13_11F	tgactaggcttggcggttcctg	Chr13_11R	tggcatcctgaagctcagtcac
PP4	A10	Chr13_12	Chr13_12F	tggttgctagtgagggatctcg	Chr13_12R	actatgccagggtttctgggtg
PP4	B10	Chr13_13	Chr13_13F	gggttggctgtccacctgtagc	Chr13_13R	ggtttaacgtgaggagtcttcc
PP4	C10	Chr13_14	Chr13_14F	gccacgttagaaaacgcaccac	Chr13_14R	catgcagcgtgatgatgggctg
PP4	D10	Chr13_15	Chr13_15F	actgaggcccaggcatgactcc	Chr13_15R	tccaaccatgttagctttggag
PP4	E10	Chr13_16	Chr13_16F	tgtgtagccctggctgtcctgg	Chr13_16R	tgcagtgtaacttcacaggagc
PP4	F10	Chr13_17	Chr13_17F	ttctgtggtgttcgttgtgcag	Chr13_17R	aggccctaaatgatcaggctcc
PP4	G10	Chr13_18	Chr13_18F	cccagccatgggtcgtttgtgg	Chr13_18R	tcaagtctccaccaagccacag
PP4	H10	Chr13_19	Chr13_19F	acccagagtgccagcatcactg	Chr13_19R	acaagagcccatgactcctgtg
PP4	A11	Chr13_20	Chr13_20F	tgagttgctcctcaaggctctc	Chr13_20R	acaggacatttgtgttagccac
PP4	B11	Chr13_21	Chr13_21F	caacaatgccctcggtctgtgg	Chr13_21R	tcagctagcctggcgtacccag
PP4	C11	Chr13_22	Chr13_22F	tagctgatgccaagaccagtcc	Chr13_22R	agtctctccagacagcactacc
PP4	D11	Chr13_23	Chr13_23F	agctgtcacgattgctacatcg	Chr13_23R	acatggaaacttgggctttgac
PP4	E11	Chr13_24	Chr13_24F	agcccttgctcaaggacaaggt	Chr13_24R	tgattgagcaggtgtgggaagg
PP4	F11	Chr13_25	Chr13_25F	tgtgggcatccacactcaccag	Chr13_25R	tcccatgccacgtttggcacag
PP4	G11	Chr13_26	Chr13_26F	cccacagacctgttgccaagac	Chr13_26R	tccaggttccagaggacgcctg
PP4	H11	Chr13_27	Chr13_27F	tgtgtgccgaatcgccatgcag	Chr13_27R	aggcgtgcaccaccactgctgg
PP4	A12	Chr13_28	Chr13_28F	aacagcacttctggcctgggac	Chr13_28R	gcgtgagaactcctggagctgg
PP4	B12	Chr13_29	Chr13_29F	ccaatgtagccttcctggtgag	Chr13_29R	cataatgggcatcactgggctc
PP4	C12	Chr13_30	Chr13_30F	tggaactggctgcgataccagc	Chr13_30R	tgagtggctaatggacactctg
PP4	D12	Chr13_31	Chr13_31F	tcagagttgctttgcaggatgg	Chr13_31R	agaggctgccagtgcttacgtg
PP4	E12	Chr13_32	Chr13_32F	ctccagaagctcacaagccagc	Chr13_32R	agtccttagctgtgctggaaac
PP4	F12	Chr13_33	Chr13_33F	gaccatccagtaccaagccctg	Chr13_33R	tagcaatgcccaccacacaagc
PP4	G12	Chr13_34	Chr13_34F	gcagttatggaccacatgcacc	Chr13_34R	actggagtgcctttgcatgtgc
PP4	H12	Chr13_35	Chr13_35F	cagtgcaggcctagtagcctgg	Chr13_35R	tcatgagccatcatgtggatgc
PP5	A1	Chr13_36	Chr13_36F	tctgtggcatgtggaccttgag	Chr13_36R	cttcagcatggccaggcaaacc
PP5	B1	Chr13_37	Chr13_37F	aattcggtgacagtgctctcag	Chr13_37R	tgaagcagctcacaactgcctg
PP5	C1	Chr13_38	Chr13_38F	tcttggctcaagtcacaggacc	Chr13_38R	ttcagcaacccactgtgtcagg
PP5	D1	Chr13_39	Chr13_39F	tgccatctgaagagcttggttg	Chr13_39R	agcattcacatctccagggtcc
PP5	E1	Chr13_40	Chr13_40F	gaacctgcctgagagatgcctc	Chr13_40R	tgaattggcttcagcaagaacc
PP5	F1	Chr13_41	Chr13_41F	tgcagtcagtccctgaggcacc	Chr13_41R	tggcaacagccaggtgtgcctg
PP5	G1	Chr13_42	Chr13_42F	caatggcctcacactggtcagg	Chr13_42R	aacactgggcaaaggtgcttcc
PP5	H1	Chr13_43	Chr13_43F	tgcatagcaactcccaggccac	Chr13_43R	tgtgagttgtctgacttgggtg
PP5	A2	Chr13_44	Chr13_44F	agtctgttcctggcatggacac	Chr13_44R	caaggactgagacacaccctgg
PP5	B2	Chr13_45	Chr13_45F	tctttctggctctgcactgctg	Chr13_45R	cctttgaccaaacacactggga
PP5	C2	Chr13_46	Chr13_46F	agttggttggaaagatgcttgc	Chr13_46R	agtgatggagaaacctttgctc
PP5	D2	Chr07_1	Chr07_1F	tgggaaacatggcagcttgcag	Chr07_1R	gtaggcaggctttgaggactcc
PP5	E2	Chr07_2	Chr07_2F	actaaggcacctgctctgtagc	Chr07_2R	tcctgggtgactgcactccagg
PP5	F2	Chr07_3	Chr07_3F	tccataggcactgctcaacacc	Chr07_3R	atagccaaggaccaccagactc
PP5	G2	Chr07_4	Chr07_4F	tgaacttccagaggtccttgag	Chr07_4R	ttaaaggcgtgtgccaccatgc
PP5	H2	Chr07_5	Chr07_5F	acggagagcccttgcatagcac	Chr07_5R	tgctgtccaaatcgtcagggtg
PP5	A3	Chr07_6	Chr07_6F	tgagcaactcaagccgtagttg	Chr07_6R	agctcccagggatccaactgca
PP5	B3	Chr07_7	Chr07_7F	tgagtccacccacgcttacagg	Chr07_7R	agactccagatccggctcagtg
PP5	C3	Chr07_8	Chr07_8F	gctgctaggtcactggaactgg	Chr07_8R	acacagagcctggtgcctggtg
PP5	D3	Chr07_9	Chr07_9F	agatcactggttggcagttgac	Chr07_9R	cctgttgcacaaagcgctcagg

PP5	E3	Chr07_10	Chr07_10F	agtctggacacgggtaacctcc	Chr07_10R	acgccacaggtcagggcactcc
PP5	F3	Chr07_11	Chr07_11F	agctgggagcttacctctcagg	Chr07_11R	tcggaattgcaccttgcgggag
PP5	G3	Chr07_12	Chr07_12F	aggcatcctaacctgctgaacc	Chr07_12R	gccaagagttcagcattggctg
PP5	H3	Chr07_13	Chr07_13F	aggtgggtgtgagctacaggtg	Chr07_13R	tggcaagaaagggctcagcagg
PP5	A4	Chr07_14	Chr07_14F	aggacagccagggagggctctg	Chr07_14R	gccaccaggccaagcccaacac
PP5	B4	Chr07_15	Chr07_15F	ttgatgcccatacaccaccagg	Chr07_15R	gaagagcctgttccacgtatgc
PP5	C4	Chr07_16	Chr07_16F	acccatgacactctctgaggtg	Chr07_16R	tctatagtaggccaaggctgtc
PP5	D4	Chr07_17	Chr07_17F	tgggtaggcgttgcttgcacac	Chr07_17R	acacctgtggtggcagacactg
PP5	E4	Chr07_18	Chr07_18F	tggtatgctggtggtccaggtc	Chr07_18R	gcacgaggtccctctgaagcac
PP5	F4	Chr07_19	Chr07_19F	agttcgcccacaacgcagttcc	Chr07_19R	tccagccaagaagggtttgagc
PP5	G4	Chr07_20	Chr07_20F	accatcaagggtcactgggctc	Chr07_20R	ccttgtcccacactgtaggtgg
PP5	H4	Chr07_21	Chr07_21F	gcgcacgcatccctacaccgga	Chr07_21R	tataaacctcaccgctacggag
PP5	A5	Chr07_22	Chr07_22F	atgaaagcctttgccagaacac	Chr07_22R	actcgacaaatagccacaaacc
PP5	B5	Chr07_23	Chr07_23F	cgaactcaccgagatccgcctg	Chr07_23R	ttaccgcccatttgttggagtg
PP5	C5	Chr07_24	Chr07_24F	tcataactacagcaagtgcctg	Chr07_24R	cccacatgttaaaggcttggtc
PP5	D5	Chr07_25	Chr07_25F	tccggtaagtgtccgccttgag	Chr07_25R	gaggctggacgtgcattccgag
PP5	E5	Chr07_26	Chr07_26F	tgggaactgaactcaggacctc	Chr07_26R	tgatgctgccattggtagcctg
PP5	F5	Chr07_27	Chr07_27F	aaccagagtgtgttgtgaccac	Chr07_27R	ttctgcggatacctctccaagg
PP5	G5	Chr07_28	Chr07_28F	tgctcttacgacacacacacac	Chr07_28R	tgtcaggacccacatgcaatgg
PP5	H5	Chr07_29	Chr07_29F	gtgatccggtgttgggatttgg	Chr07_29R	ggacctcacatcagcctacagc
PP5	A6	Chr07_30	Chr07_30F	agtctagcctacgtgctcagtg	Chr07_30R	ttgacggtctcatgtagcccag
PP5	B6	Chr07_31	Chr07_31F	tgcctttgcctccaagtgctgg	Chr07_31R	aattcccagcaaccacatggtg
PP5	C6	Chr07_32	Chr07_32F	atcagcctgtgaactagcaagc	Chr07_32R	gacctagagcttccagactcag
PP5	D6	Chr07_33	Chr07_33F	tgagcctccacatgggtgcctg	Chr07_33R	tgggtgacaatacgcatatcca
PP5	E6	Chr07_34	Chr07_34F	acggagacttcctcaacgccca	Chr07_34R	acaatgacgctttcccatgcag
PP5	F6	Chr07_35	Chr07_35F	ctccagacatgagccttgacac	Chr07_35R	agcccaggttgcctgcacatcg
PP5	G6	Chr07_36	Chr07_36F	acagggtcctccagatcgctgc	Chr07_36R	tgaactgagctacagtcccgtg
PP5	H6	Chr07_37	Chr07_37F	aacagcagttaacgcacctgag	Chr07_37R	agattctgcatgagctttgtgg
PP5	A7	Chr07_38	Chr07_38F	atagagcaggacacccaatgtc	Chr07_38R	ggtgtgtgccaccactgtccag
PP5	B7	Chr07_39	Chr07_39F	acggtggtgatcaagatcctgc	Chr07_39R	acacactgcttccaccaccagg
PP5	C7	Chr07_40	Chr07_40F	gttggagtctgcttgttggagc	Chr07_40R	tgcgttgtcagtgctgatggtg
PP5	D7	Chr07_41	Chr07_41F	acacagggccgctcacagttgc	Chr07_41R	aggtgtgctcaaccattccagc
PP5	E7	Chr07_42	Chr07_42F	tggaaaccaaatgcaggttctc	Chr07_42R	tgttcccaaggctggcatttgc
PP5	F7	Chr07_43	Chr07_43F	ctgctacaggagatcccagagc	Chr07_43R	aggaccctcagcaaagcctgtg
PP5	G7	Chr07_44	Chr07_44F	ccagcactcaggagccagccag	Chr07_44R	agttgggttctcagccagccac
PP5	H7	Chr07_45	Chr07_45F	tgagcccaggtttctggagctg	Chr07_45R	tctatgcagggttcagggtctg
PP5	A8	Chr07_46	Chr07_46F	tcaggcccagcactgatggctg	Chr07_46R	tgggtctagcccaggtttcagc
PP5	B8	Chr07_47	Chr07_47F	aggcataggtttgatctccagc	Chr07_47R	gtttgccaccatgcctgaagac
PP5	C8	Chr07_48	Chr07_48F	tgaacatccatagtgcaagtgc	Chr07_48R	cccaagcacatgaactccaagc
PP5	D8	Chr07_49	Chr07_49F	gctaggtgctggatctgagcac	Chr07_49R	tccaccacataagtgctgggac
PP5	E8	Chr07_50	Chr07_50F	cacttctggtccatgtggttgc	Chr07_50R	cttccctgagtgctgggctcac
PP5	F8	Chr07_51	Chr07_51F	acagagtgagttccaggacagc	Chr07_51R	agtttgcacccagaacccaagc
PP5	G8	Chr07_52	Chr07_52F	ttgtgtgctcaagccaaagtgg	Chr07_52R	tagggtttgctggaggccagtc
PP5	H8	Chr07_53	Chr07_53F	tgaaacccatctggagtcgctg	Chr07_53R	actcacttggcagaccaagctg
PP5	A9	Chr07_54	Chr07_54F	gaagcttgacatggctccagag	Chr07_54R	tcaggaagtgctggtagctcac
PP5	B9	Chr07_55	Chr07_55F	tggactgtgtggcccatgcgag	Chr07_55R	tgccaacctcttcgctgttgag
PP5	C9	Chr07_56	Chr07_56F	caattcccacattcacacatgc	Chr07_56R	taacccagcagccaacctctgg
PP5	D9	Chr07_57	Chr07_57F	tgtgccatgaccactgcccagc	Chr07_57R	accatcatgcctagcatctgtc
PP5	E9	Chr07_58	Chr07_58F	caaccaacagaaccgcaatgtg	Chr07_58R	ataggacaccagagaccctctg
PP5	F9	Chr07_59	Chr07_59F	tctctgtgagttcgaggctagg	Chr07_59R	tgtttctggcggtcggtgtgtc
PP5	G9	Chr07_60	Chr07_60F	agctttctgtcaaggaacaagc	Chr07_60R	tctagccgaagcctctgcagca

PP5	H9	Chr07_61	Chr07_61F	accactagctttggcatcagtg	Chr07_61R	ttgcctaacaggctggatgacc
PP5	A10	Chr07_62	Chr07_62F	actcactagcagtcagacctgg	Chr07_62R	gtctctgcctcccaagtgctgg
PP5	B10	Chr07_63	Chr07_63F	agccactaggtctggcttgagc	Chr07_63R	actggagatgtggccttgttgg
PP5	C10	Chr07_64	Chr07_64F	tgatgtgacctggctttgctgg	Chr07_64R	ggccactgttggcactgcacag
PP5	D10	Chr07_65	Chr07_65F	tggcaggtgctaatggcttgag	Chr07_65R	aggatcacagaatggtgcttcc
PP5	E10	Chr07_66	Chr07_66F	ttgcacctggaagcccaacctg	Chr07_66R	gcacatgtacaaggacatgctc
PP5	F10	Chr07_67	Chr07_67F	agccacagccctgtcctcaagg	Chr07_67R	agcagtgtggtatgtcatgtgg
PP5	G10	Chr07_68	Chr07_68F	aggcattgaccactctgtcctg	Chr07_68R	tcacaggcatctgtaacgccag
PP5	H10	Chr07_69	Chr07_69F	gcacaccgtgctggtgctagac	Chr07_69R	tggaatgcgtgccaccgtgctc
PP5	A11	Chr19_01	Chr19_01F	ccattgacccagtgaggctgac	Chr19_01R	caggacctacccacctggtagc
PP5	B11	Chr19_02	Chr19_02F	ccaccaaacagaggctggactc	Chr19_02R	accacagcattgacaccatcac
PP5	C11	Chr19_03	Chr19_03F	tgacctccacacaacccaccac	Chr19_03R	cgaactcacaaagatccacctg
PP5	D11	Chr19_04	Chr19_04F	tgggaatgctgtatgttgcgac	Chr19_04R	tggctgtagtggtgtgtgcctg
PP5	E11	Chr19_05	Chr19_05F	tcacagctgcctgtaacgccag	Chr19_05R	agagatgtccgtacgattcctg
PP5	F11	Chr19_06	Chr19_06F	tccagtgcttccaaatgccagc	Chr19_06R	tccaggacagccagtgcttcag
PP5	G11	Chr19_07	Chr19_07F	gaagctgggacctctagggtcc	Chr19_07R	acctgctacatcctcagcccag
PP5	H11	Chr19_08	Chr19_08F	tcagaggcagagtctcgctagg	Chr19_08R	agttgtggctccgcccggtgtc
PP5	A12	Chr19_09	Chr19_09F	atctccagtggtatgtgcagac	Chr19_09R	tctttgctggttgtgttggcac
PP5	B12	Chr19_10	Chr19_10F	tctgcaagtgactgcctcagag	Chr19_10R	agacctgtgtccagctcccagg
PP5	C12	Chr19_11	Chr19_11F	agccggagcccagagcttctgc	Chr19_11R	tggctcgcacatttctggaagc
PP5	D12	Chr19_12	Chr19_12F	agtcactccatgtcagccagag	Chr19_12	aggaggcttctgttgcagatcc
PP5	E12	Chr19_13	Chr19_13F	agtcttgatctggtgcttggag	Chr19_13	tgatggctgatcttggttgcca
PP5	F12	Chr19_14	Chr19_14F	atatgggcaaatcccacaccag	Chr19_14	tctgcatccaaggcaacgtcag
PP5	G12	Chr19_15	Chr19_15F	agctgacctcagaaagaccacc	Chr19_15	actgagcactctactcctgagc
PP5	H12	Chr19_16	Chr19_16F	ctctagcacacatgaagccctg	Chr19_16	actgaagagcctaagtcgtgtg
PP6	A1	Chr01_1	Chr01_1F	tgggtgccctggagtctggagc	Chr01_1R	aggacgttgtagcgccgagtcc
PP6	B1	Chr01_2	Chr01_2F	aactggctctgggcattcctgg	Chr01_2R	cctgagaccaggctcatgcgag
PP6	C1	Chr01_3	Chr01_3F	ttagtgctggaaaccaaactcg	Chr01_3R	acacagagtcacacaccatacg
PP6	D1	Chr01_4	Chr01_4F	ggacacacttgtgctgtggtgc	Chr01_4R	ctgggtggcagtttacgaccac
PP6	E1	Chr01_5	Chr01_5F	agaccatgacactgggagcctc	Chr01_5R	ggaacatggttctgcaaggcca
PP6	F1	Chr01_6	Chr01_6F	actgacaagccaccgttcacag	Chr01_6R	tgcatcttgggaggcctgtgac
PP6	G1	Chr01_7	Chr01_7F	tatgctgcaaagaggcacctcc	Chr01_7R	tggtctagccacatcagtgagc
PP6	H1	Chr01_8	Chr01_8F	tgactcctctgtaccaactgtg	Chr01_8R	aagagttgccttggtcatggtg
PP6	A2	Chr01_9	Chr01_9F	agaactggactcttgtgccagg	Chr01_9R	aatgaccaggtttccatggctg
PP6	B2	Chr01_10	Chr01_10F	atgactgccatcctggctcagg	Chr01_10R	tgctgggatttgaactcgggac
PP6	C2	Chr01_11	Chr01_11F	catgaacacgtgatgccaacag	Chr01_11R	tggagctgtcatgatggcacac
PP6	D2	Chr01_12	Chr01_12F	tcccacagctgccatgaggacc	Chr01_12R	aggtgcctctttgcagcatacc
PP6	E2	Chr01_13	Chr01_13F	tgtggacatgtttgttcgcatg	Chr01_13R	tccaagtgagatttgatggcag
PP6	F2	Chr01_14	Chr01_14F	gaggtggttatgagctgcttgc	Chr01_14R	cttgtccctgtaccaagcactg
PP6	G2	Chr01_15	Chr01_15F	agttctgggatgagcctctgag	Chr01_15R	ggtcctctgcatacggcttgtg
PP6	H2	Chr01_16	Chr01_16F	tgggtgcctatcaccatagtgc	Chr01_16R	gtatttgtctggagtctgtgacc
PP6	A3	Chr01_17	Chr01_17F	agagctgtgacctcctctggag	Chr01_17R	tggcctgttggtcctaggcagc
PP6	B3	Chr01_18	Chr01_18F	tggtttccagcacccacgtcag	Chr01_18R	acagcatgtgccaccatgcctg
PP6	C3	Chr01_19	Chr01_19F	cttcaaatgccagcacagtcac	Chr01_19R	aggttccatgagagaccctgtc
PP6	D3	Chr01_20	Chr01_20F	tctgtggcaagggatcagcctg	Chr01_20R	atggagtctctcacctacggag
PP6	E3	Chr01_21	Chr01_21F	ggtctctgtctgctgccaagag	Chr01_21R	agaagcagttgggagcatcctc
PP6	F3	Chr01_22	Chr01_22F	actcttcaaatgccagcacagc	Chr01_22R	aggttccatgagagaccctgtc
PP6	G3	Chr01_23	Chr01_23F	gctgcaagaagctcacggctac	Chr01_23	agtgtcacgtcatccaggccag
PP6	H3	Chr01_24	Chr01_24F	actgctgagccatctctccagc	Chr01_24R	tcattaccgcagagtcaccgtg
PP6	A4	Chr01_25	Chr01_25F	tcctgaacagaggactctgacc	Chr01_25R	gctagactgatactttccgacttagc
PP6	B4	Chr01_26	Chr01_26F	ccctattgcagtggaggcttgc	Chr01_26R	ctttcgggagtagagctcactg

PP6	C4	Chr01_27	Chr01_27F	acaggaggatgtctgtgactcc	Chr01_27R	tctctcgcaccagccagactgc
PP6	D4	Chr01_28	Chr01_28F	aagctcagcgagacaatccgtc	Chr01_28R	tgatgttcccatagcagattcg
PP6	E4	Chr01_29	Chr01_29F	tcacaaaccttacaaccacgga	Chr01_29R	tgacagcccatgtctgaagcct
PP6	F4	Chr01_30	Chr01_30F	atagtagccaggaactctgctg	Chr01_30R	gagaatgtttgcaagctgtgtc
PP6	G4	Chr01_31	Chr01_31F	tgagttccaggacagccagagc	Chr01_31R	cattagcacctatcagctggtg
PP6	H4	Chr3_01	Chr3_01F	agcctgtgtacagctagactgg	Chr3_01R	cacctcaaaccctggcctgtgc
PP6	A5	Chr3_02	Chr3_02F	attcttgggcatggctcacacc	Chr3_02R	tgaacacccacatcctggcgga
PP6	B5	Chr3_03	Chr3_03F	aggtgggctggtctcttcgagg	Chr3_03R	tgctcacagaaggcaccaactg
PP6	C5	Chr3_04	Chr3_04F	caatggtcacccagcacgatgc	Chr3_04R	tagtcagcacacagcaagttcc
PP6	D5	Chr3_05	Chr3_05F	acaccagtttctgcccattagc	Chr3_05R	actctgtgtggtgggcgttagc
PP6	E5	Chr3_06	Chr3_06F	gagccatgagtcccaccatgtg	Chr3_06R	agacggacacacacatgcactc
PP6	F5	Chr3_07	Chr3_07F	gactgggtcttggcagaagctc	Chr3_07R	tgggtcggctgcagctggactg
PP6	G5	Chr3_08	Chr3_08F	aggtgtgtgtcatcacgctcag	Chr3_08R	gggctccttcacaactgagtgc
PP6	H5	Chr3_09	Chr3_09F	accatctagcttgcttgctctg	Chr3_09R	atcagttgggattgggttccag
PP6	A6	Chr3_10	Chr3_10F	atgtggacggctccctgaaacg	Chr3_10R	ggtgcagtcccagctcctcagg
PP6	B6	Chr3_11	Chr3_11F	tggcctctgactccatgctgag	Chr3_11R	gggaaatgggtaacgtgggctc
PP6	C6	Chr3_12	Chr3_12F	ttggcctcacttgggatgtacc	Chr3_12R	accaggcttgatggcaggctcc
PP6	D6	Chr3_13	Chr3_13F	tgcagctagaaatacgagctcc	Chr3_13R	aggattctggtctagcgactgc
PP6	E6	Chr3_14	Chr3_14F	gtgaccctactcaactgccagg	Chr3_14R	actcgggtcatcaggcttggtg
PP6	F6	Chr3_15	Chr3_15F	tggcttctgttagggaggactc	Chr3_15R	agtatccagcacggttcccagc
PP6	G6	Chr3_16	Chr3_16F	tccctcatgtctcggaacttgg	Chr3_16R	acgtccacttcctcacggatgc
PP6	H6	Chr3_17	Chr3_17F	tggcattcggacattcccggag	Chr3_17R	cttcgggcccaaatgttccagc
PP6	A7	Chr3_18	Chr3_18F	cccatcctggagatcagagtgc	Chr3_18R	tcatgaaccaattgctgcccag
PP6	B7	Chr3_19	Chr3_19F	tctaccaacaggagcacactcc	Chr3_19R	agcagttaacctcgcatgtgca
PP6	C7	Chr3_20	Chr3_20F	tgtctgtgtgatagtgcttgtg	Chr3_20R	agcttgatgatcttgggcattc
PP6	D7	Chr3_21	Chr3_21F	atctgtgcaatgacgactcctg	Chr3_21R	ccgctactgatgacgaccagtg
PP6	E7	Chr3_22	Chr3_22F	agattccctgcctcctgagtgc	Chr3_22R	agcgtcctagtcaggctgcaca
PP6	F7	Chr3_23	Chr3_23F	ggcgttcatgggtgttgcagtc	Chr3_23R	tgttggaggtgcgtcactgagg
PP6	G7	Chr3_24	Chr3_24F	tcccatgcctctaacaacctgg	Chr3_24R	acgtcttgataggcatctgtgc
PP6	H7	Chr3_25	Chr3_25F	cttgggcaatagggacacacac	Chr3_25R	actggcatgatggcacttgagg
PP6	A8	Chr3_26	Chr3_26F	ctggtacccaggagccaaactc	Chr3_26R	tgttgggcactgccttaggtgg
PP6	B8	Chr3_27	Chr3_27F	acctgggtcaggacgcctctgg	Chr3_27R	tctcttggccaccagcagctcc
PP6	C8	Chr3_28	Chr3_28F	agttgtcctctagttgcaccag	Chr3_28R	tgtgagttcaaggctaacctgc
PP6	D8	Chr3_29	Chr3_29F	acttcagacctcactcacacac	Chr3_29R	aggagtgctatgacccactgac
PP6	E8	Chr3_30	Chr3_30F	tgtgatggtctttccgaagcag	Chr3_30R	cgcatccttagagcacagtcac
PP6	F8	Chr3_31	Chr3_31F	ctgtgcctagaagaccacagtg	Chr3_31R	catgattccctaggaagccctg
PP6	G8	Chr3_32	Chr3_32F	agtggccttgggtcggaagtcc	Chr3_32R	agcaaccctaagcaaatgcagc
PP6	H8	Chr3_33	Chr3_33F	aggccactgttcctggtcactg	Chr3_33R	acccagtcaaaggcttgaccag
PP6	A9	Chr3_34	Chr3_34F	tggtgttccattgtctagctgg	Chr3_34R	ctccatctgaatcatcctggag
PP6	B9	Chr3_35	Chr3_35F	agtgctgctagctggtcctcag	Chr3_35R	tcaacgtggcaaagtctggcct
PP6	C9	Chr3_36	Chr3_36F	agactcaggtagtcaggcttgg	Chr3_36R	agaggaacagcgcagccaacag
PP6	D9	Chr3_37	Chr3_37F	tgacctctgtccaacagtgcag	Chr3_37R	tctgacggtatcaggcacacac
PP6	E9	Chr3_38	Chr3_38F	gaggccaagtgtccagcagtgg	Chr3_38R	accaatctgcatgccatcatgg
PP6	F9	Chr3_39	Chr3_39F	tctgtgacccaatgctgagagc	Chr3_39R	tggcaacctaaggctctgtgtg
PP6	G9	Chr3_40	Chr3_40F	attgtgggtgttgccatgtctg	Chr3_40R	agaggcaatgggtgtgtgtgtg
PP6	H9	Chr3_41	Chr3_41F	accatccaccaagccaagaacc	Chr3_41R	ggattggactttcacttgtcgtg
PP6	A10	Chr3_42	Chr3_42F	tcttcccagctaaggttgccac	Chr3_42R	actgccaagacaggttctccag
PP6	B10	Chr3_43	Chr3_43F	tccctcatgcacctatggccga	Chr3_43R	tggtgggttgggcagggtactg
PP6	C10	Chr3_44	Chr3_44F	tgtgagcccttatcagcagggt	Chr3_44R	gtgcctcaaccaagcaatctgg
PP6	D10	Chr3_45	Chr3_45F	tcaggcgcattatggcacacac	Chr3_45R	tagggctgccttgtctgactgc
PP6	E10	Chr3_46	Chr3_46F	acagtcccttgagagctccctg	Chr3_46R	tcccagaagccacgtcaggcag

PP6	F10	Chr3_47	Chr3_47F	aggcttctacaggaatgcacac	Chr3_47R	caaagcccagcctcttgtgtgg
PP6	G10	Chr3_48	Chr3_48F	tcacactccatcagagggttgg	Chr3_48R	tcaacagggaacaaagctgtcc
PP6	H10	Chr3_49	Chr3_49F	aggagcattggttccagagcac	Chr3_49R	actaacgggttaccatgcttgc
PP6	A11	Chr3_50	Chr3_50F	tgctcgatacacaattgcagct	Chr3_50R	cagacagcccttgaagctgacc
PP6	B11	Chr3_51	Chr3_51F	tggccgagtggattcttggagc	Chr3_51R	tgagaccttcaggtgagcaacc
PP6	C11	Chr3_52	Chr3_52F	atcctgcgagcttgctgactgg	Chr3_52R	cctttcagccagactcagatgc
PP6	D11	Chr3_53	Chr3_53F	cgcgaaaccagacaccaagacg	Chr3_53R	acaggctgccctcgaactcagc
PP6	E11	Chr3_54	Chr3_54F	actacatggcttgggctgaacc	Chr3_54R	gggattggatacatgcaactcac
PP6	F11	Chr3_55	Chr3_55F	agaggtctgttctcaagctcac	Chr3_55R	ctgagacactatgagcttctgc
PP6	G11	Chr3_56	Chr3_56F	cccaagtgataggttgcttgtg	Chr3_56R	ggctatgtcaaagtatggctcc
PP6	H11	Chr3_57	Chr3_57F	acttgaggcttaactcccacag	Chr3_57R	caacacccaaagctgtccggtg
PP6	A12	Chr19_17	Chr19_17F	cccggaaagcgtgagtgcctgc	Chr19_17R	accaaccagagtaccgttacac
PP6	B12	Chr19_18	Chr19_18F	caggaggattctcaggtcttgc	Chr19_18R	gatgtgctctcattggcatgac
PP6	C12	Chr19_19	Chr19_19F	tccagatcagccagggctactg	Chr19_19R	acgcatgcacacacacacatgc
PP6	D12	Chr19_20	Chr19_20F	tctggcctggtgaggcacttgg	Chr19_20R	agtcactgctcattggagccag
PP6	E12	Chr19_21	Chr19_22F	catgagcactgaatgagccacc	Chr19_22R	tgctgacaagcctgacaacctg
PP6	F12	Chr19_22	Chr19_23F	agtgtctgtgtagccctgactg	Chr19_23R	tacatgagcctggcaacccgac
PP6	G12	Chr19_23	Chr19_24F	cccttgccttgggacagtgacc	Chr19_24R	tgggtgctcatatacccacacg
PP6	H12	Chr19_24	Chr19_25F	gcattagtgcggctggctctgg	Chr19_25R	gagccacattccctggtagacg
PP7	A1	Chr05_01	Chr05_01F	ccacttagcaagtgttgggtcc	Chr05_01R	agaacacagcaatgacaggcca
PP7	B1	Chr05_02	Chr05_02F	atgactccagctgatgacccac	Chr05_02R	agtgctccctgaaagttcccac
PP7	C1	Chr05_03	Chr05_03F	aggctacctccgagcgtgtgtg	Chr05_03R	gagttgatccctaggacccgtg
PP7	D1	Chr05_04	Chr05_04F	tgattggacagtgctacagctc	Chr05_04R	ggaaccatctcacgcaatgtgg
PP7	E1	Chr05_05	Chr05_05F	tctgggagagctgttgtggctg	Chr05_05R	agttgcctccgcaaggcttcag
PP7	F1	Chr05_06	Chr05_06F	gcaagcatactctactgcaagc	Chr05_06R	acagggctttgtcagaaggctc
PP7	G1	Chr05_07	Chr05_07F	tcctggaaagactgtgtgaacc	Chr05_07R	tgagctttggcaggactcatcc
PP7	H1	Chr05_08	Chr05_08F	tacaccgtgagttccaggccag	Chr05_08R	ccacactgtaagcccagaagtg
PP7	A2	Chr05_09	Chr05_09F	tggctaactcctccaggcaacg	Chr05_09R	ttggaacaggtgaacgccctcg
PP7	B2	Chr05_10	Chr05_10F	agggtgtctggtcctttggagc	Chr05_10R	cagagtgagttccaggacagtc
PP7	C2	Chr05_11	Chr05_11F	ctgcttccacttcctaagggtg	Chr05_11R	gatcctggcagatttgcccagg
PP7	D2	Chr05_12	Chr05_12F	tgccagagtttggtgttgctgc	Chr05_12R	acagcttcagggagtacggtcc
PP7	E2	Chr05_13	Chr05_13F	tctggacactccatcagggtgc	Chr05_13R	ctggaactggagccacctagtg
PP7	F2	Chr05_14	Chr05_14F	tgagaaccaccttgctgctcag	Chr05_14R	ccaaagccagggtgcctcacag
PP7	G2	Chr05_15	Chr05_15F	acgctggtgctgtcgagtgtgc	Chr05_15R	ctccgagtgtcccaaggctgac
PP7	H2	Chr05_16	Chr05_16F	tctgtggtcccagctcacccag	Chr05_16R	tgagatgcacagcctaacgcac
PP7	A3	Chr05_17	Chr05_17F	ccttcagagtgccaccagatgg	Chr05_17R	cagcaagttcaaggccagcctc
PP7	B3	Chr05_18	Chr05_18F	caccgggaggctgcttaccagc	Chr05_18R	aaggctgttgttcggcatcccag
PP7	C3	Chr05_19	Chr05_19F	tgagcctctcctgtgtagcctg	Chr05_19R	agagacaggacagacttggagc
PP7	D3	Chr05_20	Chr05_20F	aggtgcccttcctgctttgcag	Chr05_20R	gcccatgtcaggagcaacaacc
PP7	E3	Chr05_21	Chr05_21F	ctgcatgtacgcagacccacac	Chr05_21R	agcatggctgttgagcggcatc
PP7	F3	Chr05_22	Chr05_22F	actggatgtgcccacacagtcc	Chr05_22R	actcacaacagcctggaactcc
PP7	G3	Chr05_23	Chr05_23F	tgtgctgtaatgggtgtgcgtg	Chr05_23R	accagaggtgaaggcacatagc
PP7	H3	Chr05_24	Chr05_24F	tgatcagctcacagccacatgg	Chr05_24R	acatgcgccacagtgcacagac
PP7	A4	Chr05_25	Chr05_25F	accctccagtcaccaaagcctg	Chr05_25R	acaggtctgatttccagggcag
PP7	B4	Chr05_26	Chr05_26F	aatgtgccctgaggcaacttgg	Chr05_26R	tgaactcacaggatgccatcac
PP7	C4	Chr05_27	Chr05_27F	ataagcagtactcagccattgc	Chr05_27R	agcacaaagacgtgctgtgagc
PP7	D4	Chr05_28	Chr05_28F	tggtgaagtccaggtcccagtc	Chr05_28R	tgtgtgctctctggaccttagc
PP7	E4	Chr05_29	Chr05_29F	ggacctgggttcaattcccagc	Chr05_29R	ggctgaaggcgacgcgctaagg
PP7	F4	Chr05_30	Chr05_30F	agccacctctttgccactgagg	Chr05_30R	agcaagatggctcagcagactc
PP7	G4	Chr05_31	Chr05_31F	agccttgcttcaggcccatagg	Chr05_31R	gcagcaattccaggtggtttgg
PP7	H4	Chr05_32	Chr05_32F	gccagcagaagccagtgccacc	Chr05_32R	tgggtgagttggctccagatgg

PP7	A5	Chr05_33	Chr05_33F	acccacagggtggctcacaacc	Chr05_33R	caccaagcctctggctcttagg
PP7	B5	Chr05_34	Chr05_34F	acccgtgttcaacgcctaagac	Chr05_34R	tgagctgccatgtgggtgctgg
PP7	C5	Chr05_35	Chr05_35F	accagctgagatacatccctgg	Chr05_35R	caaaggctcgttccaggctgag
PP7	D5	Chr05_36	Chr05_36F	tgctcatgggagcaccaggcag	Chr05_36R	acccttggagacactgaggcac
PP7	E5	Chr05_37	Chr05_37F	tgagtggacagacagggatgtg	Chr05_37R	aactcaggacttagagcctctg
PP7	F5	Chr05_38	Chr05_38F	acacacaaaccgaggtctggtc	Chr05_38R	gcttatgagtagcagtggtgtg
PP7	G5	Chr05_39	Chr05_39F	tcagtgcctggcgcagtacctg	Chr05_39R	gccctggactcctttggagcag
PP7	H5	Chr05_40	Chr05_40F	aggtatgcaccaccatgcccag	Chr05_40R	agcattccatagggctggcttg
PP7	A6	Chr05_41	Chr05_41F	tgttctggggactctgacaagc	Chr05_41R	tcacagcctagtgtggaactcc
PP7	B6	Chr05_42	Chr05_42F	agccttcttagggcctccaagg	Chr05_42R	aagccacgtgcccagcccaagc
PP7	C6	Chr05_43	Chr05_43F	actttctgccaatgcacacagg	Chr05_43R	ttctgacctcacacgaatgtgc
PP7	D6	Chr05_44	Chr05_44F	ttgaggagcaccctcatccagg	Chr05_44R	acattgctgacttctgttgcag
PP7	E6	Chr05_45	Chr05_45F	tctggcctgtgtgcacctagtc	Chr05_45R	acaatgcttggtgactgtggtg
PP7	F6	Chr05_46	Chr05_46F	cacaagtggaatcccgtggagc	Chr05_46R	agacacttcccgacctgttacg
PP7	G6	Chr05_47	Chr05_47F	agtgatggagacaggcagatcc	Chr05_47R	cttactggcaaaggcaaacacc
PP7	H6	Chr05_48	Chr05_48F	gcctaactctatactttgaggagc	Chr05_48R	tctgtagtgggatgaggccctc
PP7	A7	Chr05_49	Chr05_49F	gccttcgtgattgcagtttgtg	Chr05_49R	cagtctccacaggacccagagc
PP7	B7	Chr16_01	Chr16_01F	acttagtgaacccagtgtcagc	Chr16_01R	tcctcttgccactaagccttgc
PP7	C7	Chr16_02	Chr16_02F	ccggaaacccacctccgaaacg	Chr16_02R	gcagagcttgaagcccaacacc
PP7	D7	Chr16_03	Chr16_03F	ctgcacattccatccaaaggtg	Chr16_03R	tgtcctgactggcagccgacac
PP7	E7	Chr16_04	Chr16_04F	tcccagggctgtttccaggtgc	Chr16_04R	aggaacctcactttgtccaagc
PP7	F7	Chr16_05	Chr16_05F	tggagtggaggagcccaggtgg	Chr16_05R	gtgagactggtcaaggctgctc
PP7	G7	Chr16_06	Chr16_06F	ctgaatggctttggaagtcctg	Chr16_06R	tgaaggtgttgggtcttgcagg
PP7	H7	Chr16_07	Chr16_07F	gtaaacctcagtgtttccgtgg	Chr16_07R	tccagacttctcacccatctgg
PP7	A8	Chr16_08	Chr16_08F	tagggctgggccggtggttctg	Chr16_08R	ctgtaatctcagccctggagtg
PP7	B8	Chr16_09	Chr16_09F	actagcagagccacttcgggag	Chr16_09R	tcccaggatcgtccagctcagg
PP7	C8	Chr16_10	Chr16_10F	aggtcagcatcgggagtcttcc	Chr16_10R	tgcattagggccacatgctctg
PP7	D8	Chr16_11	Chr16_11F	agctagtcagtttgcttcaagg	Chr16_11R	agttccaggacagccagatagc
PP7	E8	Chr16_12	Chr16_12F	aggatagccattgtgaagctgc	Chr16_12R	cactctgccactaggctacacc
PP7	F8	Chr16_13	Chr16_13F	tggtgacccgcctggcactgag	Chr16_13R	tctctaaggcttgttgggcagc
PP7	G8	Chr16_14	Chr16_14F	cttaacaggtgaacgcttatgc	Chr16_14R	acagtattgaccagccttggga
PP7	H8	Chr16_15	Chr16_15F	acagctcgtctgactggggtcc	Chr16_15R	tgggaatccagagggtcctagc
PP7	A9	Chr16_16	Chr16_16F	taatccctgcactgtgaggcag	Chr16_16R	cctaagcaggtagtgagcccac
PP7	B9	Chr16_17	Chr16_17F	tagagagggtgtccttgtgcag	Chr16_17R	aggttggctgtggctggtagac
PP7	C9	Chr16_18	Chr16_18F	actccctagccgcaggctcacc	Chr16_18R	tcacaccgtcactgtagcaagc
PP7	D9	Chr16_19	Chr16_19F	tgaagtccagttccactgatgc	Chr16_19R	acagcagcctgtagctgtcacc
PP7	E9	Chr16_20	Chr16_20F	aggagactgccaccttcaggac	Chr16_20R	agcctcgtcacagtacaggcag
PP7	F9	Chr16_21	Chr16_21F	actcatgatcacacatgctctg	Chr16_21R	taagaggccacccagctacctg
PP7	G9	Chr16_22	Chr16_22F	tgttgtgaactatgaagccagc	Chr16_22R	aggctttcgtatgcaatgtgct
PP7	H9	Chr16_23	Chr16_23F	cctctgactgtaccaaggtctg	Chr16_23R	gaacaaagccacaggctacctg
PP7	A10	Chr16_24	Chr16_24F	aggagactgccaccttcaggac	Chr16_24R	agcctcgtcacagtacaggcag
PP7	B10	Chr16_25	Chr16_25F	actcatgatcacacatgctctg	Chr16_25R	taagaggccacccagctacctg
PP7	C10	Chr16_26	Chr16_26F	gcactggctgctcttccagagg	Chr16_26R	tgtgagttcaaggccaccttgg
PP7	D10	Chr16_27	Chr16_27F	ttacttagaacgcccatgtacg	Chr16_27R	gactttgagaagtccttggagc
PP7	E10	Chr16_28	Chr16_28F	tgcctgtgcttgtgtgctggtg	Chr16_28R	tgtgtagcaccatgcctcatgg
PP7	F10	Chr16_29	Chr16_29F	tgtggaaccctggctgtcctgg	Chr16_29R	ttgcgggtgttgggcaagactc
PP7	G10	Chr15_01	Chr15_01F	agaggtgtccttgggctttgag	Chr15_01R	tgattcccaggacccacatgac
PP7	H10	Chr15_02	Chr15_02F	tgactcccacaagtgtaccacg	Chr15_02R	agccacatggatggctactctg
PP7	A11	Chr15_03	Chr15_03F	tggtgtaggtgcacgtgtgtgc	Chr15_03R	gttccactgaatcgaactgggt
PP7	B11	Chr15_04	Chr15_04F	agtctgggctggcagatccgtg	Chr15_04R	ctgctcagagtggtgcttccag
PP7	C11	Chr15_05	Chr15_05F	tctctgtagctgagggtgactc	Chr15_05R	aggtcccagtatcccaagcaag

PP7	D11	Chr15_06	Chr15_06F	tctcctggcatcctcaggtagc	Chr15_06R	agtgtgtgatgtcgtttgtgtg
PP7	E11	Chr15_07	Chr15_07F	ggtacagagccctgtgtggagc	Chr15_07R	cccgagtttcatcctcaagacc
PP7	F11	Chr15_08	Chr15_08F	caaacagccttccaggttctgg	Chr15_08R	agacagtcacattggttgtgtg
PP7	G11	Chr15_09	Chr15_09F	tggacttgacactccttgcctg	Chr15_09R	tgttaccctcaacctatggtgg
PP7	H11	Chr15_10	Chr15_10F	cacgaggaagccagctttcagg	Chr15_10R	tccacacactgggcagagcctg
PP7	A12	Chr15_11	Chr15_11F	tccttaggctttgcccagtgtc	Chr15_11R	acaaagctgtttgagctgaacc
PP7	B12	Chr15_12	Chr15_12F	tgcctgtttgcctcccaaatgc	Chr15_12R	ttaaaggcgtgagccaccactg
PP7	C12	Chr15_13	Chr15_13F	tccagagtcaggtccaggtgag	Chr15_13R	tgccctgttgctcctggctgag
PP7	D12	Chr15_14	Chr15_14F	gcgaacagtgtcacttgtgtgc	Chr15_14R	acccagaaacgatgggcagtcc
PP7	E12	Chr15_15	Chr15_15F	agccatgccacttgcacgctgg	Chr15_15R	ttgtgaggagatccactagctg
PP7	F12	Chr15_16	Chr15_16F	ttcctgacagcatgttcacagg	Chr15_16R	agcagaacaccctgctgttgag
PP7	G12	Chr15_17	Chr15_17F	aaacagcgtccctcacagcctg	Chr15_17R	ctcactccagcaccgtgacagg
PP7	H12	Chr15_18	Chr15_18F	tcaggagcagaagcagtctgtc	Chr15_18R	tctaggccagtcagggctacac
PP8	A1	Chr15_19	Chr15_19F	agtgggcatagcgcatgacagc	Chr15_19R	tgaaacctggctaacctgggac
PP8	B1	Chr15_20	Chr15_20F	accagttaggtactctgacagc	Chr15_20R	tggtctacagagcgagttccag
PP8	C1	Chr15_21	Chr15_21F	agaacccgagttgagttcccag	Chr15_21R	taatcccagcacttcagaggct
PP8	D1	Chr15_22	Chr15_22F	tcctgtagcaaatcttgggtgg	Chr15_22R	agaacccgagtcctgcatcagc
PP8	E1	Chr15_23	Chr15_23F	atctgggacagattccagccag	Chr15_23R	tcccaaagccaagcctgagtgc
PP8	F1	Chr15_24	Chr15_24F	tccaggacagccagggctgcac	Chr15_24R	gcatgaagcctttggtgagcac
PP8	G1	Chr15_25	Chr15_25F	tccaggttcccagcttggtgag	Chr15_25R	tgcatgtgaccagtagcactgc
PP8	H1	Chr15_26	Chr15_26F	accagccagccaaggtgagtcc	Chr15_26R	agggttcctgagtcacgtgtcg
PP8	A2	Chr15_27	Chr15_27F	ggcaggtctgtgaaccatgtgc	Chr15_27R	gtgccacaatgccctgggtctg
PP8	B2	Chr15_28	Chr15_28F	actgctgacatagggaacacct	Chr15_28R	agtctgaccagcagacagctcg
PP8	C2	Chr15_29	Chr15_29F	agacccaggtttggttgccagc	Chr15_29R	aacattgcctcctgaagcctgg
PP8	D2	Chr15_30	Chr15_30F	gaggcagaggcaggtggatctc	Chr15_30R	tgctgagactaaaggtgtgtgc
PP8	E2	Chr15_31	Chr15_31F	tctcccatgtttgcttcagcga	Chr15_31R	tccaaccaaccaaccaccatgc
PP8	F2	Chr15_32	Chr15_32F	tctccggcagtggctgctcaag	Chr15_32R	tttgggagtggcccaaccactg
PP8	G2	Chr15_33	Chr15_33F	cagcaaccatgtgaggtcttgg	Chr15_33R	atggaggccaggaatcagcctc
PP8	H2	Chr15_34	Chr15_34F	agctatggccctgtcctgctgg	Chr15_34R	gcccaagctacaacgatgtcag
PP8	A3	Chr15_35	Chr15_35F	agggctacacaatgagaccctc	Chr15_35R	tgaaactgtagctctgtggtgg
PP8	B3	Chr15_36	Chr15_36F	agtcactgaccgcagccagctc	Chr15_36R	ttgtatcccggtacagcaggtc
PP8	C3	Chr15_37	Chr15_37F	aggactgtgtttgccacctcag	Chr15_37R	ccctgctggacctgcttgggag
PP8	D3	Chr17_01	Chr17_01F	acgcacagtagtcaatggctgc	Chr17_01R	cgacacttgcgtaaccacaagc
PP8	E3	Chr17_02	Chr17_02F	tgaatcgcattactgtaggcag	Chr17_02R	tggagcctgggtctgggtcctg
PP8	F3	Chr17_03	Chr17_03F	tgttggagctgaatacacgcac	Chr17_03R	aggcaacccaaacaccagggtc
PP8	G3	Chr17_04	Chr17_04F	gccataaccactgcggacccac	Chr17_04R	acacacgcacctgcacacatgc
PP8	H3	Chr17_05	Chr17_05F	aggccagggtctgtgtcccgag	Chr17_05R	tggaactcgccaactccgaagc
PP8	A4	Chr17_06	Chr17_06F	gtgagcccatttgtcagtgagc	Chr17_06R	caggcaccagccagtcgctagg
PP8	B4	Chr17_07	Chr17_07F	gcaccgtctccagtggacacac	Chr17_07R	gaacaagcacacaggcatttcc
PP8	C4	Chr17_08	Chr17_08F	ttgaaggacctatgtgccttcg	Chr17_08R	aggcgtgcaccaccaccctgag
PP8	D4	Chr17_09	Chr17_09F	aggctcattctagacagactcc	Chr17_09R	acaagggtcggctgtgtgtctg
PP8	E4	Chr17_10	Chr17_10F	aggtcccgtggagtgtgcgtgc	Chr17_10R	agccaaatgtgttgcctctagg
PP8	F4	Chr17_11	Chr17_11F	acagagtgagttccaggacacc	Chr17_11R	tcacattgctcaccacatggag
PP8	G4	Chr17_12	Chr17_12F	acagagtgagttccaggacacc	Chr17_12R	tcagttctggggaccgaacctg
PP8	H4	Chr17_13	Chr17_13F	agtccatctgcctgtgcaatgg	Chr17_13R	acagggagcagcccgatgtacc
PP8	A5	Chr17_14	Chr17_14F	tctatcagatgggctctgggag	Chr17_14R	tcatccatgcaaagtaggcctc
PP8	B5	Chr17_15	Chr17_15F	acatcttgctcttgctgaggac	Chr17_15R	gtgctaccacgtccggcaaagc
PP8	C5	Chr17_16	Chr17_16F	atagaggcagaatccagagtcc	Chr17_16R	tcccagcatttgggtggcagag
PP8	D5	Chr17_17	Chr17_17F	agccacaggttggacacacctc	Chr17_17R	ttacggatggtcgtgagccacc
PP8	E5	Chr17_18	Chr17_18F	gtggatgtgctaccaagctctg	Chr17_18R	tgggcattagagagctgacctg
PP8	F5	Chr17_19	Chr17_19F	tccgtggtgttgcggcggactc	Chr17_19R	ggatcttgctgtgtagctctgg

PP8	G5	Chr17_20	Chr17_20F	agtcctgcaagccctcagcagc	Chr17_20R	tcaaggccagactgagctgtgc
PP8	H5	Chr17_21	Chr17_21F	ccatatgccagctagagcctgc	Chr17_21R	agcaatctgaaccgtagtgtgc
PP8	A6	Chr17_22	Chr17_22F	agggctctccagtggtgcccac	Chr17_22R	ctacgagtcaggagtccagctc
PP8	B6	Chr17_23	Chr17_23F	aaactaggcgcccaagcccagc	Chr17_23R	gaatcccgagggtaggccatcc
PP8	C6	Chr17_24	Chr17_24F	cacggctaactgggctcactgg	Chr17_24R	tctgaggcacagtctcgtgcag
PP8	D6	Chr17_25	Chr17_25F	acacctggcaagaggcacccac	Chr17_25R	gagggcactgatcccactaacg
PP8	E6	Chr17_26	Chr17_26F	agtgcaagtgcgcggtgctgtg	Chr17_26R	acatccagccaaggccaggtcc
PP8	F6	Chr17_27	Chr17_27F	tcccagcatttgggtggcagag	Chr17_27R	atagaggcagaatccagagtcc
PP8	G6	Chr17_28	Chr17_28F	tccaatgctcactggttcttgg	Chr17_28R	ttggattacctctgaatgccag
PP8	H6	Chr17_29	Chr17_29F	acgtcctgcagaagtccagctc	Chr17_29R	agtggcatggtagctcgacacc
PP8	A7	Chr17_30	Chr17_30F	tggccttaaacttggagccacc	Chr17_30R	tgccttgcatcagttggtggtg
PP8	B7	Chr17_31	Chr17_31F	tccatggtacctcacaagggtc	Chr17_31R	ggcttagggtcaagtccgggtc
PP8	C7	Chr17_32	Chr17_32F	acccagctctaaggttggactg	Chr17_32R	tcactaatggctacaggccacg
PP8	D7	Chr17_33	Chr17_33F	acgtcatgggcttcacacccag	Chr17_33R	gacatcgccagtgctgccaagg
PP8	E7	Chr17_34	Chr17_34F	acaccagtgcgttcgttccgag	Chr17_34R	tcggttgtcaggcggtcgcatc
PP8	F7	Chr17_35	Chr17_35F	acttcgtcaaactcacacaagc	Chr17_35R	agctgtgccagaggtcacgagc
PP8	G7	Chr02_01	Chr02_01F	agggctctctgacctgtagcag	Chr02_01R	tgagttcaagtccacactggtc
PP8	H7	Chr02_02	Chr02_02F	gttaaggctgccgccatcagag	Chr02_02R	agcctgacaacctggaacccac
PP8	A8	Chr02_03	Chr02_03F	tgatgtaagcttggtttggtcc	Chr02_03R	gatcaatgcctttcacgagcag
PP8	B8	Chr02_04	Chr02_04F	tcaagggcagatccacgcccag	Chr02_04R	tgtaccaccataccagacaacc
PP8	C8	Chr02_05	Chr02_05F	tctgagggcacttgctgagctg	Chr02_05R	actgtgcccagctctagggtgc
PP8	D8	Chr02_06	Chr02_06F	gccctaggaaattgctctggac	Chr02_06R	aagtgatcggtcatcttggcag
PP8	E8	Chr02_07	Chr02_07F	aggtggctcaggactgcctgac	Chr02_07R	tctggaggagcagccagtgctc
PP8	F8	Chr02_08	Chr02_08F	gctccaacacagtgaaagctcc	Chr02_08R	tcccaggagctagcttggcgtg
PP8	G8	Chr02_09	Chr02_09F	tgcacacctcctggctatgtgg	Chr02_09R	aggtgtctgttgcctggtgacc
PP8	H8	Chr02_10	Chr02_10F	tgtagctgggtgtggtgtcaag	Chr02_10R	tcctgctaagagtctggctgca
PP8	A9	Chr02_11	Chr02_11F	aacactcctctaggctaagtcc	Chr02_11R	ggtgacgatcgcaagctgtctg
PP8	B9	Chr02_12	Chr02_12F	tggtagctttaaggtcttgctg	Chr02_12R	ggcctatgtatctgacagttgg
PP8	C9	Chr02_13	Chr02_13F	tcagccagctcatgtgtcatgc	Chr02_13R	acattgggcagctcacaaccac
PP8	D9	Chr02_14	Chr02_14F	acgcaagaccgagaagcactcc	Chr02_14R	aacatttcgctgaagcagacac
PP8	E9	Chr02_15	Chr02_15F	tgtgcgccatggaaacctgcac	Chr02_15R	accgcagctagtcaggcatagc
PP8	F9	Chr02_16	Chr02_16F	tgtgaactgagtttgaaccagc	Chr02_16R	acacacccaaacacggcttctg
PP8	G9	Chr02_17	Chr02_17F	acctcgcatgtgttcccaggag	Chr02_17R	gtacgtatgctaagcaggttctc
PP8	H9	Chr02_18	Chr02_18F	agcttgctgaacaaacctgacc	Chr02_18R	cgaggccgttagggatctgagc
PP8	A10	Chr02_19	Chr02_19F	tttagaggcgcacagcggtcac	Chr02_19R	tcggagcatagtccctcggtgc
PP8	B10	Chr02_20	Chr02_20F	aacaggctcctacacacaggtg	Chr02_20R	agagcacttgcctgccacacac
PP8	C10	Chr02_21	Chr02_21F	atcagcacagacgcacactcac	Chr02_21R	tctctgcaagaggtttgtgcag
PP8	D10	Chr02_22	Chr02_22F	tttgcctagcctgcctgaagtg	Chr02_22R	acctctgcctcccgaatgccag
PP8	E10	Chr02_23	Chr02_23F	tcccagcgctagtgtggcagag	Chr02_23R	ttctaccagctgagctggctgg
PP8	F10	Chr02_24	Chr02_24F	ctcgccatgtgtggcagtttgg	Chr02_24R	agtcacacatagagccgtacag
PP8	G10	Chr02_25	Chr02_25F	acgctcctgatctccagaggac	Chr02_25R	tgtcaccgtgtcccttgtggag
PP8	H10	Chr02_26	Chr02_26F	tgtagctctggctgacctggag	Chr02_26R	gcgcgtacacacacgtacacac
PP8	A11	Chr02_27	Chr02_27F	tgaggtccagtacgccagacag	Chr02_27R	aggtttcacgtgtcagcaacag
PP8	B11	Chr02_28	Chr02_28F	agcactggatgtgccatgatcc	Chr02_28R	tgcatacacgctgtagcacctg
PP8	C11	Chr02_29	Chr02_29F	tgttccagcagagggctcaagc	Chr02_29R	agaggtgacagtgtgcctgtgc
PP8	D11	Chr02_30	Chr02_30F	tcagctccgtataaacctggtc	Chr02_30R	aatctgggctacccagcaagac
PP8	E11	Chr02_31	Chr02_31F	actcgtgagacacacagtcctg	Chr02_31R	gtgtcagccaacaggcacttgc
PP8	F11	Chr02_32	Chr02_32F	acaactgagctacacctccagc	Chr02_32R	acctggcacagagatgcctagc
PP8	G11	Chr02_33	Chr02_33F	atgtacgcatccacacacatgc	Chr02_33R	cctagtccactacagcaggcac
PP8	H11	Chr02_34	Chr02_34F	ttgggcctgtgtacaatgcctc	Chr02_34R	aacgagcactttccaccatgag
PP8	A12	Chr02_35	Chr02_35F	actggcttccacgcaaaggtgc	Chr02_35R	cttccgaaagagcgtgagaacc

PP8	B12	Chr02_36	Chr02_36F	tgagccaagtgtcctaggtgtg	Chr02_36R	tggtctccataactgctactgg
PP8	C12	Chr02_37	Chr02_37F	tgagtgctgccatacccaggct	Chr02_37R	atgctcatgacaagggtttgct
PP8	D12	Chr02_38	Chr02_38F	actggtgcagggcaggcactcc	Chr02_38R	tgatctcagcaactcaggatgc
PP8	E12	Chr02_39	Chr02_39F	actccagtgttttgcctctgag	Chr02_39R	caaccacatggtggctcacagc
PP8	F12	Chr02_40	Chr02_40F	aggtcctgagttcaagtcccag	Chr02_40R	aacactggccatccatccgcag
PP8	G12	Chr02_41	Chr02_41F	agggaatcagatccctcaaagc	Chr02_41R	aagccagacaaaggccatcctc
PP8	H12	Chr02_42	Chr02_42F	agtaattccaggtcagtccagc	Chr02_42R	agctcaccctctctggtggagc
PP9	A1	Chr02_43	Chr02_43F	tgtgcagccgaggtgggcttcc	Chr02_43R	tggcatcctatgacaccaacac
PP9	B1	Chr02_44	Chr02_44F	ccaatggcttaggcactgttgg	Chr02_44R	tgggttcagtgagagaccttgc
PP9	C1	Chr02_45	Chr02_45F	agtggcctgagatgcacccagc	Chr02_45R	tggtacaggttatgcccagagg
PP9	D1	Chr02_46	Chr02_46F	gaacccatgtgaagccctatgg	Chr02_46R	acctaaggcccagtgagactcc
PP9	E1	Chr02_47	Chr02_47F	agtgtggttggtcgggcatttc	Chr02_47R	tctgacaccctcttgtggtgtg
PP9	F1	Chr02_48	Chr02_48F	gtgttatggagcacaccagcac	Chr02_48R	tggacattgctttgtgttgcac
PP9	G1	Chr02_49	Chr02_49F	cagggaccaaactcaggtcacc	Chr02_49R	gtgacaaccaacgaggtcttcc
PP9	H1	Chr02_50	Chr02_50F	gttgtgagctgtcatgtgggtg	Chr02_50R	agaccaggctgacctggaactc
PP9	A2	Chr02_51	Chr02_51F	cacggcttgtaaggctgaactg	Chr02_51R	gtccaatgtggttcgttgacag
PP9	B2	Chr02_52	Chr02_52F	tgcttggaagtgaacccaggtc	Chr02_52R	ttgtgagtgtcaggaaaccctc
PP9	C2	Chr02_53	Chr02_53F	tgagtgctcttaaccgctgagc	Chr02_53R	tgctgttgtcaccgtcacagag
PP9	D2	Chr02_54	Chr02_54F	tcaaccacgagacagatggtgc	Chr02_54R	agtcaggagcttgtggcagtcg
PP9	E2	Chr02_55	Chr02_55F	gaaacctgtgccctcgtggagc	Chr02_55R	tcccatgaagcctagcctgagc
PP9	F2	Chr02_56	Chr02_56F	gctagggcttgatgtgtgtagc	Chr02_56R	gcaagagcctccattgcgcctg
PP9	G2	Chr02_57	Chr02_57F	agttctgggagccagagtcgct	Chr02_57R	gtatagggcatttccagggctc
PP9	H2	Chr02_58	Chr02_58F	actcccaggcccagcgtctgag	Chr02_58R	aggatccaggacttcccagagg
PP9	A3	Chr02_59	Chr02_59F	agcacccttacacgcatgcaac	Chr02_59R	aggatccaggacttcccagagg
PP9	B3	Chr02_60	Chr02_60F	agaattatccagctcatggctg	Chr02_60R	accatactggccaccaacccag
PP9	C3	Chr02_61	Chr02_61F	tgatctgggagttgagcagcac	Chr02_61R	ccctaatggtgcacgcatgtgc
PP9	D3	Chr02_62	Chr02_62F	accctggaaagctgcaaacctg	Chr02_62R	tgcacacagcatccgggacctc
PP9	E3	Chr02_63	Chr02_63F	tgtgtcgtacagctgttccagg	Chr02_63R	gactgtacgccatcacaacagg
PP9	F3	Chr02_64	Chr02_64F	gcaatctgcagggttgaggacc	Chr02_64R	catgcattgtagtgttgactgg
PP9	G3	Chr02_65	Chr02_65F	ctaataaagcccatgacgccac	Chr02_65R	gagtttggcatcagtatgcacc
PP9	H3	Chr02_66	Chr02_66F	ccaagggcataagcaatagcct	Chr02_66R	ttaaggcaccatctctcagcag
PP9	A4	Chr14_01	Chr14_01F	atcgccaaggcactctggggtg	Chr14_01R	tgagtgctacactctgggaacc
PP9	B4	Chr14_02	Chr14_02F	aggtagggattctccggtcagc	Chr14_02R	tcaggtttgggagcaagcacct
PP9	C4	Chr14_03	Chr14_03F	agcatctcaaggaggcaactgc	Chr14_03R	aagtgcctaattgagctcctgc
PP9	D4	Chr14_04	Chr14_04F	tgagcctgtgagtcgtgggtgc	Chr14_04R	acacccacacccagcatcacag
PP9	E4	Chr14_05	Chr14_05F	gaccaggttcttaaccatgctg	Chr14_05R	agacctccacatgtgtgtctgg
PP9	F4	Chr14_06	Chr14_06F	cgtgtgccactacaccgggcag	Chr14_06R	aggcaaactttctgcaagggct
PP9	G4	Chr14_07	Chr14_07F	aaggcggagccaggccgatctc	Chr14_07R	attgtttgttggtgcgtgtgtg
PP9	H4	Chr14_08	Chr14_08F	tgacaggatcataggcttccac	Chr14_08R	acagacccacatcctaggcgtc
PP9	A5	Chr14_09	Chr14_09F	ccagaggtgactggcgttatcc	Chr14_09R	agggtgatcctgacacccgtgc
PP9	B5	Chr14_10	Chr14_10F	acaggaccacatcccatgactg	Chr14_10R	tgtcagccaggacacacgtagc
PP9	C5	Chr14_11	Chr14_11F	tctgtagctccagttccagagg	Chr14_11R	gattgccaagttctatggcagc
PP9	D5	Chr14_12	Chr14_12F	atctaggcagaagccagttcac	Chr14_12R	agcaaaccgatacgaggttctc
PP9	E5	Chr14_13	Chr14_13F	tagcccatgtataccatgccag	Chr14_13R	ggagccctggttgatctgggac
PP9	F5	Chr14_14	Chr14_14F	agctgagcgtgacagcaacctc	Chr14_14R	gtgcggcttgtgtgcgtatctg
PP9	G5	Chr14_15	Chr14_15F	gtgtgtccttatggctgtgacc	Chr14_15R	tcccaggtaaatgtgaccaacc
PP9	H5	Chr14_16	Chr14_16F	actgccattactgctgacatgg	Chr14_16R	gacatcggtgtgagccatctgc
PP9	A6	Chr14_17	Chr14_17F	aggcttctgctgcctctgacag	Chr14_17R	agccttccgactccaggttgtg
PP9	B6	Chr14_18	Chr14_18F	tgagaatgggagcatgtgcagg	Chr14_18R	acgttgaaggactctgacatgc
PP9	C6	Chr14_19	Chr14_19F	gcacctacaacaagcagctacc	Chr14_19R	tcagcttggagcttgctggagg
PP9	D6	Chr14_20	Chr14_20F	gtaccttgggcagtaggcccac	Chr14_20R	atcagttgctaacccgaggtgc

PP9	E6	Chr14_21	Chr14_21F	tgggttggatttcccagtcagg	Chr14_21R	tcctgtcggcagcacacacacg
PP9	F6	Chr14_22	Chr14_22F	gtagatggcagtgctctactgg	Chr14_22R	tcactggcccacttcctgaagc
PP9	G6	Chr14_23	Chr14_23F	tgcccaagcaagggaccagacc	Chr14_23R	ttggaccatcaaggtggttcag
PP9	H6	Chr14_24	Chr14_24F	tgaaagccagtatgagttgctg	Chr14_24R	ctgtagagggtgtgtgtgcctg
PP9	A7	Chr14_25	Chr14_25F	cctggaggctggtgggtagcac	Chr14_25R	tgctagggaagtcaggtggacc
PP9	B7	Chr14_26	Chr14_26F	tcacagagccaaataccagtgc	Chr14_26R	acaccatgaacccaagtccctg
PP9	C7	Chr14_27	Chr14_27F	tggccctaaactcaatcgccag	Chr14_27R	tcgcctgcctctacctcccgag
PP9	D7	Chr14_28	Chr14_28F	tgagaatgggagcatgtgcagg	Chr14_28R	gggttttcctgtctcacagagg
PP9	E7	Chr14_29	Chr14_29F	tgagaatgggagcatgtgcagg	Chr14_29R	gaacgttggttaggactctgac
PP9	F7	Chr14_30	Chr14_30F	acaatgaggttgagcttctgtg	Chr14_30R	gtaggcttggcatggcattcag
PP9	G7	Chr12_01	Chr12_01F	tcggtggttgtaagctgccctg	Chr12_01R	tgggaactgaacccaggtgtcc
PP9	H7	Chr12_02	Chr12_02F	actaccgagctaaatctccagc	Chr12_02R	gaacacttcgtgcccgtgtgtg
PP9	A8	Chr12_03	Chr12_03F	tgaccagactgtcccagaactc	Chr12_03R	agcatgggctgctcctctggag
PP9	B8	Chr12_04	Chr12_04F	acacacgagaaacctcctctgc	Chr12_04R	accgttgctaggtgaacgcgac
PP9	C8	Chr12_05	Chr12_05F	gctctctcaaatcgatggtctc	Chr12_05R	ctttcatggtaccagaaggctc
PP9	D8	Chr12_06	Chr12_06F	ccatctctccaggcccatggtg	Chr12_06R	tggtcagaacatgaatccctgc
PP9	E8	Chr12_07	Chr12_07F	tccagatggcaaatgcttggct	Chr12_07R	ctggtagagcaagctccaagac
PP9	F8	Chr12_08	Chr12_08F	tatggtgtgcgtgtacgtgtgc	Chr12_08R	taccttcgagggcactgggcac
PP9	G8	Chr12_09	Chr12_09F	aggctcatgtatctgaccactg	Chr12_09R	tgcatccagatccacaggcagc
PP9	H8	Chr12_10	Chr12_10F	ccatcacagagctggacaaacc	Chr12_10R	gggagtctgaccctggttctgg
PP9	A9	Chr12_11	Chr12_11F	tctgctctgatgcacctgcacg	Chr12_11R	cggcttggaacttgtgaggctg
PP9	B9	Chr12_12	Chr12_12F	tgaagagcatgttgtctgaagc	Chr12_12R	acttgggacagctcacaaccac
PP9	C9	Chr12_13	Chr12_13F	tgtttgcagtttgcctggctac	Chr12_13R	tctgacttccgggaaccacctg
PP9	D9	Chr12_14	Chr12_14F	gcagaccgaggcaggtaccctc	Chr12_14R	agtaagagcctacattgcctgc
PP9	E9	Chr18_01	Chr18_01F	tctacgcccggaagctttgtcg	Chr18_01R	acaaatagcgggcctcggagtc
PP9	F9	Chr18_02	Chr18_02F	tccacgtgtgtgacccttcagg	Chr18_02R	agtccagggttgctgacacacc
PP9	G9	Chr18_03	Chr18_03F	tgtctgtggaccgtggtgtacc	Chr18_03R	tgaccagctccaggtgactagc
PP9	H9	Chr18_04	Chr18_04F	tgcaacattagtgcctgtatgg	Chr18_04R	acccacaccatcatggacacac
PP9	A10	Chr18_05	Chr18_05F	tgaagtgacaaacgaggtcaac	Chr18_05R	catgccatactatgacccagag
PP9	B10	Chr18_06	Chr18_06F	tgaccataccggctattggca	Chr18_06R	ccaaggttcccaacagagcctg
PP9	C10	Chr18_07	Chr18_07F	tggtgttacctgagggctgctg	Chr18_07R	ggaattcactgctcaactcagc
PP9	D10	Chr18_08	Chr18_08F	tctgcatgagttcaagaccagc	Chr18_08R	gctagagcaggacctccattgg
PP9	E10	Chr18_09	Chr18_09F	tgactgatgcctccaggaatgg	Chr18_09R	actaggcaaacctgccctaagg
PP9	F10	Chr18_10	Chr18_10F	acgtgtgtgtgcgcgcgtatgc	Chr18_10R	caagaccacgttgcagtgttcc
PP9	G10	Chr18_11	Chr18_11F	tctgcactgaatatctgcatcg	Chr18_11R	aatgacagctcaggctgtactc
PP9	H10	Chr18_12	Chr18_12F	ttcaatgtgctaccatgcaacg	Chr18_12R	aagctgctgtctgcggtctgag
PP9	A11	Chr18_13	Chr18_13F	atccttgccaaccatcaggcac	Chr18_13R	ggctttcaagatggctcatctg
PP9	B11	Chr18_14	Chr18_14F	ttattcaaaccaccacagcagg	Chr18_14R	agctacctgtctgccatacacc
PP9	C11	Chr18_15	Chr18_15F	tgacacacttcctccaacaagg	Chr18_15R	caagtcaccaaaccaagctgtc
PP9	D11	Chr08_01	Chr08_01F	tcctgaagcagcaccagtgacc	Chr08_01R	agtccagggaagtgcaactgtc
PP9	E11	Chr08_02	Chr08_02F	accttagccagacccaaggacc	Chr08_02R	agcaccgttgtctgcactttgg
PP9	F11	Chr08_03	Chr08_03F	gctctgagaactggtgtgtgtg	Chr08_03R	tcctccttagaatccagctacag
PP9	G11	Chr08_04	Chr08_04F	acattgccatggatgcctgagt	Chr08_04R	actccaccagggcacacactgc
PP9	H11	Chr08_05	Chr08_05F	ccttgttggaggaagtgtgtcg	Chr08_05R	tgctgtgaagagacgccatgac
PP9	A12	Chr08_06	Chr08_06F	tgggtctgcctgccctagatgc	Chr08_06R	ggccctgagatagccacatcac
PP9	B12	Chr08_07	Chr08_07F	agatgtgctgccctctgtgagg	Chr08_07R	acttgaccggcccagctgagtg
PP9	C12	Chr08_08	Chr08_08F	tgcctctgtcagggtgggacct	Chr08_08R	agctcacaggccagcgaacctg
PP9	D12	Chr08_09	Chr08_09F	agattgtccaagcatgctctgg	Chr08_09R	tctgtcaagtggacatcgtgag
PP9	E12	Chr08_10	Chr08_10F	gatcaaggcccaggctacacag	Chr08_10R	tgccgacctagaggctgctctg
PP9	F12	Chr08_11	Chr08_11F	tggaacactcaagcaattccca	Chr08_11R	tagctcccaaggacagttgacg
PP9	G12	Chr08_12	Chr08_12F	acttagagcacagaatgcccac	Chr08_12R	tcaaaggcatggaccatcatgc

PP9 H12	Chr08_13	Chr08_13F	acttggtaccgcatcagcagac	Chr08_13R	tggcccatcaagtggaaggctc
PP10 A1	Chr08_14	Chr08_14F	tggttaggtccatacctggagg	Chr08_14R	ctgcaaaggctgacgcatccac
PP10 B1	Chr08_15	Chr08_15F	gcgcacaggccaatctgccgca	Chr08_15R	ttcttccaggtggcaagatgct
PP10 C1	Chr08_16	Chr08_16F	gaacttgctctgttccaagctg	Chr08_16R	cgcgcctgggtacgtgtttgtg
PP10 D1	Chr08_17	Chr08_17F	catgtctgagtcggcctgtagc	Chr08_17R	agagtcactgtatatctctggctg
PP10 E1	Chr08_18	Chr08_18F	tgtaccatgcttgtggcatctc	Chr08_18R	ggggtcagaactcacagtcctg
PP10 F1	Chr08_19	Chr08_19F	ctatgcatgcttgtgcaggtgc	Chr08_19R	tggcctctgatgaccacaaacg
PP10 G1	Chr08_20	Chr08_20F	acaaccgtgaccaacccattgg	Chr08_20R	attcggaaccttgctaagcctg
PP10 H1	Chr08_21	Chr08_21F	ccatcaccctaacagggcttcc	Chr08_21R	gcgcacaggccaatctgccgca
PP10 A2	Chr08_22	Chr08_22F	agaaagcagcttacagtcccag	Chr08_22R	tggattgggatgagccttgaac
PP10 B2	Chr08_23	Chr08_23F	agggcatcagctcccactaagg	Chr08_23R	tggaatcacacagctgagccac
PP10 C2	Chr08_24	Chr08_24F	aggactccataccagcctgagc	Chr08_24R	gctaacccggtgggaggcactc
PP10 D2	Chr08_25	Chr08_25F	actgcgataccctggctggtgg	Chr08_25R	tgcccagcgggaggtgcatgtc
PP10 E2	Chr08_26	Chr08_26F	gaccaccatacatgcatcatgg	Chr08_26R	tggcaaagccatttgcaggagc
PP10 F2	Chr08_27	Chr08_27F	caccagccaggaatgttccagc	Chr08_27R	cctgatgtgaatgccctgtgtg
PP10 G2	Chr08_28	Chr08_28F	gacaaggcaagctacctggcac	Chr08_28R	actgtgggaactcctttggacg
PP10 H2	Chr08_29	Chr08_29F	actgcggtgatgacggtcgtgc	Chr08_29R	ctgtcccgatctggagacttcc
PP10 A3	Chr08_30	Chr08_30F	tcagcccggagagagccacagc	Chr08_30R	acaagggtcttgtggacatgtg
PP10 B3	Chr08_31	Chr08_31F	aagcacatgccaccacacctgg	Chr08_31R	atcccaagacctcacatggcag
PP10 C3	Chr08_32	Chr08_32F	tgtatatgggctttgcatgtgg	Chr08_32R	acccagtctcctagactcccag
PP10 D3	Chr08_33	Chr08_33F	aggtgtgtgccaccatgtcagc	Chr08_33R	tccagggcatcaagggctgcac
PP10 E3	Chr08_34	Chr08_34F	gttccaggacagtcaaggttgc	Chr08_34R	ggtgtgagttgccaagcctgac
PP10 F3	Chr08_35	Chr08_35F	gctcactggcctatcagcctgg	Chr08_35R	acgagtgagccatcaagcaagc
PP10 G3	Chr08_36	Chr08_36F	tccaatagcatctggttgatgg	Chr08_36R	ggcatcaagccttcttggagga
PP10 H3	Chr09_1	Chr09_01F	tctctgggagtttgaggtcagc	Chr09_01R	tgcctcagccttctcagtggag
PP10 A4	Chr09_2	Chr09_02F	tctgttgcatgtgtgcatgtgc	Chr09_02R	aggagcctggccacctgagtcc
PP10 B4	Chr09_3	Chr09_03F	acgagggtatccctgagtgtgg	Chr09_03R	gttctgtgccgagggaactcac
PP10 C4	Chr09_4	Chr09_04F	caagccatgctatcaaccacac	Chr09_04R	ccaggttacatattgagaccctg
PP10 D4	Chr09_5	Chr09_05F	gagtgacagagatggtggcagc	Chr09_05R	acataggtcccgtgccttgtcc
PP10 E4	Chr09_6	Chr09_06F	caagtgacactcaacttgccag	Chr09_06R	cgaggccatcctctcgaactgg
PP10 F4	Chr09_7	Chr09_07F	acgtttcaggacagtgtttgca	Chr09_07R	tgcctctgccttctgagtgcag
PP10 G4	Chr09_8	Chr09_08F	acgtgtctacccactcacatgc	Chr09_08R	tcctagtgctgtagtggtgctg
PP10 H4	Chr09_9	Chr09_09F	gtacgcggtaatgcacatcctg	Chr09_09R	tccagaatgcaccaggtctgca
PP10 A5	Chr09_10	Chr09_12F	tgtctcccgactatgagcactg	Chr09_10R	caacttactcagttcccagatgtc
PP10 B5	Chr09_11	Chr09_11F	acacagccacctacgacttcag	Chr09_11R	atgaagagttccaggctccagc
PP10 C5	Chr09_12	Chr09_12F	actcactgtgtagaccagactg	Chr09_12R	tgcactatgtcctaggctccac
PP10 D5	Chr09_13	Chr09_13F	gccagctgaggtacatgacacc	Chr09_13R	agggtacagctcgtgtgtggtc
PP10 E5	Chr09_14	Chr09_14F	aagccatgccttcttggaggtg	Chr09_14R	ttctttaaagggcactaggcag
PP10 F5	Chr09_15	Chr09_15F	tcagtacccatatgatggcgac	Chr09_15R	actaggcttggtggcaagcacc
PP10 G5	Chr09_16	Chr09_16F	tcaggagttctagaccaacctg	Chr09_16R	tcgccatgcaacgctgacgacc
PP10 H5	Chr09_17	Chr09_17F	cctgaaggctgtctctgactgg	Chr09_17R	gtctgggctacatgagaccctg
PP10 A6	Chr09_18	Chr09_18F	tccaccaaagcacgaatgcacc	Chr09_18R	tgtcagtgtgtagcggtctagg
PP10 B6	Chr09_19	Chr09_19F	tcccgagagtggtcggacagtc	Chr09_19R	tgtttgatgctcggtttcctgg
PP10 C6	Chr09_20	Chr09_20F	tgcaggaagactcccggacttg	Chr09_20R	agactccacttgggacaactgc
PP10 D6	Chr09_21	Chr09_21F	tgtctgtcagcgggctccctgc	Chr09_21R	tctgtgggagtctccggtggga
PP10 E6	Chr09_22	Chr09_22F	atggagtgcattggaccctctg	Chr09_22R	tggtctgcttggaaggcgtgct
PP10 F6	Chr09_23	Chr09_23F	acagatgccttgagtcagcacc	Chr09_23R	agggtctttgctcactggaggt
PP10 G6	Chr09_24	Chr09_24F	tcttaggctggagagatggctc	Chr09_24R	acagtggatcccagcatagtgc
PP10 H6	Chr09_25	Chr09_25F	ctcgaactcagaaatccgcctg	Chr09_25R	aaagttgacctttggcctacac
PP10 A7	Chr09_26	Chr09_26F	tgcagttgaaacccttgcagag	Chr09_26R	acttcctggtcagctagctcag
PP10 B7	Chr09_27	Chr09_27F	ttagccacactaggcaagtcac	Chr09_27R	agttcccattaaccacagtagc

PP10 C7	Chr09_28	Chr09_28F	aggttgtcaccaagcctgatga	Chr09_28R	tcgggctctccaaagaccctgg
PP10 D7	Chr09_29	Chr09_29F	tcttagcaccacatggcagctc	Chr09_29R	tctagcctgtgtcaagctgacg
PP10 E7	Chr09_30	Chr09_30F	atgcaaaccgaagcaaacctcc	Chr09_30R	aggtcaactgtctccagcaagc
PP10 F7	Chr09_31	Chr09_31F	agaagcactgaagcagtgtgtg	Chr09_31R	agatgtctccatggtcaagagc
PP10 G7	Chr04_01	Chr04_01F	acagaccctgtagacccatcac	Chr04_01R	ggtgtaccactaccctaagcac
PP10 H7	Chr04_02	Chr04_02F	tctggtccacagagcaagttcc	Chr04_02R	ggtcagtgcacggccatacctg
PP10 A8	Chr04_03	Chr04_03F	gcttgaggcttctgggttgctg	Chr04_03R	tgtggctgttcccacacagcag
PP10 B8	Chr04_04	Chr04_04F	agttctgagcatcccacagtgc	Chr04_04R	tgcgacagacttcgctttgcag
PP10 C8	Chr04_05	Chr04_05F	agctgtgccctgggttcaggtc	Chr04_05R	ctgatccaggctgaccacagtg
PP10 D8	Chr04_06	Chr04_06F	tgtgtctgttgttgggtgactg	Chr04_06R	ccagtacgaggacaacctggtc
PP10 E8	Chr04_07	Chr04_07F	aggctccaccacggagcacctc	Chr04_07R	ggaggttgagtccagcctagtc
PP10 F8	Chr04_08	Chr04_08F	tgcacacatacaaacgtaccac	Chr04_08R	gcaaagtggttctgttagcagc
PP10 G8	Chr04_09	Chr04_09F	tgacacacacctgtcgtcgcag	Chr04_09R	tgggcagcgttcgaccactgag
PP10 H8	Chr04_10	Chr04_10F	tgatgtgtcaggctcggcaagc	Chr04_10R	acctgggcctcagcttagccag
PP10 A9	Chr04_11	Chr04_11F	tctgacaggtctgcttcacctg	Chr04_11R	acaaagtggccaggctccagtc
PP10 B9	Chr04_12	Chr04_12F	tcaaggagtcacagtggtgctg	Chr04_12R	taagccatggtctcccgcacag
PP10 C9	Chr04_13	Chr04_13F	tgaatctgtgacgttggcccag	Chr04_13R	ccagctgaatgtggtggttcac
PP10 D9	Chr04_14	Chr04_14F	gacctccaacatcagccttagg	Chr04_14R	gtcaggcttggtggctagcacc
PP10 E9	Chr04_15	Chr04_15F	tctgcttgctaagtgtgaatgc	Chr04_15R	atgcaaccttaagatggtgctg
PP10 F9	Chr04_16	Chr04_16F	agctcttgctgtcctggaactc	Chr04_16R	agctgttgcctgtctgtggacc
PP10 G9	Chr04_17	Chr04_17F	gccagcatctgcctgggatgtg	Chr04_17R	gtcaagcctacaacatccaagc
PP10 H9	Chr04_18	Chr04_18F	gattcctgggattgtgggtctc	Chr04_18R	actgatgtctacagtgtctccag
PP10 A10	Chr04_19	Chr04_19F	aggactccaccactggcatcag	Chr04_19R	acgcctttaatcccagcactcg
PP10 B10	Chr04_20	Chr04_20F	acttacgtgctcccacacccag	Chr04_20R	acccttgcggagtctgtcggag
PP10 C10	Chr04_21	Chr04_21F	actgctctgtgggctccaatgc	Chr04_21R	tgcacagccatttcaaagccag
PP10 D10	Chr04_22	Chr04_22F	tctcccgagtcctccgcacagg	Chr04_22R	tgtttcccaggctggcaatctg
PP10 E10	Chr04_23	Chr04_23F	acggcagagctgggactccagc	Chr04_23R	tgctagtcagtgtccctaggac
PP10 F10	Chr04_24	Chr04_24F	acagcaccactaactgggaacc	Chr04_24R	acaagcctcaggtgcgatttgc
PP10 G10	Chr04_25	Chr04_25F	tgtgagccaccatgtgggcacg	Chr04_25R	tcatctggcttgtgtgtggcag
PP10 H10	Chr04_26	Chr04_26F	tgctactgtgctggaccgttgg	Chr04_26R	ggctctgtcaaaggcctccgtg
PP10 A11	Chr04_27	Chr04_27F	tctaagcacctatgtcaaggca	Chr04_27R	gggatctaccttaggtcatcagg
PP10 B11	Chr04_28	Chr04_28F	tccacttccagggccactagtc	Chr04_28R	tgattgcgtagactctgggtgc
PP10 C11	Chr04_29	Chr04_29F	tgcccaagtttgatccctggag	Chr04_29R	acctcccaagtgctagggttgc
PP10 D11	Chr04_30	Chr04_30F	aggatatggcctttgaagcagc	Chr04_30R	tcagaaggtctaggaagtctgc
PP10 E11	Chr04_31	Chr04_31F	tgaataatgtggcactgacctg	Chr04_31R	agagataaggccagccattctg
PP10 F11	Chr04_32	Chr04_32F	acagagtccacggaggctttgg	Chr04_32R	agaatgccttactcagaaccca
PP10 G11	Chr04_33	Chr04_33F	tgtgggtgtgcgagcatgtacc	Chr04_33R	agcctgatgcagtggctcacag
PP10 H11	Chr04_34	Chr04_34F	atgtaggcaaatacccatgcac	Chr04_34R	tgactcccagatgtgtgctgtg
PP10 A12	Chr04_35	Chr04_35F	tcccagagccagagcttccagc	Chr04_35R	tcctgggatctccgaagtctgg
PP10 B12	Chr04_36	Chr04_36F	agacctgagttcaggattcagc	Chr04_36R	tgtacttgcacctcctgggcac
PP10 C12	Chr04_37	Chr04_37F	tggcctcgaactcagagacctg	Chr04_37R	tgcctaaggacacacatgcacg
PP10 D12	Chr04_38	Chr04_38F	tgcaagctggcctgttgactgg	Chr04_38R	aaggttagctttcgtagtgcag
PP10 E12	Chr04_39	Chr04_39F	agaaacggatgatgcagttgtc	Chr04_39R	ttagtgatccaaagcccaggtg
PP10 F12	Chr04_40	Chr04_40F	ctgtgggttgaaggccattctg	Chr04_40R	caacgagtccttcgtcacgtgc
PP10 G12	Chr04_41	Chr04_41F	aggacccaggtttggttcccag	Chr04_41R	agtcacacaccactacacctgg
PP10 H12	Chr04_42	Chr04_42F	aaaccaccagccacttggcctg	Chr04_42R	ttcgtgtgtcacgtgtccaagg

Supplement 3 Primer sequences of the microsatellite loci genotyped along the candidate regions.

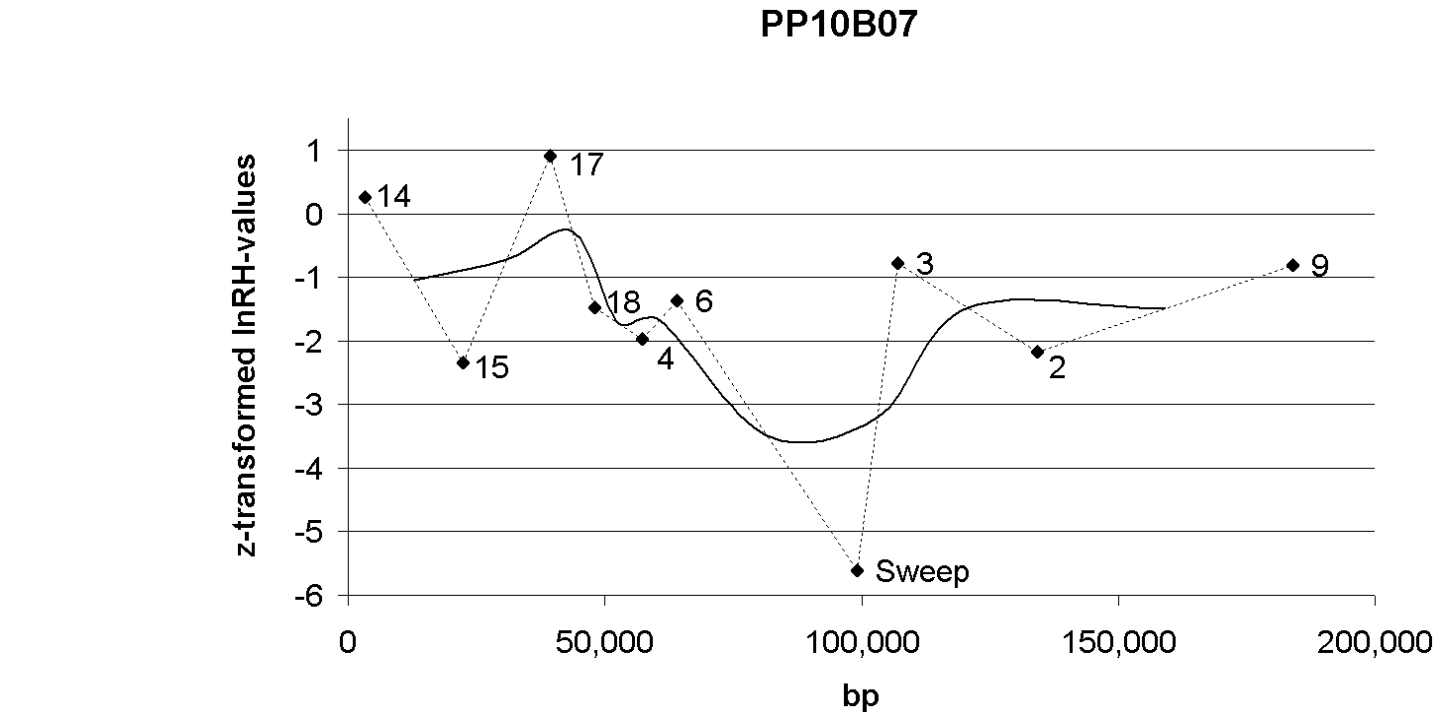
Markername	Primer	Sequence	Primer's Name	Sequence
PP9A07_12	PP9A07_12_F	tgccacgggaaccctctcggtg	PP9A07_12_R	tggctttcagtacactggacag
PP9A07_14	PP9A07_14_F	aggctgacgcacatggcggctg	PP9A07_14_R	tcctgaagccagacctcctgag
PP9A07_16	PP9A07_16_F	gtggaccgtgatgcaccttagc	PP9A07_16_R	ccagtcagtctagcttaggcag
PP9A07_17	PP9A07_17_F	acaacacagatggacatccctg	PP9A07_17_R	tcacttggcttacagaagtgct
PP9A07_18	PP9A07_18_F	agccatcagcttacttgagcag	PP9A07_18_R	cctgggatctgctgtgtgatgg
PP9A07_3	PP9A07_3_F	tcaaggtccagggctgagtctg	PP9A07_3_R	atgcaagcctgttgacctgagc
PP9A07_4	PP9A07_4_F	agacctccaaagcccggatcag	PP9A07_4_R	tgaccaccagccaaggaccctg
PP9A07_5	PP9A07_5_F	agccttctgtctgcctgagtgc	PP9A07_5_R	cagattcagctggactggctgg
PP9A07_6	PP9A07_6_F	atggtcgctgggcagagctgtc	PP9A07_6_R	cactaggccatgtctgcaaacc
PP9A07_8	PP9A07_8_F	ttactaagcctgccttgagagg	PP9A07_8_R	caaagattgtgtcctgggaacc
PP9A07_9	PP9A07_9_F	accgatggcttgagttcggtcc	PP9A07_9_R	tttggttgcacatatgcgtgtg
PP9A07_10	PP9A07_10_F	agattgtggtcttggtttctgg	PP9A07_10_R	aggaaatggcttccaccatacc
PP10B07_14	PP10B07_14_F	tgtctccaggcactggcatagg	PP10B07_14_R	tggatgcctgcgcacctagcag
PP10B07_15	PP10B07_15_F	tggcttgactacacatgcacac	PP10B07_15_R	ggcctcagacaggcagcaatcc
PP10B07_17	PP10B07_17_F	acttcgatgctttccgtcagtg	PP10B07_17_R	tgtcaggcatctcccaactgtg
PP10B07_18	PP10B07_18_F	tgtcaggctttagctgagttgg	PP10B07_18_R	agctgactgaggccgtggtgtc
PP10B07_4	PP10B07_4_F	agccttgggtcaggctcagtgg	PP10B07_4_R	tgtgcagagcattgatggctgc
PP10B07_6	PP10B07_6_F	accagggcagtcacagtgttcc	PP10B07_6_R	actgggcatttgtgtgtccgtg
PP10B07_3	PP10B07_3_F	atagacgtgctctgcatgactg	PP10B07_3_R	aggagccttgtgactacagcca
PP10B07_2	PP10B07_2_F	acctgccttgtaggcctgggac	PP10B07_2_R	tggtcctgaggcaccaaccctg
PP10B07_9	PP10B07_9_F	aaactcagggctttgggcgtgc	PP10B07_9_R	tctggacctgcttccaacacac
>PP7F04_5	PP7F04_5_F	agacacctggattgtcctctgg	PP7F04_5_R	tgtagactcacagggcacctgg
>PP7F04_7	PP7F04_7_F	gttgctatgggcatgcttgcac	PP7F04_7_R	agatgtcacccagcacacctgg
PP7F04_2	PP7F04_2_F	accaggagtccaaggctaacct	PP7F04_2_R	ccaaatccacacaagcacgcac
PP7F04_3	PP7F04_3_F	ttgttgtctagtgtacggttgg	PP7F04_3_R	aggctgacaaacacctgcaagc
PP7F04_4	PP7F04_4_F	tacctctgtgtgcacgtgagtg	PP7F04_4_R	ggaaggagctatgggtccacct
PP5F12_3	PP5F12_3_F	aggcaccagtaggtgcacatcc	PP5F12_3_R	tccagacaaggcagtgcaagct
PP5F12_4neu	PP5F12_4neu_F	tgcagatggacagccagcatcg	PP5F12_4neu_R	acacactgagcaggttgcgttc
PP5F12_8	PP5F12_8_F	ctggcaagactcagcctcacag	PP5F12_8_R	gtccagacaatccctcacagac
PP5F12_5	PP5F12_5_F	cttaacgaagtttccagtggct	PP5F12_5_R	caagatgctgaaacacacacac
PP5F12_6	PP5F12_6_F	ccacaagcatccgagaaccctg	PP5F12_6_R	tgccatcaagccacatccaagc
PP5F12_7	PP5F12_7_F	agtaaggtgggaaatgtggacc	PP5F12_7_R	ggccagcctgctctacatggca
PP8D01_18	PP8D01_18_F	tgtggtgtacctgctacacgag	PP8D01_18_R	tgccatagtccaggcctctgtg
PP8D01_20	PP8D01_20_F	agccaagtaccaagtgctcgtc	PP8D01_20_R	gcagtgagtttcatggtagcac
PP8D01_7	PP8D01_7_F	tgtgggacctccagacaccctg	PP8D01_7_R	tgaggcccgcttgcctacagag
PP8D01_1	PP8D01_1_F	agatgccttgcaaccacttgtg	PP8D01_1_R	gagacactgcctacctggcagc
PP8D01_2	PP8D01_2_F	gttggccacctaggaactggtc	PP8D01_2_R	aattcccagcaaccacatggtg
PP8D01_3	PP8D01_3_F	aggttcacacacatgacagctc	PP8D01_3_R	agcaccacctggctgagtccag
PP8D01_4	PP8D01_4_F	ttgttgttgctaggttggttgg	PP8D01_4_R	tgagcttgataccagcttggtc
PP8D01_10	PP8D01_10_F	tcagtttgctgttagccacgag	PP8D01_10_R	tcaccattgctagccatcagag
PP8D01_11	PP8D01_11_F	tcagatcccacacacggaggtc	PP8D01_11_R	agtgggagacctgagcctgtcc

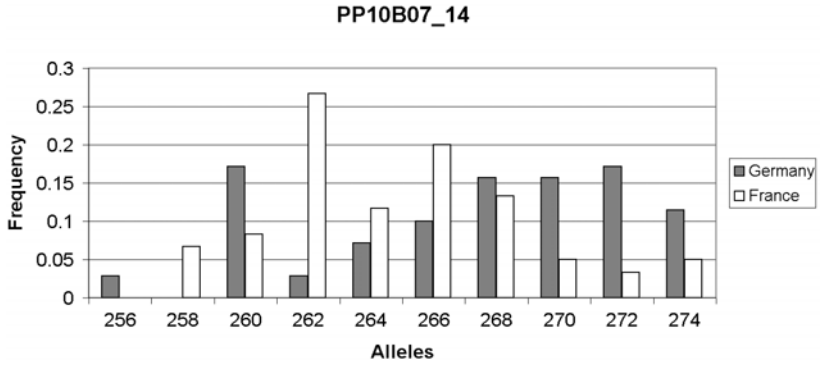
PP8D01_12	PP8D01_12_F	aagccgagtgtggtgttcatgc	PP8D01_12_R	aagccttaggaccaccgctcag
PP8D01_15	PP8D01_15_F	cagattgagttccaggacaacc	PP8D01_15_R	agttgccctgttgtggatgtgc
PP8D01_16	PP8D01_16_F	tgcaccgcacaaatgccacctg	PP8D01_16_R	ggattgacacatggcccactgg
PP4H12_8	PP4H12_8_F	actctgtgacctcccaggttag	PP4H12_8_R	ggccacaagccagccaagtctg
PP4H12_1	PP4H12_1_F	tgcatcctaccaagcaagccac	PP4H12_1_R	cccaggccaacttgcgttacag
PP4H12_10	PP4H12_10_F	ctagccctggagagtgctgacc	PP4H12_10_R	gaagttccctttggcttgctgg
PP4H12_3	PP4H12_3_F	tgggtgttgccctaagcccagg	PP4H12_3_R	agcaagcctgtgctcacacagc
PP4H12_12	PP4H12_12_F	aatgggcatgtttcaagtgtgc	PP4H12_12_R	ctgctgtgacctggctggaacc
PP4H12_2	PP4H12_2_F	accttcaagactggatctggtg	PP4H12_2_R	tttccagcctgttgggtttagc
PP4H12_5	PP4H12_5_F	agggccaagaagtgggagtctg	PP4H12_5_R	tcaggactctacagcagtggct
PP4H12_4	PP4H12_4_F	cctcacaggaagtcatcacagc	PP4H12_4_R	acgcctttaatcccagcactcg
PP3H03_2	PP3H03_2_F	tctaggacggccagggctacac	PP3H03_2_R	acggttgctatggcactgttgc
PP3H03_3	PP3H03_3_F	tgccaggacattcctgagcacc	PP3H03_3_R	agaccctgagccctgggaccac
D10Mit23	D10Mit23_F	tgagccctgccgtgtgcctcag	D10Mit23_F	caagccctaggttcccatgagc
PP3H03_5	PP3H03_5_F	tgctgaaggacctccacattgg	PP3H03_5_R	agctcctcctgcctcccaagag
PP3H03_6	PP3H03_6_F	attgtcctgttgggaaccttgg	PP3H03_6_R	aggtctcaaggctgggctgctg
PP3H03_7	PP3H03_7_F	cattcacaggatccatgtccag	PP3H03_7_R	aggcaacatggtctccggaacc
D1Mit305	D1Mit305_F	gtgggaaccttctactatttctatgc	D1Mit305_R	gtgtacctcctttctgtttatggg
PP6E03_1	PP6E03_1_F	agaccaatgtctttggacacag	PP6E03_1_R	gcacttcccaacacaaaggtcc
PP6E03_3neu	PP6E03_3neu_F	ccttgcttgcagagagtcccac	PP6E03_3neu_R	tgggcattgaaacctcagaacc
PP6E03_2	PP6E03_2_F	ggacagacacctgaggttgtcc	PP6E03_2_R	gacagacacctactctgtgtgg
D1Mit440	D1Mit440_F	tccacacaaggtgtcctctg	D1Mit440_R	gctcaggtgacctccaaaac
PP6E03_5	PP6E03_5_F	cacccattgacagctccaagac	PP6E03_5_R	ccacatgatggttcacaaccac
PP6E03_6	PP6E03_6_F	tgtcacatgggctcccgcaagc	PP6E03_6_R	tctggatggtctagcaagcacc
PP6E03_7	PP6E03_7_F	atgccaaaccttagctctggag	PP6E03_7_R	acactttgacaccgacccacag
PP6E03_9	PP6E03_9_F	acactggactagacccagaacg	PP6E03_9_R	agactgttgtgcacctctgtgg

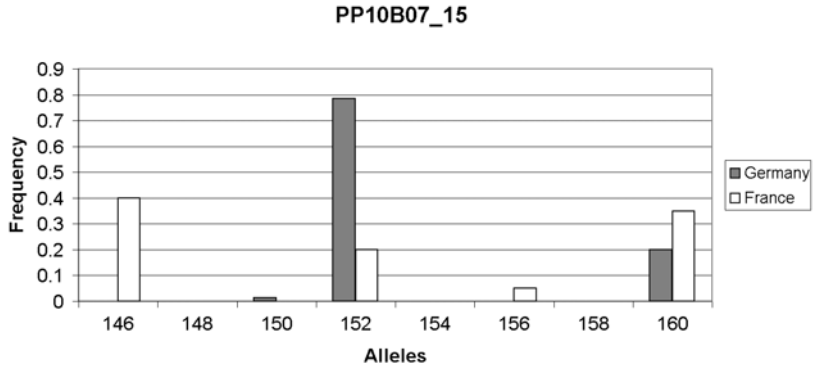
Supplement 4 Primer sequences of the loci genotyped along the 'neutral regions'.

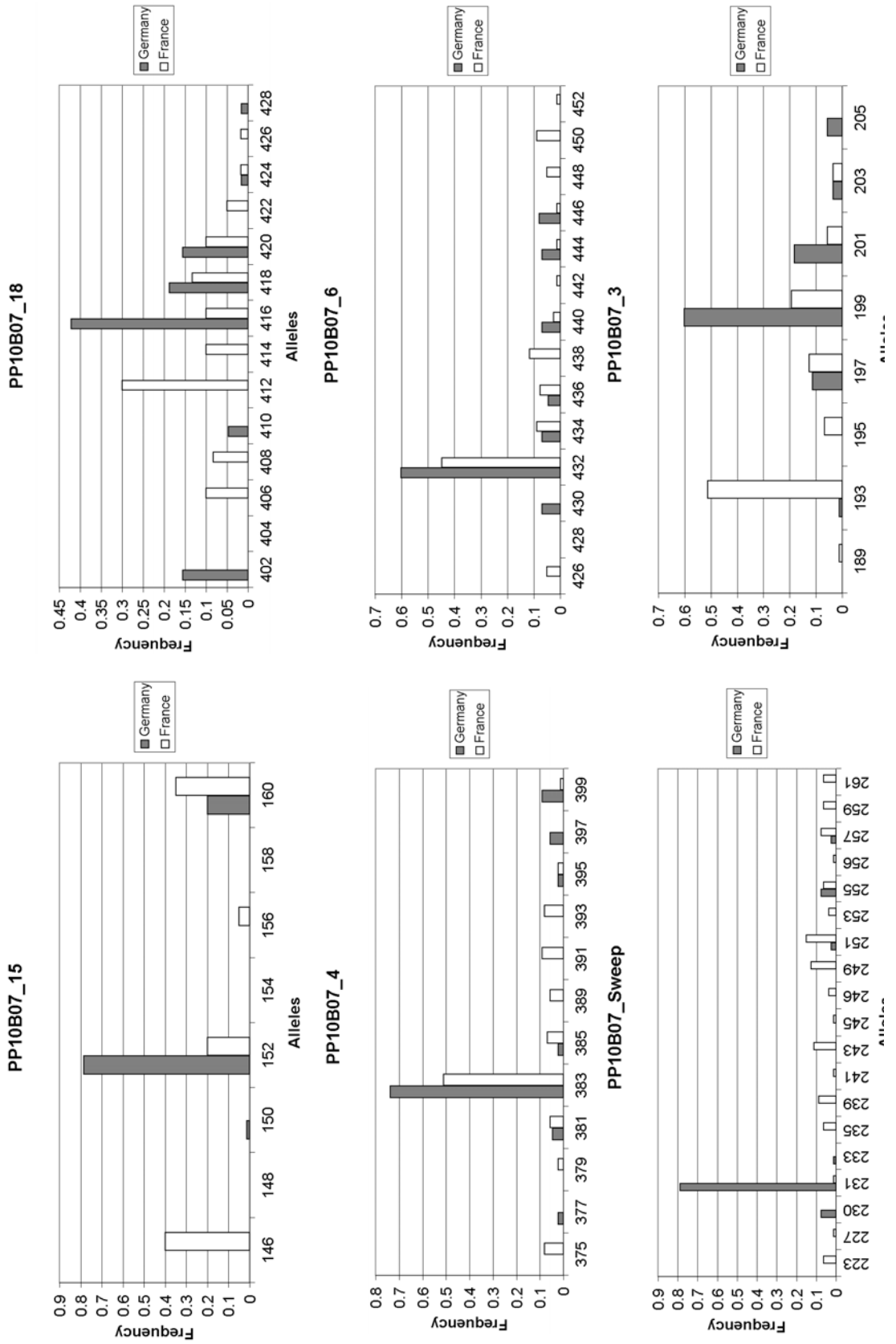
Marker	Chr	Primer's Name	Sequence	Primer's Name	Sequence
N1_01	3	N1_1F	agcaagaattgttcccgaatgc	N1_1R	aggcagttggagcagcttgagg
N1_02	3	N1_2F	gcctgcagccatccaaagctgg	N1_2R	ccagttcagagcaggcatttgg
N1_03	3	N1_3F	tctctgctgggagcacactcag	N1_3R	acccagcacttgggtcttcagg
N1_04	3	N1_4F	accagatcgcaggagatggctc	N1_4R	agagggttgtatgcctcctgag
N1_05	3	N1_5F	tcctgatgtctgccaaccagag	N1_5R	agagcggcagccatgccttcga
N1_07	3	N1_7F	tcagccttggtgagagcacctg	N1_7R	agcaggaagttagctctgtagc
N1_10	3	N1_10F	accacgcacacaaacaagcctc	N1_10R	agggagcactatccttgcagtg
N1_11	3	N1_11F	agggcagctttggctacagacc	N1_11R	tgactgaccaaccagcaaccag
N2_01	6	N2_1F	acccagattgttgagagccacc	N2_1R	tgtctttaggggacggacacac
N2_04	6	N2_4F	agtatgctgcttgatgctcctg	N2_4R	tctcaaggcacacagcacctgc
N2_05	6	N2_5F	aacaagttatgcccacacaagg	N2_5R	tgacttccacacaatcagcctg
N2_11	6	N2_11F	aggtgtgtaccaccatacgtgg	N2_11R	tgcaaaactgtagccgacatgg
N3_01	8	N3_1F	tccagacctcagccttgaccag	N3_1R	tgtgttgagtcaggatctgtgg
N3_02	8	N3_2F	aaccttggtccagccttagagg	N3_2R	tgagagcctattcagggttagc
N3_04	8	N3_4F	gaggactgactggtctagcacc	N3_4R	ggaatgtgtgtgtgcagatgac
N3_08	8	N3_8F	ggagatggctagatgtcgctgc	N3_8R	gctttgggaacagattccggtg
N3_09	8	N3_9F	tgttgccacatacccagcaaag	N3_9R	tcttgacaccggaggttccaca
N3_12	8	N3_12F	tcccactgatgtccaacaaagc	N3_12R	tggttccaaactgacagggctg
N4_01	17	N4_1F	tgtaccttgcctggatgcagag	N4_1R	gagcatggccttatctgacagc
N4_02	17	N4_2F	gtatgcagtggaaacacctgct	N4_2R	gaagcaagccttcctaagctgc
N4_06	17	N4_6F	ggttctgatagaggcatgttgtc	N4_6R	tgtcccaagttcagacacctgg
N4_07	17	N4_7F	ctataaagttgctgttgcccga	N4_7R	agtcaccagtcagatcaggcag
N4_08	17	N4_8F	tgggatagggcgactccagagg	N4_8R	tgcaaacccaatagacacctcc
N4_09	17	N4_9F	tgtattctatcccatgcccgag	N4_9R	gtgaacacacgcgcacacatgc
N4_13	17	N4_13F	actaggggacttgtacccatgc	N4_13R	tccaagctcacagttgcctgag
N4_15	17	N4_15F	aagctgctgattaccaccaaag	N4_15R	ctagggtcttctaccacaatgg
N4_18	17	N4_18F	tcaactgcctcaggggttgtcc	N4_18R	agctattgtagcccaactggct
N5_02	4	N5_2F	acgctgacaccattgcgtacac	N5_2R	catcttgtggagagagcggttc
N5_03	4	N5_3F	ttaatcttgcccatccatgagc	N5_3R	accttcctctacacacacagac
N5_04	4	N5_4F	tttgggtggaaagtccatgagc	N5_4R	aagacagctttctgcttgagac
N5_07	4	N5_7F	cccatcactggaggtctcatcg	N5_7R	gggcctagatgtcctgagagct
N6_03	12	N6_3F	accagcagcccaatcacagcag	N6_3R	taattccaagcagtggtactgg
N6_05	12	N6_5F	tgtaccaggatatcatgcaggt	N6_5R	tagagccttcaccacaacatgc
N6_08	12	N6_8F	ggtttctttgtgtccgtgaagg	N6_8R	caggtgtcagtgttcttgcctg
N6_09	12	N6_9F	tcatgcaaacacagccattgtc	N6_9R	cctaatccttctccaggagcac
N6_10	12	N6_10F	aggtgtccattaatgccctgtc	N6_10R	gggcatgtaataccttgaggac
N8_01	7	n8_1F	tccaagactttgtgggtgccag	N8_1R	agcaccatgtctgcctgaatgc
N8_02	7	n8_2F	accttaaggcatgcaaaccctg	n8_2R	tggtactgaccaaggccctttg
N8_03	7	n8_3F	agatgcacctaatcctccagag	n8_3R	agcctgcagaaagcagttctcc
N8_09	7	n8_9F	acatagatggttgtgatggtcc	n8_9R	ttcaggaccttggtcaggagtc
N8_13	7	n8_13F	atctggctttatgttgagggct	n8_13R	agtcctcagacctaggacccag
N8_19	7	n8_19F	tgaagaaacaagctggagtagc	n8_19R	ttaaaggcgtgtgccaccatgc
N9_05	11	n9_5F	acagcactcccgtggactcagg	n9_5R	agttagttcaggacatgcagtc
N9_08	11	n9_8F	ttcttgcgtgcctgctgaggac	n9_8R	aactggccacatggcctttacc
N9_13	11	n9_13F	tcaatgccattgtccagaggtc	n9_13R	aacttcggtgtttgctaaggca
N9_14	11	n9_14F	tccaagttcagatggacacacc	n9_14R	tctttgtgcatcagttccctgg
N9_16	11	n9_16F	tcctcctctgtgagctgcactc	n9_16R	gttgtttgcagatgtgtgtgtg
N9_17	11	n9_17F	ttcagtgtgcaagagtgcttgc	n9_17R	ggataggctcaacccatccagg

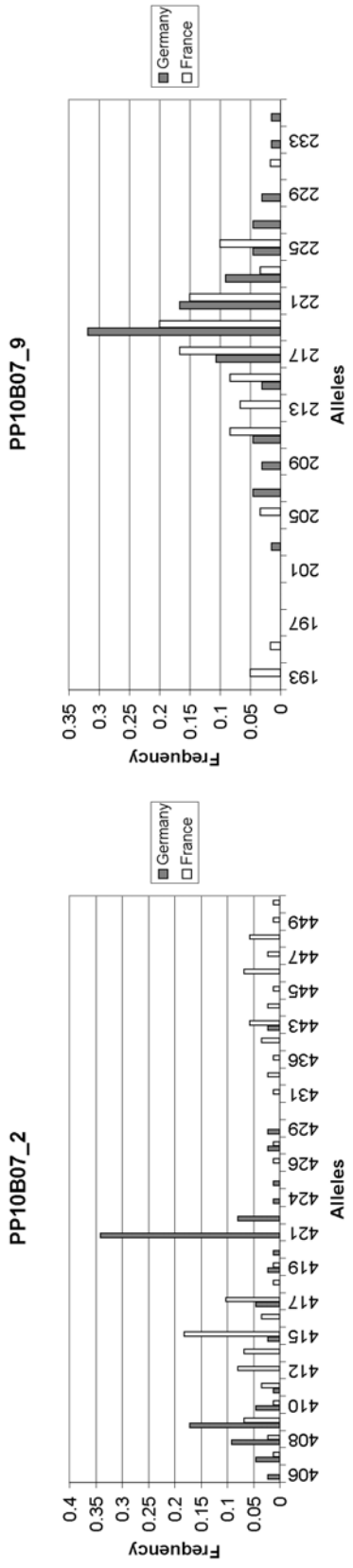
Supplement 5 Allele frequencies of the loci genotyped in candidate region PP10B07 (German sweep).

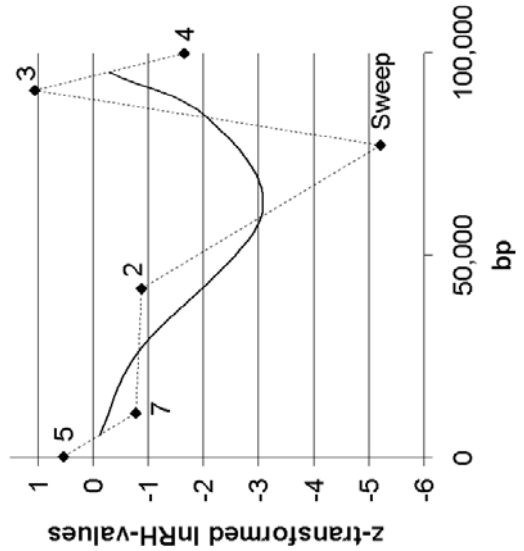




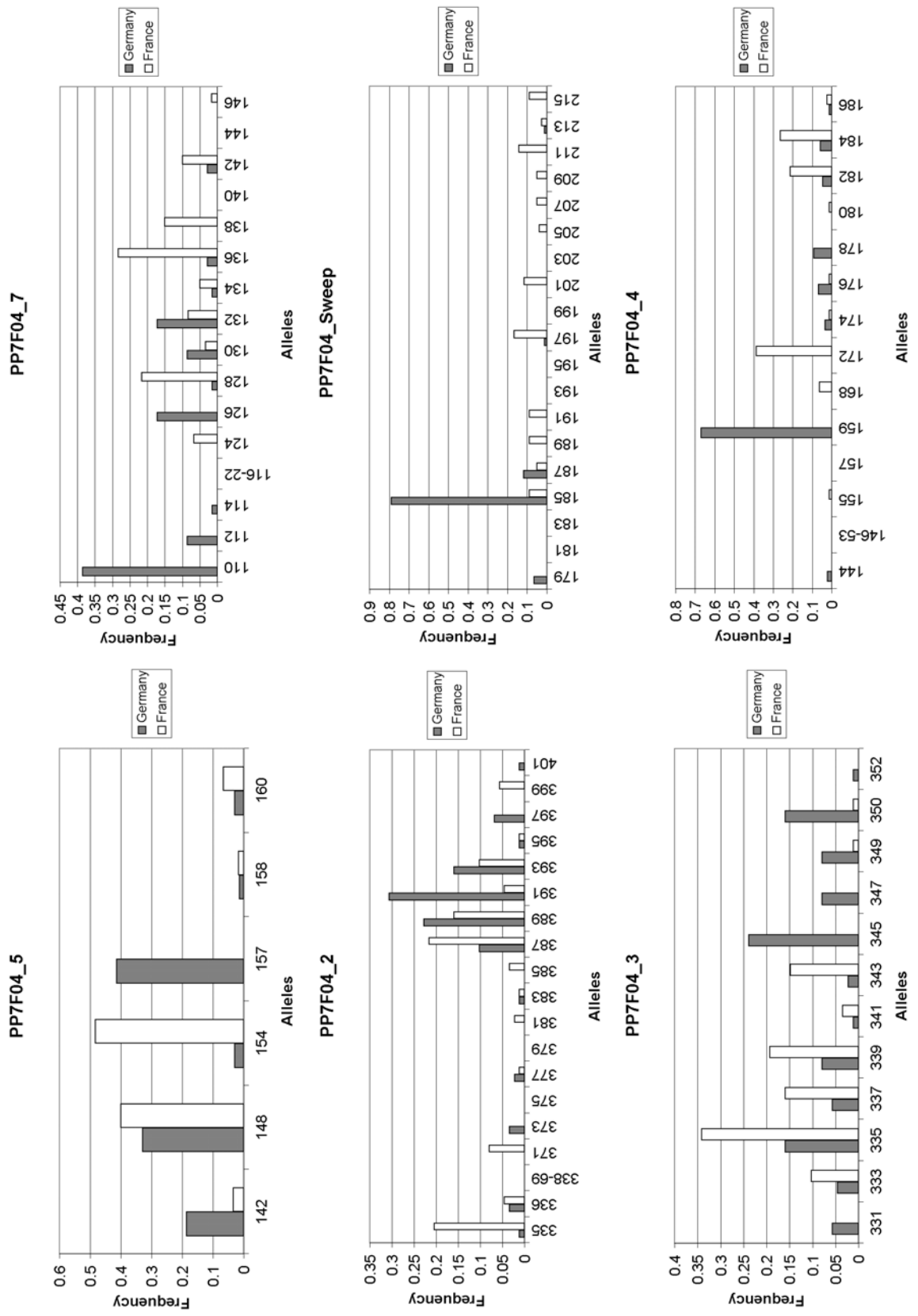



PP10B07_18
PP10B07_6
PP10B07_3
PP10B07_15
PP10B07_4
PP10B07_Sweep
Germany
France
Frequency
Alleles

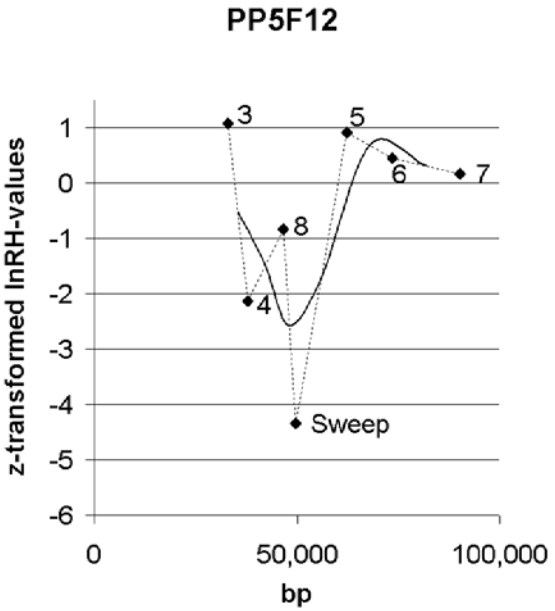


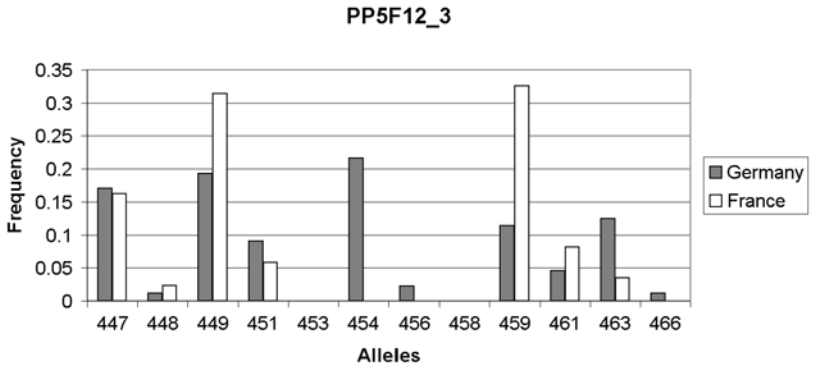


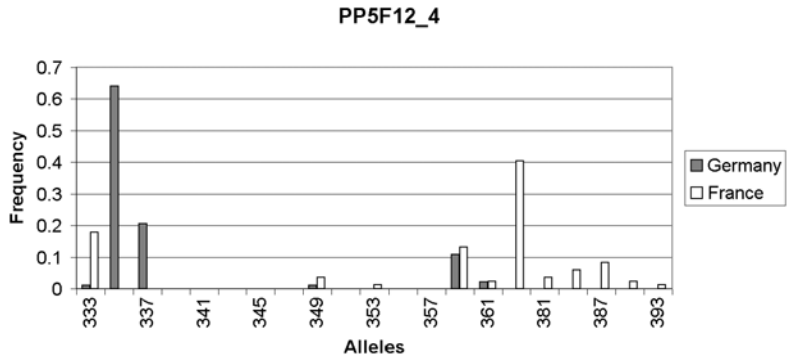
Supplement 6 Allele frequencies of the loci genotyped in candidate region PP7F04 (German sweep).

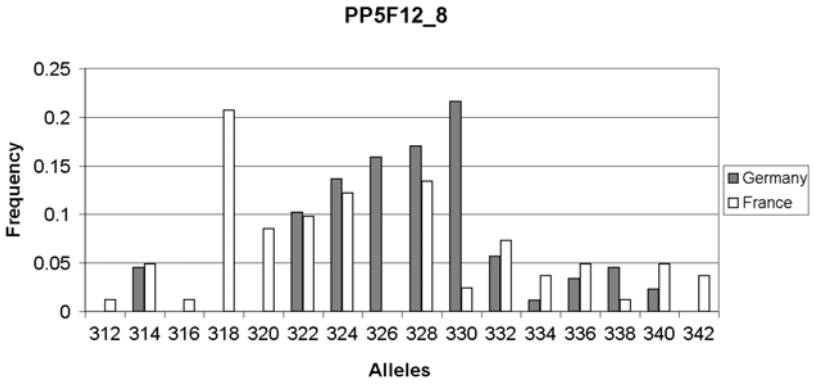

PP7F04_7
PP7F04_Sweep
PP7F04_4
PP7F04_5
PP7F04_2
PP7F04_3
Germany
France
Frequency
Alleles

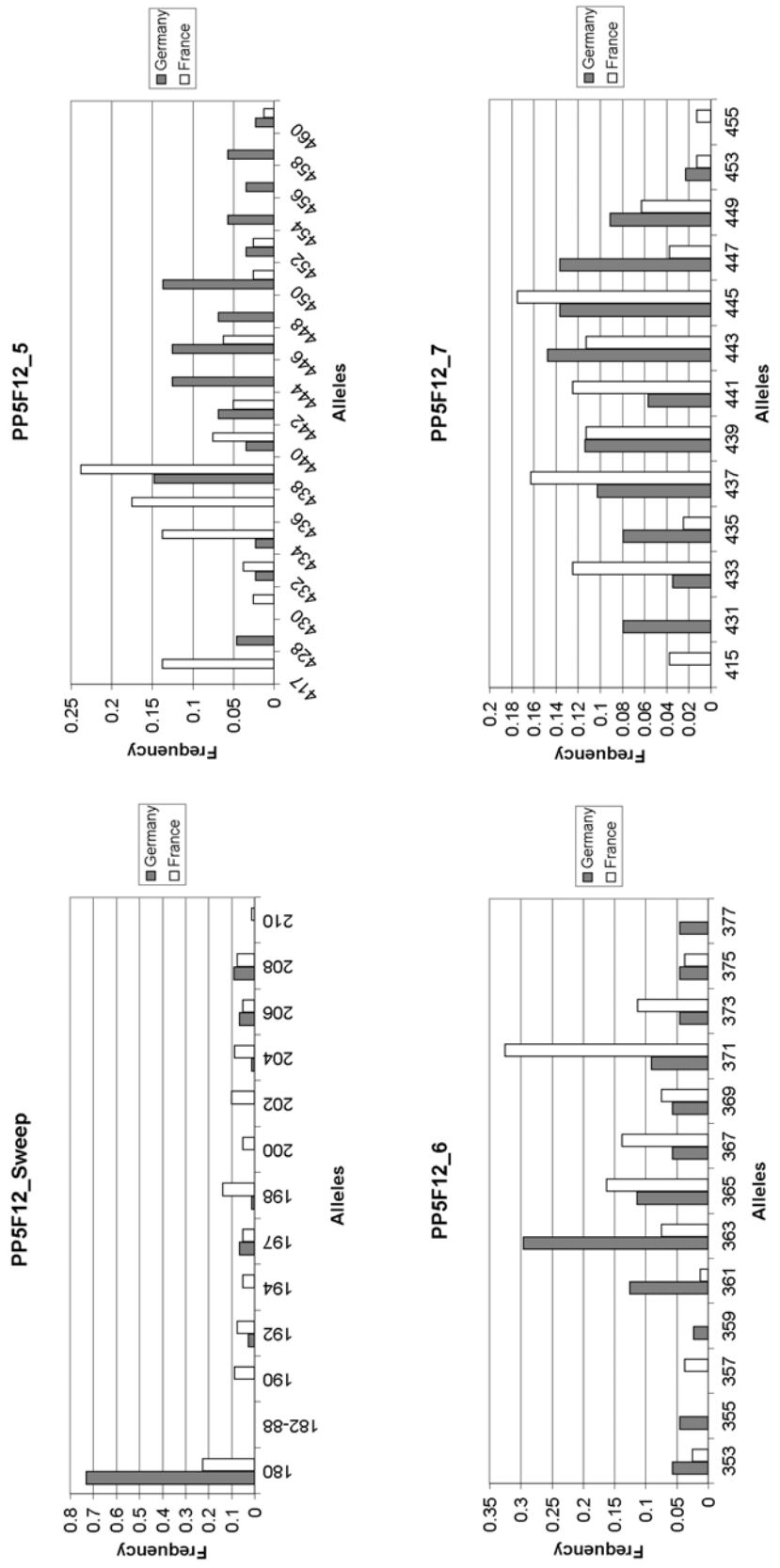
Supplement 7 Allele frequencies of the loci genotyped in candidate region PP5F12 (German sweep).

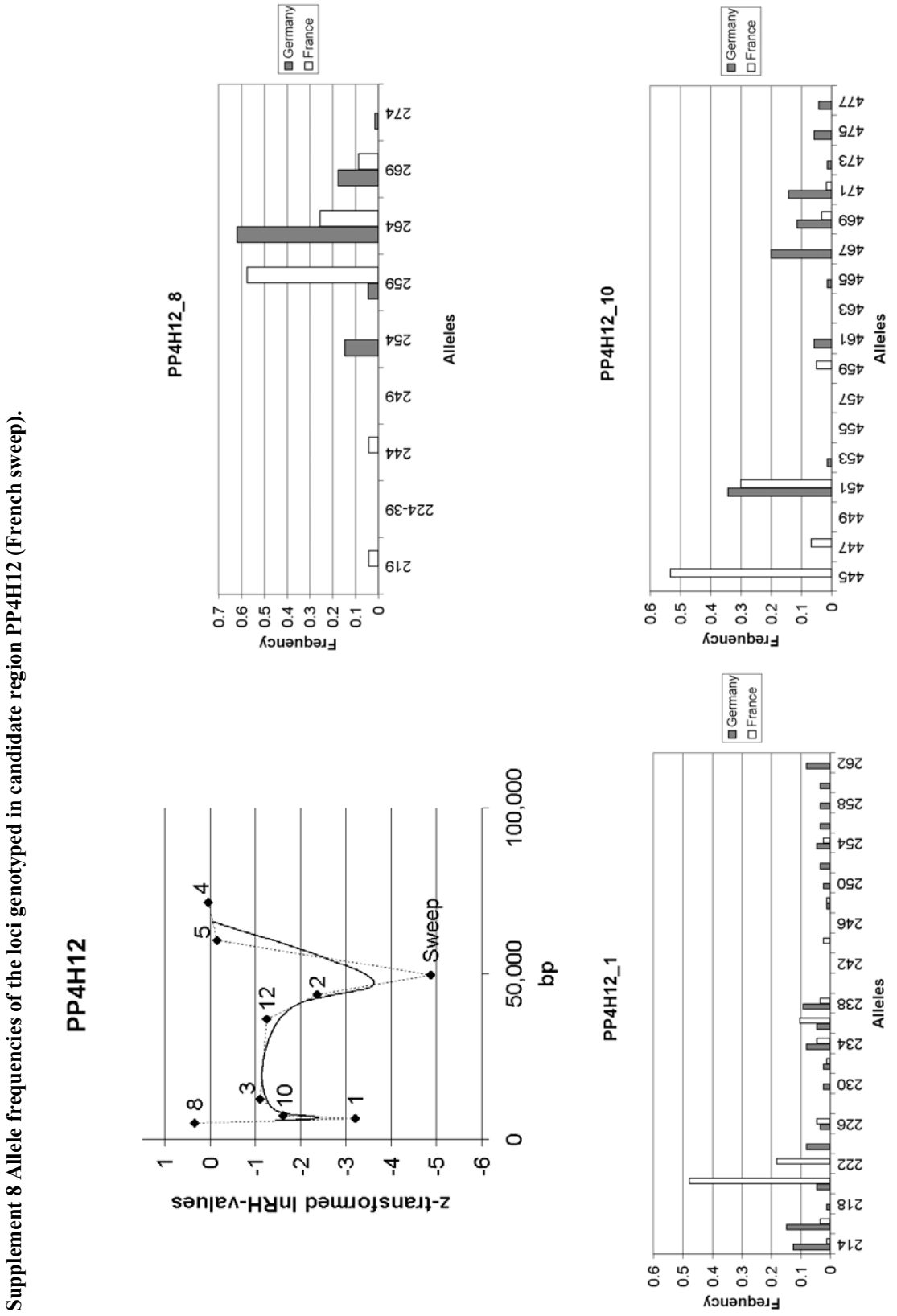




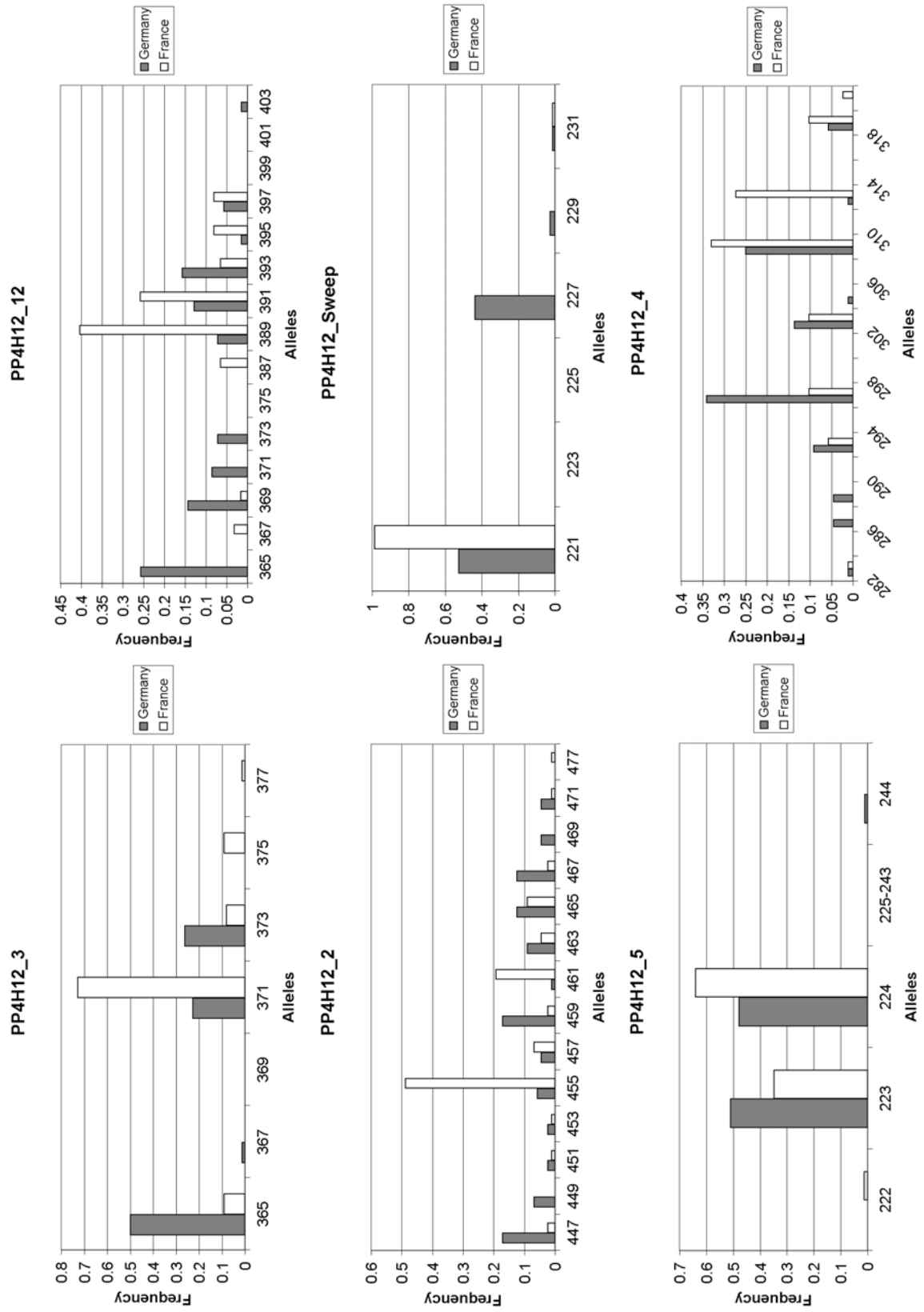




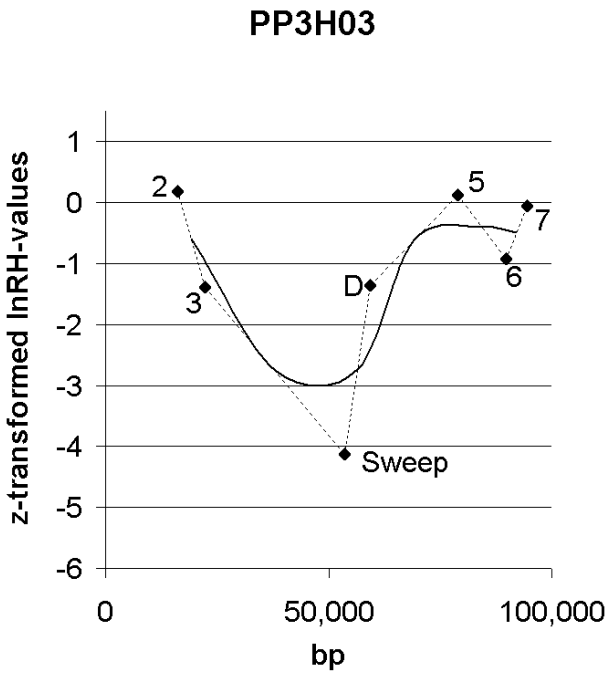

PP5F12_5
Germany
France
Frequency
Alleles
PP5F12_7
Germany
France
Frequency
Alleles
PP5F12_Sweep
Germany
France
Frequency
Alleles
PP5F12_6
Germany
France
Frequency
Alleles

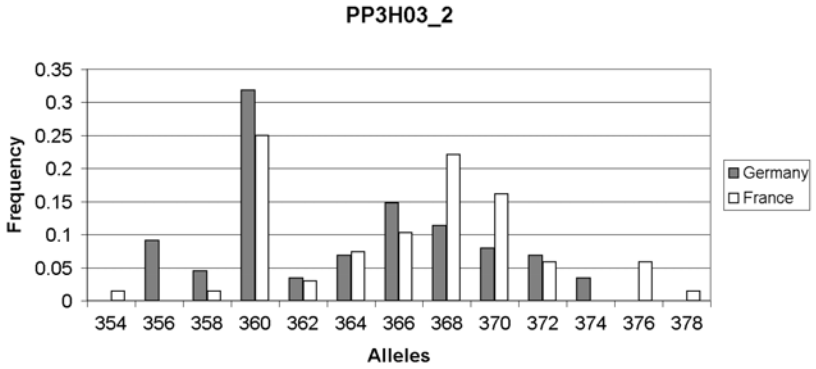


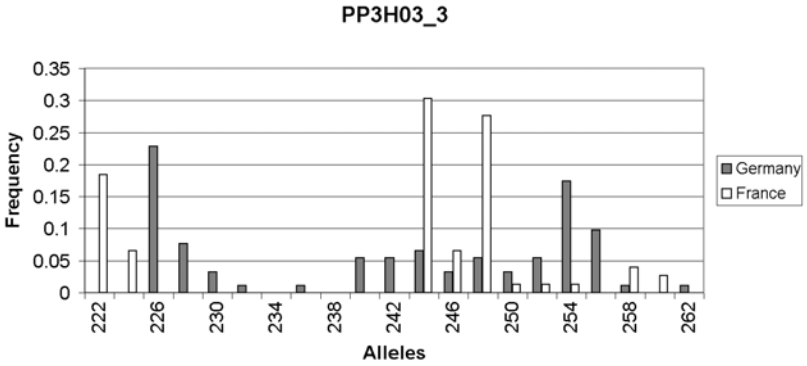
Supplement 8 Allele frequencies of the loci genotyped in candidate region PP4H12 (French sweep).

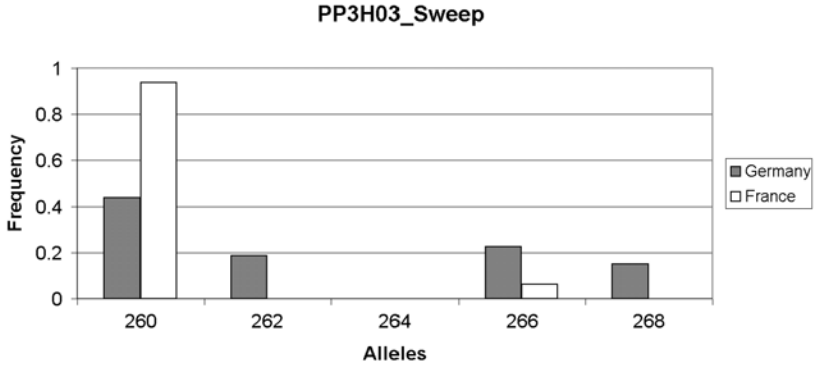

PP4H12_12
PP4H12_Sweep
PP4H12_4
PP4H12_3
PP4H12_2
PP4H12_5
Germany
France
Frequency
Alleles

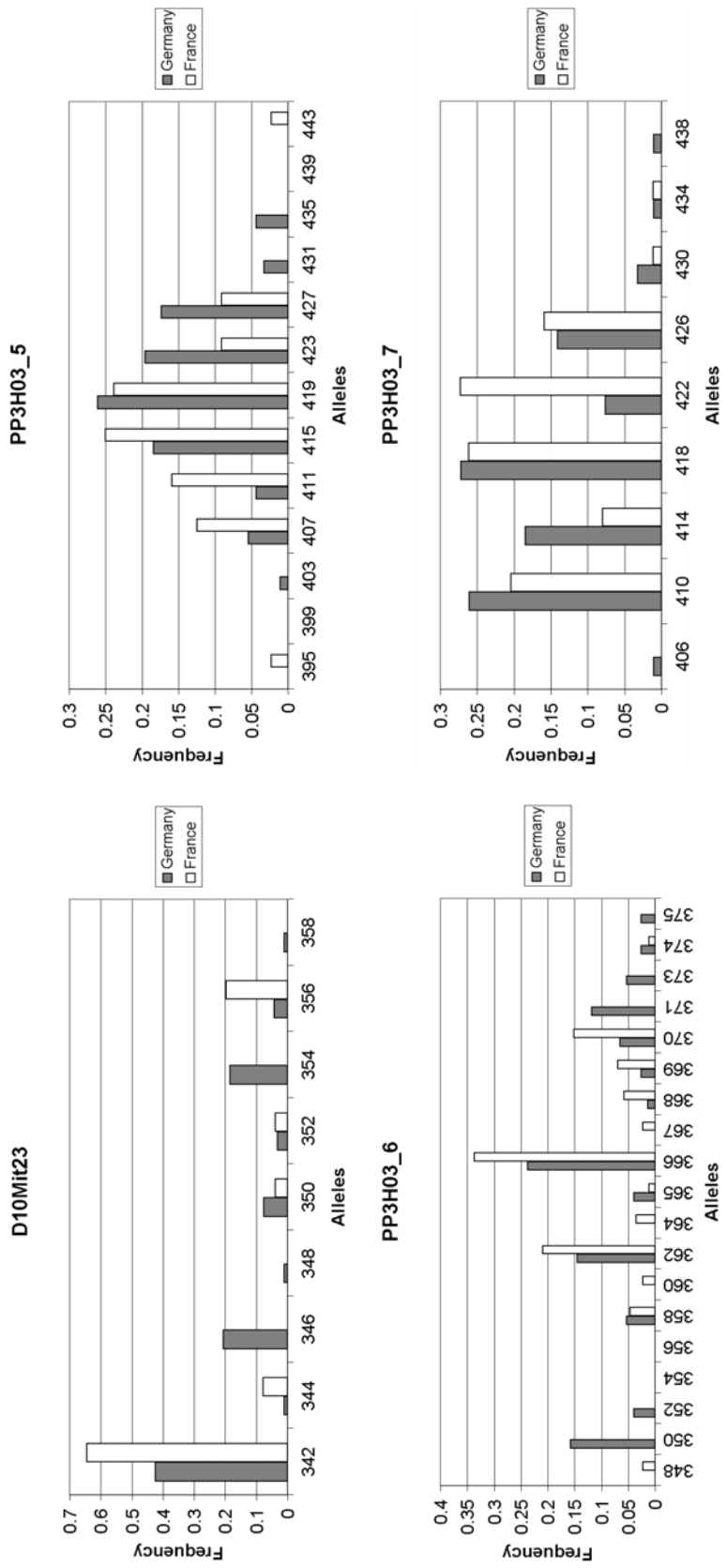
Supplement 9 Allele frequencies of the loci genotyped in candidate region PP3H03 (French sweep).



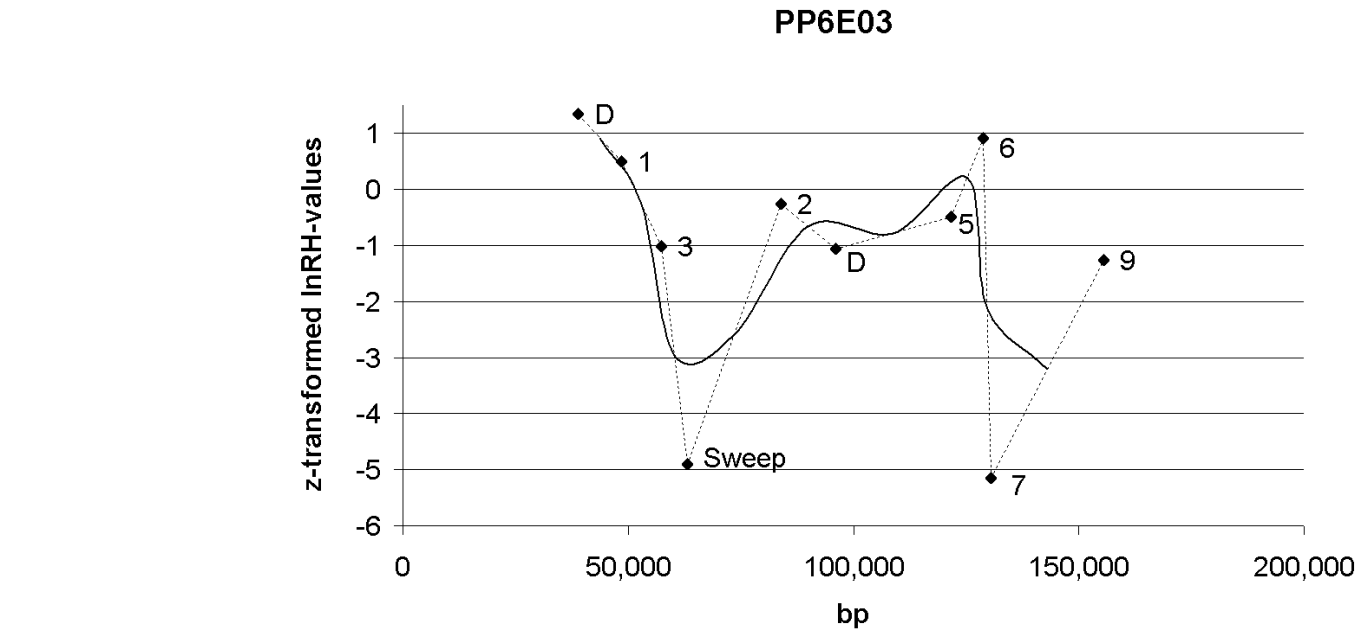


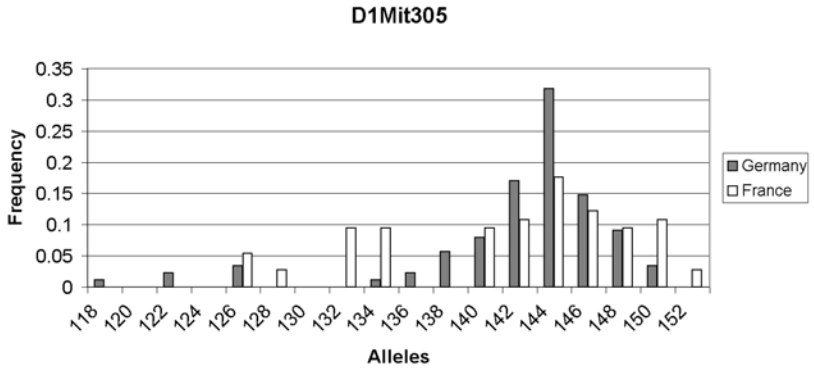


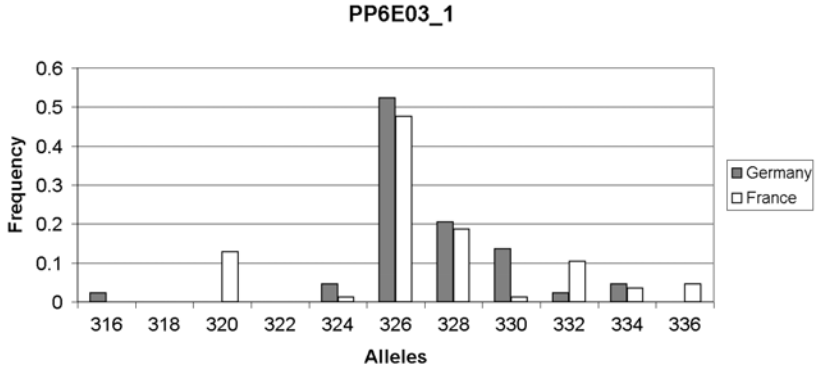


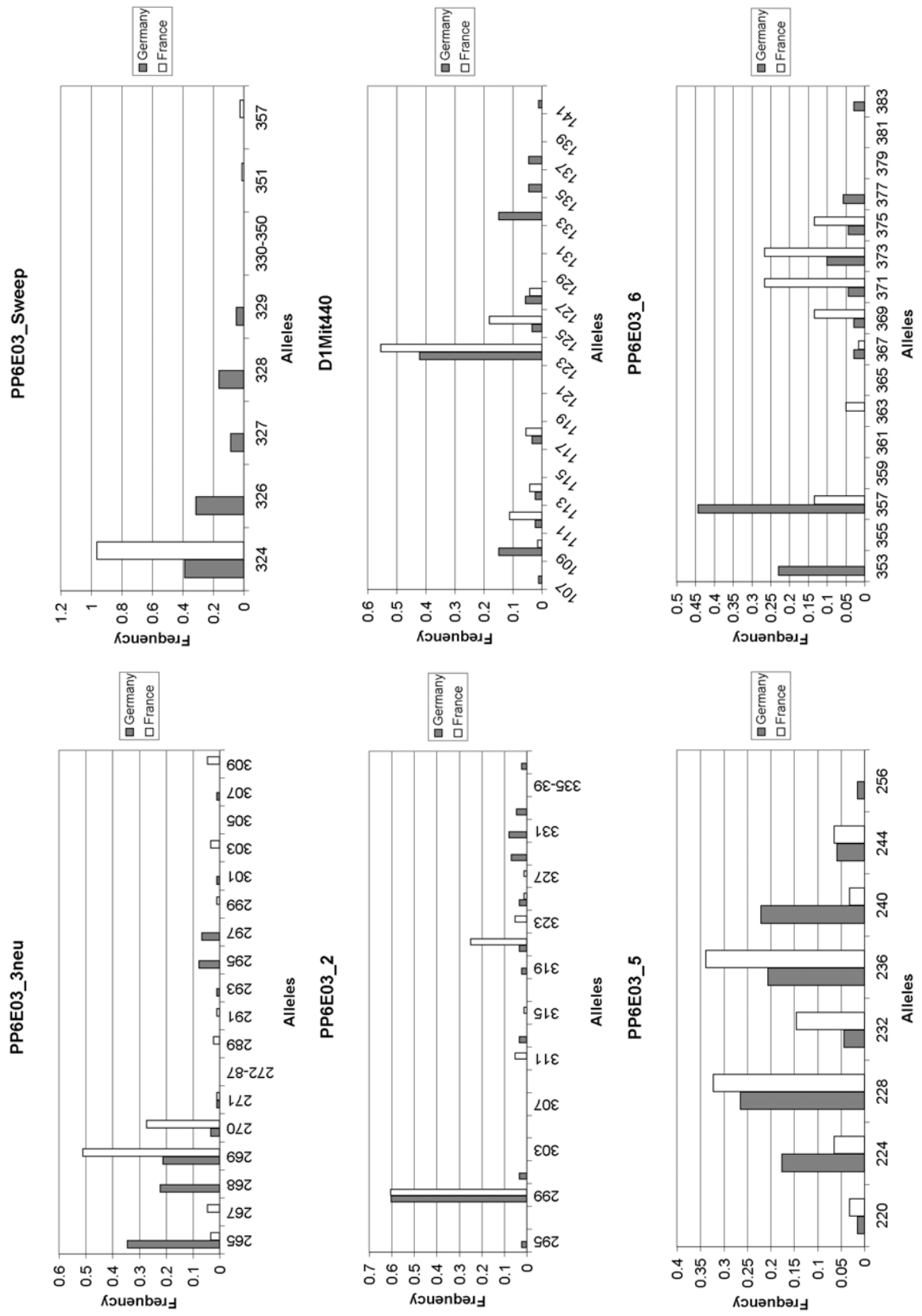

PP3H03_5
Germany
France
Frequency
Alleles
PP3H03_7
Germany
France
Frequency
Alleles
D10Mit23
Germany
France
Frequency
Alleles
PP3H03_6
Germany
France
Frequency
Alleles

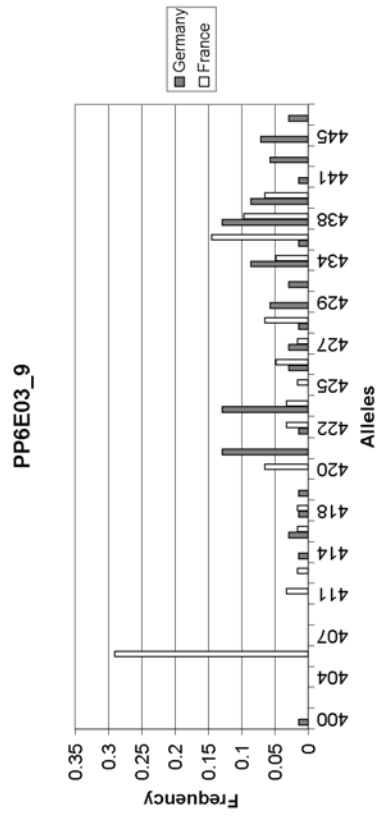
Supplement 10 Allele frequencies of the loci genotyped in candidate region PP6E03 (French sweep).

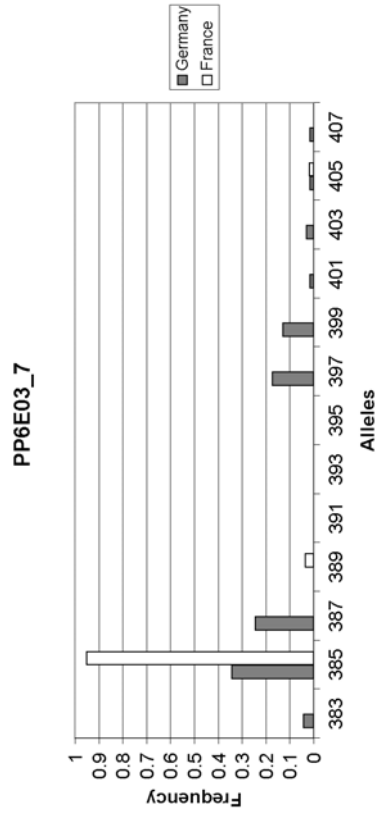







PP6E03_Sweep
D1Mit440
PP6E03_6
PP6E03_3neu
PP6E03_2
PP6E03_5
Germany
France
Frequency
Alleles


PP6E03_9
Germany
France
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Alleles
400
404
407
411
414
418
420
422
425
427
429
434
438
441
445


PP6E03_7
Germany
France
Frequency
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Alleles
383
385
387
389
391
393
395
397
399
401
403
405
407

6 Digital Supplement

- PDF files of the pool pattern of all 960 microsatellite loci
- Row data of pool patterns
- PDF of a table containing the sequences of all amplified fragments

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 Herrn Prof. Dr. Diethard Tautz betreut worden.

Köln, den 29.03.2006

Teilpublikationen:

Die folgende Publikation basiert auf Teilen dieser Arbeit und entspricht Kapitel 2:

Thomas, M, Ihle, S, Ravaoarimanana, Y, Kraechter, S, Wiehe, T, Tautz, D. 2005. Microsatellite variability in wild populations of the house mouse is not influenced by differences in chromosomal recombination rates. Biological Journal of the Linnean Society **84**: 629-635.

Lebenslauf

Name:	Meike Thomas
Anschrift:	Michael-Leveilly-Str. 6 53127 Bonn
Geburtsdaten:	13.09.1976 in Traben-Trarbach
Staatsangehörigkeit:	deutsch

Schulausbildung:

1983 – 1987	Konrad-Adenauer Grundschule in Emmelshausen
1987 – 1996	Staatliches Kant Gymnasium in Boppard
1996	Abitur

Hochschulausbildung:

WS 1996/97 – SS 2002	Studium der Biologie Rheinische Friedrich-Wilhelms-Universität Bonn
26. Juni 2002	Abschluss des Diploms in Biologie Forschungsmuseum Alexander Koenig Abteilung für Herpetologie Prof. Dr. Wolfgang Böhme Rheinische Friedrich-Wilhelms-Universität Bonn
April 2003 – Juni 2006	Promotion am Institut für Genetik Lehrstuhl für Evolutionsgenetik Prof. Dr. Diethard Tautz, Universität zu Köln
01. Juni 2006	Vorraussichtlicher Abschluss der Promotion