Functional characterization of adaptively relevant genes in the house mouse (*Mus musculus* L.)

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Natascha Hasenkamp

Kiel, 2014

Erster Gutachter: Prof. Dr. Diethard Tautz

Zweiter Gutachter: Prof. Dr. Hinrich Schulenburg

Tag der mündlichen Prüfung: 02.07.2014

Zum Druck genehmigt: 02.07.2014

gez. Prof. Dr. Wolfgang Duschl, Dekan

Table of contents

List of figures III

List of tables V

Danksagung VII

Zusammenfassung IX

Summary XI

Declaration XIII

General introduction 1

A brief introduction to molecular evolution 1

Identification and analysis of adaptively relevant genes 4

The wild mouse (*Mus musculus* L.) as a model system 9

Scope of the thesis 12

Chapter 1 - Natural variation in the cell-surface receptor *Xpr1* in wild populations of the house mouse 15

Chapter 2 - Characterization of functional allelic variation in *Dmrt1* in wild populations of the house mouse 33

Bibliography 57

Eidesstattliche Erklärung 69

Appendix 71

S1 (General): 71

S2 (Chapter 1): 72

S3 (Chapter 2): 102

List of figures

1.1	Genetic signatures of natural selection……....	4
1.2	Phylogeny of the genus *Mus*…….	10
1.3	Today's geographical distribution of the genus *Mus*……..	11
2.1	Overview map of analyzed wild mouse populations……	18
2.2	Genomic overview of *Xpr1*……	21
2.3	Map of the haplotype frequencies of *Xpr1* in wild populations……	24
2.4	Haplotype network of *Xpr1* exon sequences……..	25
2.5	Bayesian tree of P-MLV sequences from wild-derived mice……...	26
3.1	Overview of the genomic organization of *Dmrt1*……	37
3.2	Phylogenetic relationships of analyzed species and their *Dmrt1* variants…….	41
3.3	Frequencies of *Dmrt1* genotypes in wild populations……...	42
3.4	Comparison of difference measures for the microarray data analysis...	43
3.5	Venn-Diagrams showing the overlaps of detected genes across tested postnatal stages……..	47
S2.1	Allele frequency distribution of microsatellite loci around *Xpr1*……...	89
S2.2	Amino acid sequence alignment of P-MLVs from wild-derived mice….	101
S3.1	Microsatellite allele frequency distribution of all loci in the genomic region of *Dmrt1*…….	102

List of tables

2.1	Identified *Xpr1* haplotypes.....	23
2.2	Comparison of SNPs in the extracellular loops of *Xpr1*…….	24
3.1	Overview of test results from microarray analysis…….	45
3.2	Enriched GO terms in the most differently transcribed genes between mice carrying the Dmrt1-N or –S allele……...	46
S1	Origin of wild-caught mice…….	71
S2.2	Primers for microsatellite loci in the genome region of *Xpr1*……...	72
S2.3	PCR and sequencing primers for *Xpr1*…….	73
S2.4	Overview of SNPs in the *Xpr1* haplotypes……..	74
S2.5	Overview of SNPs in the P-MLV provirus receptor-binding domain…..	87
S3.1	PCR and sequencing primers for *Dmrt1*…….	104
S3.2	Primers for microsatellite loci in the genome region of *Dmrt1*…….	104
S3.3	Transcript levels of *Dmrt1* for different alleles and stages……...	104
S3.4	List of the most differently transcribed genes between mice carrying the *Dmrt1*-N or –S allele at different postnatal stages……..	105
S3.5	Overview of shared genes across analyzed stages in brain and testis.	109

Danksagung

Mein Dank gilt all jenen, die mich im Laufe meiner Dissertation unterstützt haben und es so ermöglichten, dass diese Arbeit in ihrer heutigen Form zustande gekommen ist.

Vor allem möchte ich mich bei Herrn Prof. Dr. Diethard Tautz dafür bedanken, dass er es mir ermöglichte meine Promotion in seiner Arbeitsgruppe anzufertigen. Durch seine Diskussionsbereitschaft und Ideen hat er entscheidend zum Gelingen des Projektes beigetragen und ich habe in der Zeit in seinem Labor viel lernen können. Auch den beiden anderen Mitgliedern meines IMPRS-Thesis-Kommittees, Prof. Dr. John Baines und Prof. Dr. Matthias Leippe, möchte ich für ihre Vorschläge und Diskussionsbeiträge danken.

Dr. Meike Teschke möchte ich dafür danken, dass sie mich dazu ermutigt hat den Schritt von der Naturschutzbiologie in die Evolutionsgenetik zu machen und mich vor allem in der Anfangszeit sehr unterstützt hat.

Auch möchte ich mich ganz herzlich bei allen dauerhaften und zeitweisen Mitgliedern der Abteilung Evolutionsgenetik bedanken, die unterschiedlichste Beiträge zu der hier vorliegenden Arbeit geleistet haben.

Terry Solomon hat mit ihrer großen Begeisterung, Offenheit und Lernbereitschaft das *Xpr1*-Projekt wesentlich geprägt und es hat mir viel Freude gemacht mit ihr zu Arbeiten. Auch Heinke Buhtz, Cornelia Burghardt, Dr. Emilie Hardouin und Dr. Miriam Linnenbrink danke ich herzlich für die lustige, abwechslungsreiche Zeit im Labor und die fortwährende moralische Unterstützung. Ohne Connies und Heinkes Erfahrungsschatz, Organisationsgeschick, Sorgfalt und Humor, mit denen sie stets zur Stelle waren, wäre vieles nur schwer möglich und bei weitem nicht so schön gewesen. Emilie und Miri hatten stets ein offenes Ohr für Fragen und Sorgen und zwei offene Augen beim Korrekturlesen, sodass wir manches Problem über einem Kaffee lösen konnten. Außerdem möchte ich mich bei Emilie für die zahlreichen, unmöglichen Ohrwürmer bedanken, die gemeinsame Arbeiten im Labor so mit sich brachten. Sunna Ellendt möchte ich, neben dem Korrekturlesen, auch für die großartige moralische Unterstützung danken. Bei Dr. Angelika Börsch-Haubold und Dr. Pedrag Kalajdzic möchte ich mich für ihre nette Gesellschaft im Büro und die vielen hilfreichen Gespräche bedanken. Vor allem Angelika hat mir mit ihren Ideen und Anregungen in Bezug auf Labormethoden und Datenauswertung sehr geholfen.

Darüberhinaus möchte ich Elke Blohm-Sievers und Nicole Thomsen für die hervorragende Arbeit mit den Microarrays danken. Auch Dr. Alexander Pozhitkov, Dr. Till Czypionka und Rafik Neme danke ich in diesem Zusammenhang für die Beantwortung unzähliger Fragen rund um das Thema Microarraydesign und –datenanalyse. Proben für diese Analysen konnte ich dank Christine Pfeifle und Heike Harre sammeln, die mich bei allen mausigen Angelegenheiten sehr unterstützt haben, wie etwa Zuchtplanung und der praktischen Einführung in den Umgang mit sprunghaften Mäusen. Bei Till Sckerl bedanke ich mich für die sehr zuverlässige Betreuung der Mauszuchten und Mithilfe bei den ersten Gehversuchen auf dem Gebiet der ultraschallbasierten Schwangerschaftsdiagnostik bei Wildmäusen.

Zu guter Letzt, danke ich meiner Familie und Freunden, allen vorran Volker, Tante Inge, MELA, JUMT, Nicole und David, für den ständigen Rückhalt und die Unterstützung in allen Lebenslagen. Ohne sie wären mein Studium und diese Dissertation kaum möglich gewesen. Bei David möchte ich mich besonders bedanken, weil er, neben dem Schreiben sehr hilfreicher Skripte für die Arrayanalyse, auch die alltäglichen Höhen und Tiefen einer Doktorarbeit geduldig mitgemacht hat (ich weise allerdings nach wie vor jede Schuld von mir was den Bart betrifft).

Zusammenfassung

Ein zentrales Ziel der Evolutionsbiologie ist es, die molekulare Basis von Anpassungsprozessen zu verstehen und zu charakterisieren. Eine Voraussetzung für diese Prozesse ist, dass vererbbare Variation in Form von Allelen mit funktionalem Effekt auf den langfristigen Reproduktionserfolg von Individuen vorhanden ist. Daher wurden in den letzten Jahren viele Studien durchgeführt um Gene zu identifizieren, die unter positiver Selektion stehen und damit aktiv in Anpassungsprozesse involviert sind. Durch diese umfassenden Datensätze wurden viele Kandidatengene identifiziert, doch im Verhältnis dazu gibt es wenig Informationen über ihre allelische Variation und Funktion. In der vorliegenden Studie, habe ich Daten genutzt, die auf genomweiten Analysen zu Selektionsereignissen in vier wilden Hausmauspopulationen (*Mus musculus* L.) beruhen. Zwei Kandidatengene namens *Xpr1* und *Dmrt1* wurden aufgrund ihrer gut beschriebenen, generellen Funktion ausgewählt. Sie wurden hier hinsichtlich des Vorkommens von potenziell funktionalen Allelen, deren Variation, der Verbreitung eventuell assoziierter positiver Selektionsereignisse und funktionaler Effekte untersucht.

Beide Gene wurden in 11 Populationen und mehreren Außengruppen vor allem im Hinblick auf Variation in der proteinkodierenden Sequenz, als mögliches Ziel positiver Selektion, analysiert. Die untersuchten *M. m. domesticus*-Populationen stammten aus Deutschland, Frankreich und dem Iran, während die *M. m. musculus*-Populationen aus der Tschechischen Republik und Kasachstan stammten. Zusätzlich zur Sequenzanalyse wurden Mikrosatelliten in einem Bereich von 200 kb um beide Gene typisiert, um eine mögliche Ausbreitung des positiven Selektionsereignisses über mehrere Populationen feststellen zu können. Die Analyse der potenziellen funktionalen Effekte der Allele war abhängig von der jeweilig beschrieben Genfunktion. *Xpr1* ist ein Zelloberflächenrezeptor, der insbesondere für seine Rolle bei der Infektion von Zellen durch bestimmte Mausleukämieviren (X/P-MLVs) bekannt ist, die Leukämie und Lymphome verursachen können. Der Transkriptionsfaktor *Dmrt1* ist in Vertebraten hoch konserviert und bei Mäusen wichtig für die Geschlechtsdifferenzierung.

Beide Gene zeigten Variation in der kodierenden Sequenz. Für *Xpr1* konnten 14 Allele nachgewiesen werden, die in Form von Haplotypen auftraten. Zwei dieser Allele zeigten Signaturen positiver Selektion und ein Introgressionsereignis eines dieser Haplotypen vom Iran nach Südfrankreich wurde detektiert. Außerdem fanden

sich Signaturen balancierender Selektion auf zwei Haplotypen in drei anderen Populationen aus Frankreich. Die Analyse von möglichen allelspezifischen Funktionsunterschieden konzentrierte sich hier auf eine potenzielle Assoziation zwischen Rezeptorallelen und Variation in transkribierten P-MLVs in gezüchteten Wildmäusen. Es konnten jedoch weder eine solche Assoziation gefunden werden, noch infektiöse Virenpartikel isoliert werden. Zusammen mit Daten aus früheren Studien lassen sich diese Ergebnisse so interpretieren, dass die *Xpr1*-Variation in natürlichen Hausmauspopulationen wahrscheinlich von andauernden adaptiven Prozessen geprägt ist. Es scheinen allerdings auch andere Faktoren als nur durch MLV induzierte Krankheiten diese Prozesse anzutreiben.

In *Dmrt1* hingegen wurde ein einzelner Aminosäureaustausch von Asparagin (N) und Serin (S) gefunden. Daten aus anderen Studien legten nahe, dass dieses Gen in den hier untersuchten Wildmauspopulationen weder alternativ gespleißt noch differenziell exprimiert wird. Das N stellt wahrscheinlich die abgeleitete Form des Allels dar und verursacht eine neue potenzielle Glykosylierungsstelle. Außerdem wurde eine mögliche Signatur positiver Selektion in fünf Populationen in Frankreich identifiziert. Die Analyse von allelspezifischen, funktionalen Effekten konzentrierte sich auf mögliche Unterschiede in der Regulation von Zielgenen. Eine vergleichende Microarray-basierte Untersuchung von Mäusen früher postnataler Altersstadien mit N- und S-Genotypen deutete an, dass *Dmrt1*-Zielgene vermehrt differenziell transkribiert wurden, wenn *Dmrt1* anwesend war. Dies, zusammen mit den populationsgenetischen Daten, weist darauf hin, dass es anscheinend einen funktionalen Unterschied zwischen dem N- und S-Allel gibt, der möglicherweise unter positiver Selektion steht. Allerdings werden weitere funktionale Analysen notwendig sein, um den eigentlichen vorteilhaften Effekt des N-Allels zu ergründen.

Summary

One of the main themes in evolutionary biology is the characterization of the molecular basis of adaptation. A prerequisite for adaptation is the presence of heritable variation in the form of functionally distinct alleles which have differential effects on the lifetime reproductive success of an organism. That is why a lot of effort has been made in recent years to identify genes that are under positive selection and thus involved in adaptive processes in natural populations of different species. Through comprehensive genetic screens a large number of such candidate genes have been identified. However, knowledge about the allelic variation, its molecular function and phenotypic effects remains still rare. In this study, I took advantage of data about candidate genes for selective sweeps from previous hitchhiking mapping studies in four European wild mouse (*Mus musculus* L.) populations and of the extensive knowledge about mice through years of lab mouse research. Two candidate genes, namely *Xpr1* and *Dmrt1*, with a generally well characterized function were analyzed to gain information about potential functional alleles, their variation and extent of the selective sweep across different populations and possible functional effects of the different alleles. *Xpr1* is a cell-surface receptor which is especially known for its function in infection of cells by polytropic and xenotropic murine leukemia viruses (X/P-MLVs), which can cause leukemia and lymphomas in mice. The transcription factor *Dmrt1* is a highly conserved member of the sex determination cascade in vertebrates and required for normal gonadal differentiation in mice.

For both genes, I focused on coding sequence variation as potential target of positive selection in 11 wild mouse populations and several outgroups. Analyzed *M. m. domesticus* populations came from France, Germany and Iran, while the *M. m. musulus* populations came from Czech Republic and Kazakhstan. Additionally, microsatellites in a region of 200 kb around both loci were analyzed to survey the extent of the originally identified selective sweeps across populations. For the analysis of potential functional effects of the identified alleles, information about the general function of the genes was used.

The two genes showed coding sequence variation among the populations. For *Xpr1*, 14 different alleles were identified which occurred as haplotypes. Two of these alleles were associated to a signature of positive selection and an introgression of one of these haplotypes from Iran to Southern France was observed. Furthermore, evidence for balancing selection on two haplotypes in three

other populations was found. Analysis of potential functional effects of the receptor alleles focused in associations between these alleles and variation in transcribed P-MLVs in captive mice. However, no association could be found and no infectious virions were detected. Together with finding from previous studies, the here obtained results suggest that *Xpr1* variation in wild mouse populations is shaped by ongoing adaptive evolution, but that to current knowledge other factors than disease induction by XPR1-dependent MLVs might drive this process.

For *Dmrt1*, one amino acid exchange between Asparagine (N) and Serine (S) was observed, while data from other studies suggested that no differential expression or alternative splicing occurred. N seemed to be the derived allele and generated a potential N-linked glycosylation site. It was also found to be potentially under positive selection in five populations from France. The functional analysis focused on potential differences in target gene regulation due to the different *Dmrt1* alleles. In a comparative microarray-based analysis across early postnatal stages I found that differential transcription of *Dmrt1* target genes was increased between the different alleles when *Dmrt1* was present. This result together with the population genetic data suggests that there is possibly a functional difference between the N and S allele that could be the target of positive selection, though more analyses will be necessary to gain insights into the potential beneficial nature of the N allele.

Declaration

In this thesis, I present my doctoral research. The design of the whole project was done together with my supervisor Prof. Dr. Diethard Tautz and the interpretation of the data was achieved through numerous discussions. The major parts of the practical laboratory work and data analysis were carried out by me. Till Sckerl supported the maintenance of the mouse breeding pairs, that were set up for this project, while the sampling was my responsibility. Further contributions are specified below for the two chapters:

Chapter 1:

DNA extraction from stock samples, as well as Sanger sequencing and microsatellite genotyping of all samples were mostly performed by Terry Solomon after I trained her in the methods and we analyzed the data together. She also designed primer pairs for the P-MLV analysis and helped with the first sampling test runs. The laboratory work and analysis concerning the transcribed P-MLVs and the final analysis of the sequencing and microsatellite data was done by me. This chapter was written by me with contributions from Diethard Tautz.

Chapter 2:

I took care of maintenance of the mouse breeding pairs during the time when cages had to be checked daily for pups. DNA extraction from stock samples, sequencing and microsatellite genotyping was partly performed by Heinke Buhtz. RNA extractions were carried out by me, including the quality assessment. Nicole Thomsen and Elke Blohm-Sievers conducted the microarray laboratory work. The arrays were designed by me and I also did the data analysis. The latter was supported by David Haase by Python scripts for data filtering. This chapter was written by me with contributions from Diethard Tautz.

General introduction

The analysis of the molecular basis of adaptation is of great interest in evolutionary biology. Consequently, during the past years a lot of effort has been made to identify genes that actively contribute to ongoing adaptation processes in wild populations.

A brief introduction to molecular evolution

The term molecular evolution describes the study of changes in DNA through time. It aims to understand biological changes at the molecular and cellular level by using principles of evolutionary biology and population genetics. This field is of central interest as adaptive and neutral evolution can only occur if there is variation in lifetime reproductive success among individuals and at least a part of the variation is heritable. Changes in the frequencies of alleles which are responsible for a trait represent a record on the action of selection or genetic drift. In other words, evolution works with phenotypes that influence lifetime reproductive success, but all evolutionary change in these phenotypes must have a molecular basis (Jobling et al., 2004; Stearns and Hoekstra, 2005).

First, it is important to know which evolutionary forces shape the molecular variation in natural populations and what their influence is. There are four such forces that are described in the following, namely mutation, gene flow, genetic drift and natural selection (Mitchell-Olds et al., 2007; Oleksyk et al., 2010; Stearns and Hoekstra, 2005).

Mutation and gene flow can introduce variation to populations at which mutation is the ultimate force that generates new variation. The latter occurs randomly, i.e. there is no systematic relationship between the phenotypic effect a mutation causes and the actual needs of the organism in which it occurs (Nachman and Crowell, 2000; Stearns and Hoekstra, 2005). Rates of mutation vary across genomic regions and marker types that are applied to assess genetic diversity in populations. Single nucleotide polymorphisms (SNPs) for example follow a more or less constant mutation rate of about 2.5×10^{-8} in humans (Nachman and Crowell, 2000), $2.1.x\ 10^{-8}$ in mice (Nachman, 1997) and 3×10^{-9} in insects (Andolfatto and Przeworski, 2001). In contrast, the mutation rates of microsatellites, as another often applied marker system, are highly locus specific, strongly correlated with the number of repeats at a locus and several orders of

magnitude higher than that of SNPs (reviewed in Ellegren, 2004). Other than mutation, gene flow cannot generate new variation but can introduce new variants to a population. This is the case when for example organisms are born in one place and move to another where they have offspring that survives to reproduce (Stearns and Hoekstra, 2005).

Complementary to the two forces that can generate new variation in a population, genetic drift and natural selection both act on this variation.

Genetic drift shapes neutral evolution and describes changes in allele frequency due to random drawing of gametes that will form the next generation (Hartl and Clark, 2006). It is the dominant force if a genetically determined trait has no influence on lifetime reproductive success of the organism carrying it, i.e. it has no influence on the number of offspring that survive and reproduce compared to the population average. Thus, it acts randomly on variation and generally plays a significant role in evolutionary change at the molecular level (Oleksyk et al., 2010). Its impact is especially large in small populations and often leads to the loss of new alleles which occur in low initial frequency. This relationship is described by the term mutation-drift equilibrium which means that the probability of fixation of an allele is its observed frequency, and the expected fixation time of a new allele equals four times the effective population size measured in generations (Nei, 1987).

Natural selection shapes adaptive evolution and acts when a genetically determined trait has an influence on lifetime reproductive success, i.e. when the individual carrying it has more or fewer offspring than the average number in the population that survive and reproduce. The resulting evolutionary change will move the trait and the associated allele frequencies from generation to generation in the direction of increased adaptation. This direction is induced through the nature of correlation between the trait value and reproductive success under the given environmental factors (Stearns and Hoekstra, 2005). This correlation results in different forms of possible selection that will be described in the following.

There is negative selection that is also often termed as purifying selection and constitutes the circumstance that deleterious alleles have a decreased chance of being transmitted to the next generation. This leads to a reduction in their frequency over time and finally to their loss (Bromham, 2008; Oleksyk et al., 2010). Commonly, the consequence is a conservation of DNA sequences. In this context, the term background selection is often used (Fig. 1.1) and refers to selection against recurrent deleterious alleles that also eliminates neutral or weakly selected

variants at closely linked sites in the genome (Charlesworth, 2009; Charlesworth et al., 1993; Nachman, 2001).

Positive Darwinian selection can indicate either balancing selection or directional selection (Hedrick, 2007). Balancing selection describes a process where genetic polymorphism in the form of multiple alleles at a locus is actively maintained in a population. It is often discussed in the context of heterozygote advantage, also termed overdominance, or frequency-dependent selection (Andrés, 2001; Hamilton, 2009; Hedrick, 2007; Mitchell-Olds et al., 2007; Oleksyk et al., 2010).

Directional or positive selection is a process that leads to an increased frequency of one allele which provides a reproductive advantage compared to other present alleles in the population. In this context, the term selective sweep is used as opposed to background selection if positive selection drives the beneficial allele and closely linked neutral alleles to high frequency in a population. This leads to a reduction in genetic diversity similar to what is observed under background selection (Fig. 1.1). The fixation time of a beneficial allele depends on the effective population size and the fitness advantage provided by the allele which is usually expressed as selection coefficient. The value of this coefficient quantifies the decrease in allele frequency of other than the beneficial allele from one generation to the next. Positive selection is acting if the product of four times the selection coefficient and effective population size is larger than one ($4N_e s > 1$) (Hartl and Clark, 2006). The curve describing the frequency increase of a beneficial allele across generations is commonly S-shaped. This means that changes are slow at low and high frequencies of the respective allele and fast at intermediate frequencies. Consequently, rare or new beneficial alleles are expected to increase relatively slowly in the beginning (Hartl and Clark, 2006; Stearns and Hoekstra, 2005).

Positive selection, in the sense of directional selection, plays a major role in phenotypic diversification and local adaptation (Rieseberg et al., 2002). Thus, there is great interest to identify genomic regions that have been subjected to positive selection and have been actively involved in adaptation processes (Harr et al., 2002; Sabeti et al., 2002).

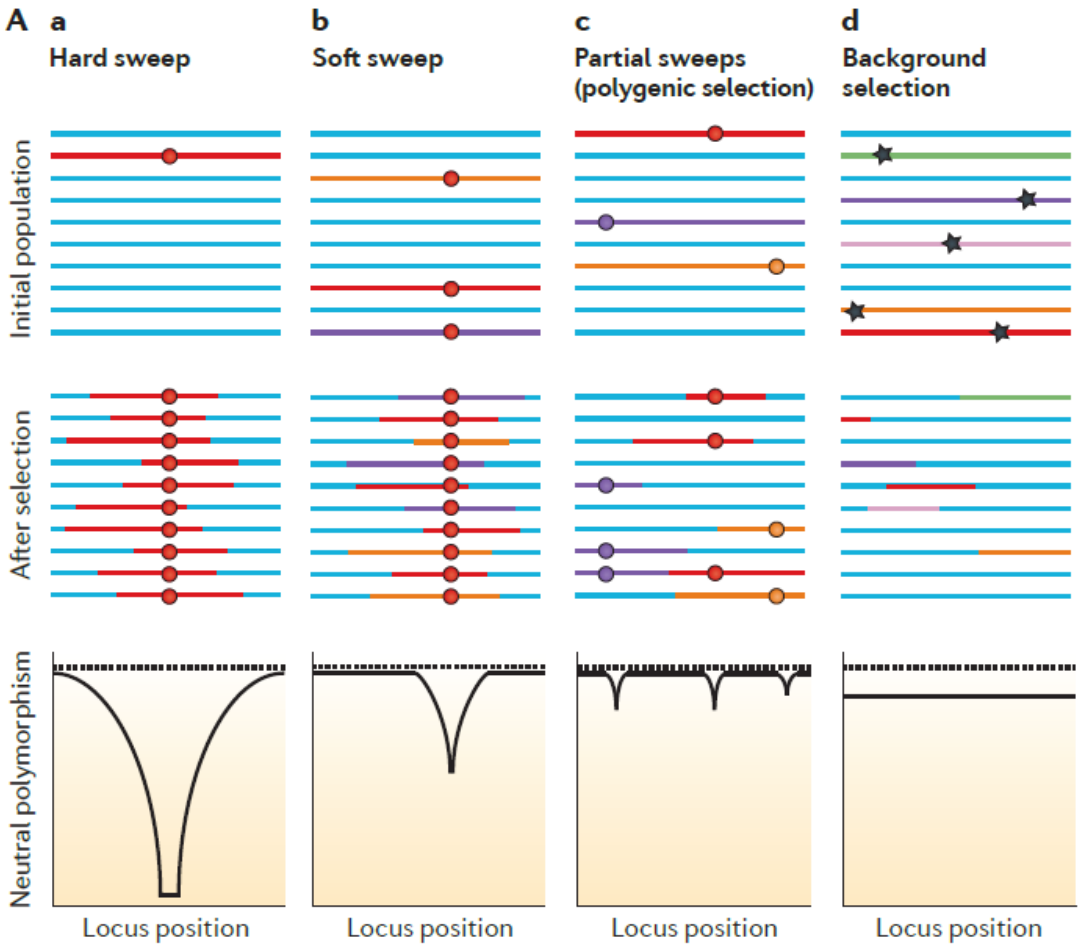


Figure 1.1: Genetic signatures of natural selection on an allele (filled dot) and its effects on neutral polymorphisms (differently colored bars) in the neighboring region which are influenced by the hitchhiking effect. (a) Represents the signature of positive selection on new or rare alleles (hard sweep). Few alleles are physically linked to the beneficial mutation when it is targeted by positive selection, and due to this the genetic polymorphism in the region will be extremely reduced. (b) shows the signature of positive selection on standing variation (soft sweep). Several neutral alleles are physically linked to the beneficial variant and thus, neutral polymorphism will be decreased less. (c) represents selection on a polygenic trait, which means that the selection intensity is spread across several variants and individual alleles experience weaker selection. (d) shows the case of background selection where neutral alleles that are linked to deleterious alleles are driven to extinction. (Taken from Cutter and Payseur (2013)).

Identification and analysis of adaptively relevant genes

Evolutionary processes like positive selection leave characteristic footprints in the genome. If a beneficial allele is targeted by positive selection not only this allele increases in frequency in the population but also the neighboring alleles which are physically linked due to their chromosomal position. This effect is called hitchhiking effect and provides the basis for various population genetic approaches to detect positive selection events. In large populations, hitchhiking can reduce the neutral diversity much more than random genetic drift and it was proposed as one

explanation why very abundant species do not show correspondingly high levels of genetic diversity (Maynard Smith and Haigh, 1974).

The size of the concerned region with reduced genetic diversity and the power to detect it as influenced by positive selection, depends on the recombination rate and intensity of selection. The intensity is positively correlated with the window size whereas the recombination rate is negatively correlated with it (Fay and Wu, 2000; Harr et al., 2002). The timeframe in which such a pattern of positive selection will be observable depends on the mutation rate of the applied marker. After a beneficial allele has reached fixation, its genetic footprint in the neighboring region will be gradually lost by reconstituting genetic variability through mutation and subsequent recombination. This process leads to so called recovery patterns which are characterized by an excess of rare and new mutations at low frequencies. Altogether, the footprint of positive selection is thus characterized by a locally reduced level of genetic diversity (Maynard Smith and Haigh, 1974), an altered allele frequency spectrum due to recovery (Braverman et al., 1995; Tajima, 1989) and a locally increased extend of linkage equilibrium (Kim and Stephan, 2002; McVean, 2007; Przeworski, 2002).

Knowledge of the characteristic signatures of positive selection provides the opportunity to directly use genomic information to detect its targets without using information about phenotypic variation (Sabeti et al., 2002; Schlötterer, 2003; Stearns and Hoekstra, 2005). This concept can be implemented by different approaches. Traditionally, most tests for selection have concentrated in comparing a set of variable markers within a gene region against assumed neutral expectations. In other words, these statistical tests aim to discriminate between natural selection and genetic drift as the most likely cause for the observed differences in homologous DNA sequences between populations or species (Oleksyk et al., 2010; Schlötterer, 2003; Stearns and Hoekstra, 2005).

An example for such a test is Tajima's D which tests for specific changes in the allele frequency distribution i.e. the accumulation of rare alleles in a region. In order to this, the local nucleotide diversity is compared with the number of segregating sites. If the nucleotide diversity is significantly lower than the number of segregating sites, the value becomes negative and is interpreted as a sign for positive selection whereas a positive value indicates balancing selection (Tajima, 1989). Another commonly applied statistics is Wright's F_{st} which detects regions of increased differentiation between populations which can be interpreted as a sign for differing local selective pressures (Wright, 1951). Both above mentioned methods and many

more have often been applied in the context of candidate gene approaches which require an a priori knowledge about the function of a gene and its potential involvement in adaptation processes. Because of the suspected function homologous sequences are compared within or across populations to check for a possible significant departure from neutral variation patterns (Schlötterer, 2003).

These standard tests of neutrality that use mostly locus-specific sequence polymorphism data are often sensitive to past demographic events as underlying assumption for neutrality, like constant population sizes are often violated in natural populations (Nielsen, 2001; Otto, 2000; Przeworski, 2002). For example, a population expansion would also increase the proportion of low frequency variants and thus mimic the effect of a selective sweep as identified by Tajima's D. On the other hand, a population bottleneck could produce an excess of intermediate frequency variants which would result in a signature similar to that induced by balancing selection (Nielsen, 2005).

Opposing to candidate gene approaches, hitchhiking mapping or mulitlocus screens refer to genome-wide surveys for genes that have been under positive selection in natural populations. In this case, no a priori assumptions about phenotypic effects are made and these screens have become more common in a range of species with increasing availability of polymorphism data (Harr et al., 2002; Oleksyk et al., 2010; Sabeti et al., 2007; Schlötterer, 2003; Staubach et al., 2012; Storz et al., 2004). Besides the advantage of identification of genes or adaptively relevant alleles that have no known or easily visible phenotypic effect, this also contributes to our knowledge about the overall influence of positive selection on the genome variation. The latter will also support a more complete functional characterization of the genome complementary to phenotypic-based approaches (Barton, 2000; Schlötterer, 2003).

The markers of choice for multilocus screens are often SNPs which have been successfully applied to identify selective sweeps in big data sets for example from humans and mice by means of different statistics. The focus of these statistics is mostly on changes in allele frequency or allelic differentiation among populations in regions of extended linkage disequilibrium. The "extended haplotype homozygosity" (EHH) test for example, searches for sites where the change in allele frequency since the start of a putative selection event occurred too quickly to be due to random drift. The start if the putative selection event is assessed by the derived allele frequency, and the speed of the allele frequency change is measured by the extend of linkage disequilibrium around the tested allele (Sabeti et al., 2002, 2007).

Another test based on this idea is the “cross-population composite likelihood ratio” (XP-CLR) test which tests for allelic differentiation that must have occurred more quickly than just due to genetic drift. Again the speed of the differentiation is assessed by the linkage disequilibrium in the region that is expected to be more extended under selection than under neutrality (Chen et al., 2010). This test has been applied in humans and natural populations of the mouse (Chen et al., 2010; Staubach et al., 2012).

Additionally to SNP-based methods, approaches have been developed that use multiple microsatellite markers to detect recent selective sweeps. Though these loci are neutrally evolving, they can also be affected by a hitchhiking effect and it has been shown that their repeat number is strongly influenced by selection (Slatkin, 1995). They are less dense in the genome than SNPs but are multiallelic and thus highly informative (reviewed in Ellegren, 2004; Schlötterer, 2003). A problem arises from the fact that no uniform background mutation rate can be assumed for these loci as can be done for many changes on DNA sequence level. The mutation rate of microsatellites appears to vary widely among loci and alleles, which prevents testing across different loci (reviewed in Ellegren, 2004; Slatkin, 1995). To circumvent this problem, hitchhiking approaches based on this type of marker rely on comparing variability levels of the same marker in different populations. The variability is measured in terms of variance in repeat number (RV) or expected heterozygosity (RH). Furthermore, a large number of unlinked loci is included in the analysis which helps to integrate the demographic history of the studied populations und thus makes this approach very robust to various demographic scenarios and reliable for the identification of recent selective sweeps (Kauer et al., 2003; Schlötterer, 2002). Such microsatellite screens have been applied to natural populations of the fruit fly *Drosophila melanogaster* (Harr et al., 2002) and mouse subspecies *M. m. domesticus* and *M. m. musculus* (Büntge, 2010; Ihle et al., 2006; Teschke et al., 2008).

The comprehensive application of SNP- or microsatellite-based multilocus screens has revealed a high number of selective sweeps in populations of humans and mice. A genome-wide SNP-based hitchhiking mapping in human populations from Africa, Asia and Europe suggested about 300 strong candidate regions (Sabeti et al., 2007). In mice, SNP- and microsatellite-based approaches in four different populations of two mouse subspecies from Czech Republic and Kazakhstan (*M. m. musculus*) and from Germany and France (*M. m. domesticus*) indicated 200 to 300 positive selection events per lineage (Staubach et al., 2012;

Teschke et al., 2008). These numbers probably represent an underestimate as the hitchhiking mapping will mostly detect positive selection events on new and rare alleles (hard sweep) and not on standing variation (soft sweep) (Fig. 1.1). In the case of microsatellite studies, also only very recent events can be traced as the sweep signatures will be lost relatively fast due to the high mutation rates (Staubach et al., 2012; Teschke et al., 2008).

After the identification of selective sweeps, questions about the function of the involved genes and their allelic variation in this context can be addressed. There are several studies that have worked on such candidate genes to investigate the molecular basis and associated phenotypic traits that are the actual targets of positive selection. In the following, two examples for studies on the molecular basis of local adaptation in deer mice (*Peromyscus*) are briefly described.

The first example by Storz et al. (2007) is a population genetic analysis of a two-locus α-globin polymorphism that is apparently involved in adaptation to oxygen-availability along an elevation gradient. The molecular basis of this adaptation has been characterized as functionally distinct protein alleles due to five amino acid exchanges in strong linkage disequilibrium. These alleles seem to be maintained as long-term balanced polymorphism and change in frequency from low to high altitude. Furthermore, they were predicted to change the oxygen affinity of α-globin which probably represents a physiological adaptation to high-altitude hypoxia.

Another well studied example for a molecular change that is involved in local adaptation is the charge changing amino acid replacement in the *melanocortin-1 receptor* (*Mc1r*) in beach mice. The mutation decreases the function of the receptor and is apparently involved in controlling pigmentation differences between the lightly colored beach mice and their darker colored mainland conspecifics. More detailed analysis also revealed epistatic interactions between this locus and the *Agouti* candidate gene which is a ligand of MC1R and has alleles which differ in expression levels. The different observed phenotypes probably evolved due to selection for crypsis as the lighter fur provides an advantage for the beach mice on light-colored sand. Furthermore, analyses have shown that the genetic basis of the light coloration seems to have evolved independently in different parts of the species range and is controlled by variation at different loci (Hoekstra et al., 2006; Steiner et al., 2007).

Although there are more examples like these two, knowledge about the molecular basis of adaptation processes remains relatively rare compared to the amount of

identified candidate genes. It still remains a challenge of central interest in evolutionary biology to combine population genetic approaches with information on molecular function, phenotypic variation and potential ecological consequences. Though this type of approach is mostly restricted to a gene-by-gene basis, it will help to gain insights into how natural selection influences phenotypic traits through molecular changes and to understand the basis of adaptation in natural populations.

The wild mouse (*Mus musculus* L.) as a model system

The house mouse (*Mus musculus*) has become one of the major model organisms especially for genetics and medical research. Easy handling in the laboratory, small size and short generation times already make it an attractive study organism. Additionally, it is also a mammal which makes studies often comparable to higher organisms like humans. Due to this attractiveness the house mouse has been well studied for decades including its phylogeny and history which also makes it interesting for evolutionary biologists (Boursot et al., 1993, 1996; Guénet and Bonhomme, 2003).

Extensive genomic and molecular data is available for the house mouse. Since 2002 the complete genome sequence as derived from the lab strain C57BL/6J is accessible (Mouse Genome Consortium 2002). Then in 2011, complete genome sequences of altogether 17 inbred strains became available which showed extensive genetic and functional variation among these strains (Keane et al., 2011). Additionally to the genetic information, gene function is studied systematically by the former Knock-out Mouse Consortium and since 2011 by the International Mouse Phenotyping Consortium through generation of targeted knockout mutations in embryonic stem cells. Mutant mouse lines are then tested through an extensive phenotyping pipeline to describe the biological function of each gene.

About the phylogeny and history of the house mouse we know that it originates from the Indian subcontinent. There, *M. musculus* split into the three major subspecies *M. m. castaneus*, *M. m. musculus* and *M. m. domesticus* about 0.5 MYA (Fig. 1.2).

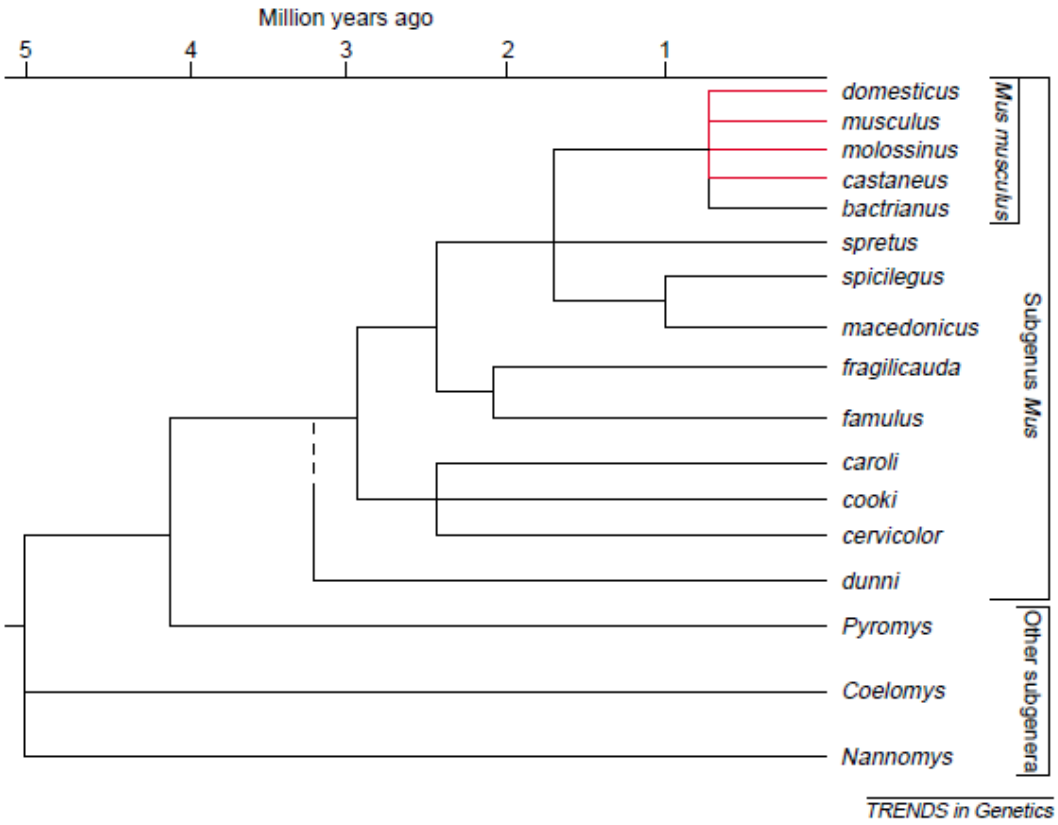


Figure 1.2: Phylogeny of the genus *Mus* (take from Guénet and Bonhomme (2003)).

Today, *M. m. castaneus* is distributed all over South East Asia, *M. m. musculus* can be found all over Northern Asia and Eastern Europe, while *M. m. domesticus* populations range from the Near East to Northern Africa and Western Europe. From Western Europe the house mouse has then spread all over the world with the help of human transport. The three lineages are not completely genetically isolated from each other and form natural hybrid zones in contact areas. *M. m. domesticus* and *M. m. musculus* meet in Central Europe in a narrow, stable hybrid zone as well as in the Caucasus and a region southeast of the Caspian Sea. *M. m. castaneus* and *M. m. musculus* have a hybrid zone in China and another one in Japan where they hybridize extensively and form a unique population which is often referred to as *M. m. molossinus* (Yonekawa et al., 1988). There is no natural contact zone for *M. m. domesticus* and *M. m. castaneus*, but they can produce fertile offspring in the laboratory (Fig. 1.3).

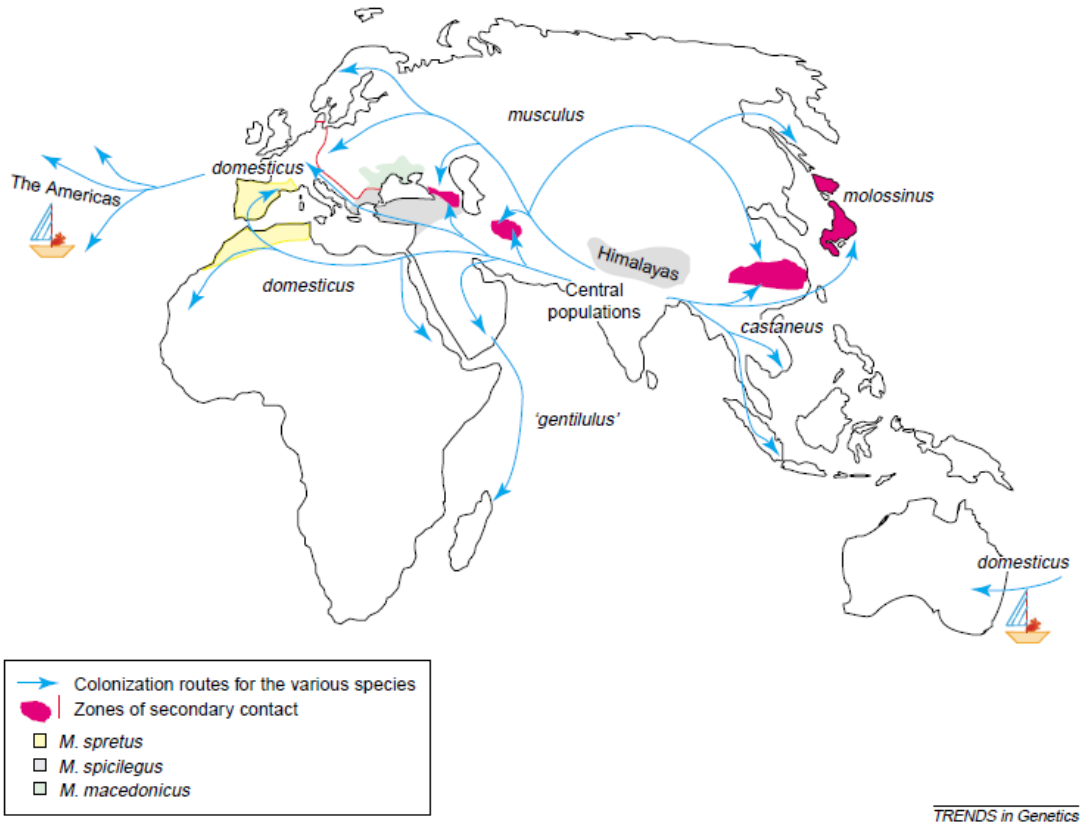


Figure 1.3: Today's geographical distribution of the genus *Mus* including secondary contact zones and colonization routes (taken from Guénet and Bonhomme (2003)).

The colonization of Western Europe by *M. m. domesticus* originating from the Near East, happened through the Mediterranean route about 3,000 years ago. This process is documented by detailed paleontological records (Cucchi et al., 2005). Humans already settled in Europe 6,000 BP, but the mice followed the Neolithic diffusion about 5,000 years later. This delay has been attributed to two possible reasons. The first reason is the limited maritime exchange between the Eastern and Western Mediterranean Basin until 1,000 BP which probably only allowed for a weak migratory flow before the increase. The second possible reason is the absence of suitable ecological niches. Neolithic commensal niches were probably poorly represented and less stable in Western Europe than in the Near East until 1,000 BP and favored by the wood mouse. Then, with an increase in human settlements more stable commensal niches became available that were then favored by the house mouse. As a result, a massive invasion of Europe by house mice took place 3,000 years ago (Cucchi et al., 2005).

As well as this study, previous studies on the population genetics of wild mouse populations and the occurrence of selective sweeps are based on samples from wild populations (Büntge, 2010; Ihle et al. 2006; Linnenbrink, 2012;

Staubach et al., 2012; Teschke et al., 2008). The studied *M. m. domesticus* populations come from different locations in France, Germany and Iran, while the *M. m. musculus* populations originate from Czech Republic and Kazakhstan. A study of Ihle et al., 2006 on several populations of both subspecies has shown that they are genetically distinct which allowed the identification of population-specific selective sweeps (Büntge, 2010; Staubach et al., 2012; Teschke et al., 2008).

Altogether, the extensive background knowledge on genes and gene function, well documented population history, data from previous population genetic analyses and screens for selective sweeps provide an ideal basis to study molecular evolution.

Scope of the thesis

In this study two candidate genes were analyzed which were previously identified as potential targets for positive selection events in wild populations of the house mouse by SNP- and microsatellite-based hitchhiking mapping approaches (Büntge 2010, Staubach et al. 2012). The presence of putative sweep signatures at these genes implied that there are functionally distinct alleles present in nature. Both genes were chosen as candidates for this study as their general function or adaptive potential has already been well characterized in previous studies. This provided the basis to address the following questions in this study:

(1) What are the potential functional alleles of the two genes that are targeted by positive selection in wild populations?

(2) How are these alleles characterized i.e. do they for example occur in the form of expression changes or as protein-coding change?

(3) How is the geographic variation of these alleles and the associated sweep signature in different populations and subspecies?

(4) Are there functional differences between the identified alleles when looking at gene-specific traits?

In the first chapter, I describe a population genetic analysis of the gene *Xpr1* (*xenotropic and polytropic retrovirus receptor 1*) which is particularly known for its key role in the interaction with murine leukemia viruses (MLVs). Thus, here I aimed at identifying potential associations between the different receptor alleles and present MLVs.

The second chapter deals with a similar population genetic analysis of the transcription factor *Dmrt1* which is known to be involved in sex differentiation in mice. Consequently, here I focused on potential effects of the different alleles on target gene transcription.

Chapter 1

Natural variation in the cell-surface receptor Xpr1 in wild populations of the house mouse

Abstract

The interaction between the cell-surface receptor *Xpr1* and the xenotropic and polytropic murine leukemia viruses in mice is a suitable model to study the molecular basis of adaptive evolution and the factors shaping this process. X/P-MLVs can cause disease in mice such as leukemia and lymphomas and use XPR1 as receptor to infect host cells. Previous studies already demonstrated substantial co-evolution between the interaction partners and that *Xpr1* is targeted by positive selection in wild mouse populations. To complement the previous studies, I analyzed 11 wild mouse populations of the two subspecies *Mus musculus domesticus* and *M. m. musculus* for *Xpr1* coding sequence variation and associated variability in neighboring microsatellites. Furthermore, I analyzed the P-MLV receptor-binding domain variation of transcribed proviruses in captive, wild-derived mice from three different origins to assess possible functional effects of the allelic variation of *Xpr1*. The *Xpr1* sequence analysis revealed the presence of 14 different *Xpr1* haplotypes that were characterized by several SNPs and varied in abundance and frequency among populations. I identified signatures of positive selection in a Czech *M. m. musculus* population, and in an ancestral *M. m. domesticus* population from Iran on different haplotypes. In the latter population, the advantageous haplotype was characterized by a missing potential N-glycosylation motif in one of the extracellular loops which is involved in virus interaction. There also seemed to be introgression of this haplotype between the Iranian and a Southern French population. Additionally, I found signatures of balancing selection in three other French populations. The analyzed proviruses of captive mice exhibited a variation pattern which is consistent with the mouse population history, and I did not detect any infectious XP-MLV virions. Altogether, our findings and previous studies suggest that *Xpr1* variation in wild mouse populations is shaped by ongoing adaptive evolution, but that to current knowledge disease induction by XPR1-dependent MLVs does not seem to be the main factor driving this evolution.

Introduction

Host-virus interactions are an important driver of evolutionary processes and the characterization of their molecular basis is of great interest in evolutionary biology (Woolhouse et al., 2002). An extensively studied example is the interaction between xenotropic and polytropic murine leukemia viruses (XP-MLVs) and the xenotropic and polytropic retrovirus receptor 1 (XPR1).

Xpr1 is highly conserved in metazoans on the amino acid sequence level and expressed in various cell types. The gene encodes a cell-surface receptor with eight transmembrane domains which putatively form four extracellular loops (ECL). It belongs to the group of G protein-coupled receptors and has been shown to function in the export of inorganic phosphate (Battini et al., 1999; Giovannini et al., 2013; Tailor et al., 1999; Vaughan et al., 2012). In contrast to its only recently described native function, the receptor was already described as P-MLV susceptibility gene in mice in 1983 (Kozak, 1983, 2010). Subsequent studies then demonstrated that XPR1 mediates infection of cells by both P- and X-MLVs in a variety of hosts (Chesebro and Wehrly, 1985; Cloyd et al., 1985; Kozak, 1985; Lyu and Kozak, 1996). XP-MLVs belong to the gammaretroviruses and are distinguished according to their host tropism, interference pattern and pathogenicity (Chesebro and Wehrly, 1985; Cloyd et al., 1980, 1985; Hartley et al., 1977; Yan et al., 2010). Generally, MLVs can cause leukemia and lymphomas, but their pathogenicity is highly variable and dependent on virus strain and host background (reviewed in Kozak, 2010). Sources of this variation are the generation of pathogenic P-MLVs (mink cell focus forming MLVs) by recombination events, genetic variability in the receptor binding domain of the viral Env glycoprotein and representation of proviral elements in the respective genome (Battini et al., 1992; Cloyd et al., 1980; Fischinger et al., 1975; Hartley et al., 1977; Kozak, 2013; Stoye et al., 1991).

Co-evolution between XPR1 and XP-MLVs has been suggested in various studies based on lab strains and samples of wild-caught mice from scattered locations. Altogether, five MLV-restrictive alleles of *Xpr1* have been identified which for example lead to the resistance of *Mus musculus castaneus* against P-MLV infection or of some lab strains against X-MLV infection. Mutagenesis and phylogenetic analyses identified XPR1 residues which are important for virus interaction and showed that the receptor has been under re-current positive selection (Kozak, 1985; Lyu and Kozak, 1996; Marin et al., 1999; Tailor et al., 1999;

Yan et al., 2007, 2010). Furthermore, in the genome-wide SNP analysis of Staubach et al. (2012), *Xpr1* was found to show a difference in haplotypes between a French and a German population which could suggest that positive selection was acting on it.

Altogether, there are a lot of data available which show that there is functional variation among *Xpr1* alleles and that this plays an important role in the adaptation of mice to infections by XP-MLVs (reviewed in Kozak, 2010; Kozak 2013). However, information about the receptor variation on this level in the wild is still rare. In this study, I analyzed the variability of *Xpr1* alleles and associated microsatellite loci on a finer scale compared to previous studies in European wild populations of two mouse subspecies. I followed a sampling scheme that allowed us to assess the variation in unrelated individuals from different geographic regions. Furthermore, I tested captive mice originating from different wild populations for a possible association between population-specific XPR1 alleles and transcribed P-MLV-related RBD (receptor-binding domain) variants.

Material and Methods

Mice and sampling

Population samples

122 samples from two subspecies and 11 distinct wild mouse populations were analyzed to assess patterns of allelic variation of *Xpr1*. Samples were collected by Ihle et al. (2006) and Linnenbrink et al. (2013) according to the same sampling strategy (Tab. S1, Fig. 2.1). The samples represented distinct sampling sites to ensure that individuals were unrelated (Ihle et al. 2006).

Samples from the breeding stock

From the wild-derived breeding stock in Plön (Germany), I collected ear punches of two *Mus spretus*, *Mus spicilegus* and *Mus musculus castaneus* individuals to complement the population sampling. Additionally, I collected feces samples and ear punches of six Ger_{CB}, six Dom-IR and 12 Fra_{MC} mice to analyze variation in the receptor binding domain (RBD) of P-MLVs with regard to allelic differences in *Xpr1*. The animals had been kept in the stock over several generations and the breeding followed an outcrossing scheme. The feces samples were stored at -80°C for later RNA extraction and analysis of P-MLV variation. The ear punches were transferred to HOM buffer and stored at 4°C until DNA extraction and *Xpr1* allele determination.

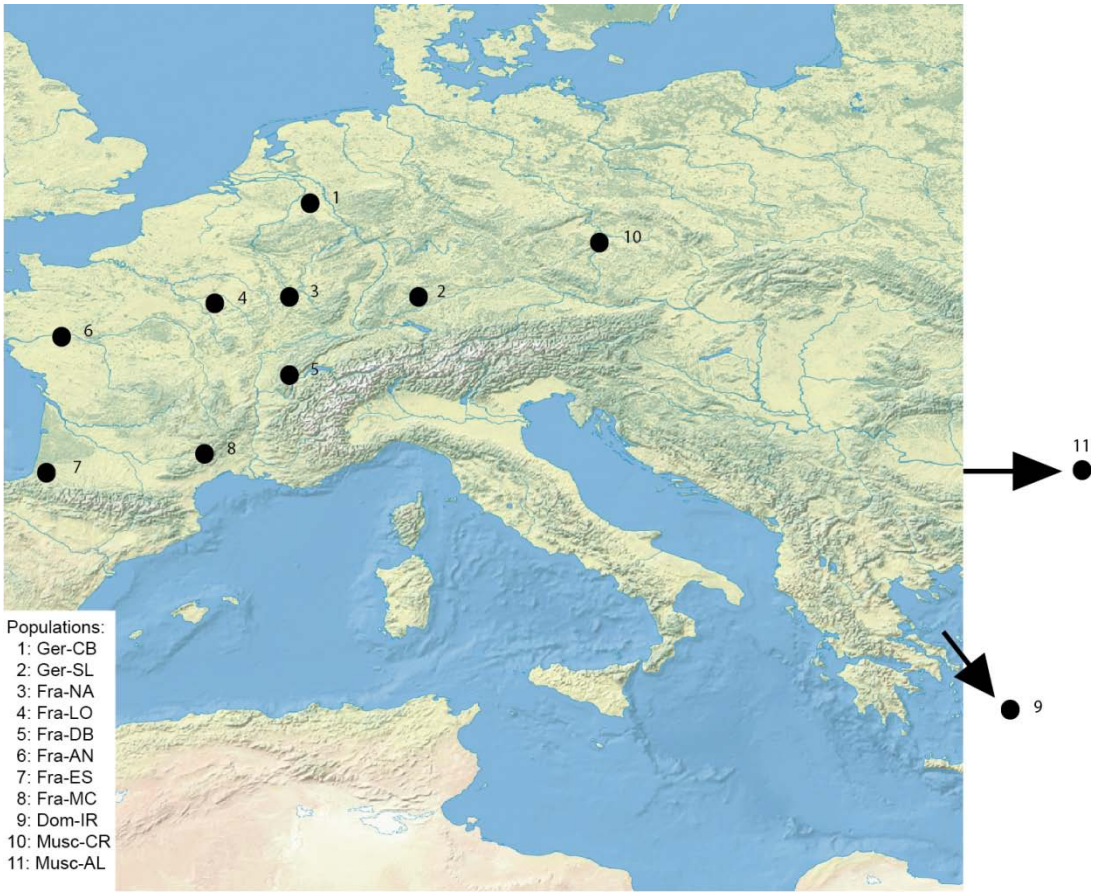


Figure 2.1: Overview map of the tested wild mouse populations.

Experimental crosses

To test whether the identified virus variants occurred in the form of infectious particles or as transcribed proviruses, I conducted a crossing experiment with two Iranian x German pairs. I chose this combination as mice from the two populations exhibited easily distinguishable virus variants. The males were removed from cages as soon as the females were visibly pregnant. Before birth, the females were moved to fresh cages, and thus the pups were only in contact with the mothers. The feces samples were collected from the adult mice before and after mating, while the pups were sampled upon weaning. Therefore, I accounted for eventual virus transmission between the parents, from mother to pups or inheritance of the viruses. The feces samples were stored at -80°C until RNA extraction.

Extraction of nucleic acids

DNA from the population samples was extracted as described in Ihle et al. (2006) or Linnenbrink et al. (2013), respectively. Extraction of DNA from ear punches of mice from the breeding stock was done by a standard salt extraction procedure. Until further processing, DNA was stored at -20°C and diluted to 5 ng/µL for PCR.

RNA from feces samples was extracted using TRizol (ambion) in combination with the PureLink RNA Mini Kit (ambion) following the protocol by the manufacturer. The samples were stored at -80°C.

Molecular analysis of DNA and RNA

cDNA synthesis

The RNA from the feces samples was reverse transcribed for later PCR and Sanger sequencing. The first strand cDNA synthesis was performed with help of the MMLV High Performance Reverse Transcriptase (epicentre). This was done by using 350 ng of extracted RNA and following the manufacturer protocol. Afterwards, cDNA was stored at -20°C.

Microsatellite genotyping

I typed nine microsatellite loci within a region of 200 kb around *Xpr1* (Fig. 2.2). The loci were identified applying the software “tandem repeats finder” (Benson 1999) and primers were designed with “Primer3Plus” (Untergasser et al. 2007). The forward primers were labeled with Hex at the 5’ end and four primer pairs were pooled per reaction. Care was taken that the pooled primer pairs yielded distinct product size ranges (Tab. S2.2). PCRs were carried out by means of a multiplex PCR kit (Qiagen) in 5 µL final volumes and using 5 ng DNA as template. Amplification conditions were as follows: 95°C for 15 min followed by 28 cycles at 95°C for 30 s, 60°C for 1.30 min, 72°C for 1.30 min with a final extension at 72°C for 10 min. Afterwards, PCR products were diluted 1:20 in water and 1 µL was transferred to 10 µL Hidi formamide and 0.1 µL 500 Rox size standard (Applied Biosystems). The subsequent denaturation step was performed with the following incubation times: 90°C for 2 min and 20°C for 5 min. Product sizes were automatically determined by a 3730 DNA Analyzer (Applied Biosystems), and alleles were called by means of the GeneMapper v4.0 software (Applied Biosystems).

Sanger sequencing

First, I identified variation in *Xpr1* alleles by sequencing parts of the coding sequence supposed to be variable which were exon 4 and two putative extracellular loops (ECL3 and 4; Fig. 2.2). For this purpose, six primer pairs (Metabion) that were designed with “Primer3Plus” were used (Tab. S2.3). Secondly, I analyzed the variation in the receptor binding domain of the surface unit of the viral envelope gene of P-MLVs.

PCR reactions were carried out in 10 µl final volume with a multiplex PCR kit Qiagen and following cycling conditions: 95°C for 15 minutes followed by 35 cycles of 95°C for 30 s, 60°C for 1.30 min, 72°C for 30 s/90 s and 10 min at 70°C for elongation time. Exo-Sap purification (USB Corp.) was performed with the following incubation: 37°C for 20 min and 80°C for 20 min. Cycle sequencing reaction was done with help of the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Reaction parameters were 96°C for 1 min followed by 29 cycles of 96°C for 10 s, 55°C for 15 s and 60°C for 4 min. Then, sequencing products were purified with the BigDye XTerminator Purification Kit (Applied Biosystems). Sequences were automatically generated by a 3730 DNA Analyzer (Applied Biosystems). Independent base calling and analysis was done using CodonCodeAligner v4.0.2 (CodonCode Corp.).

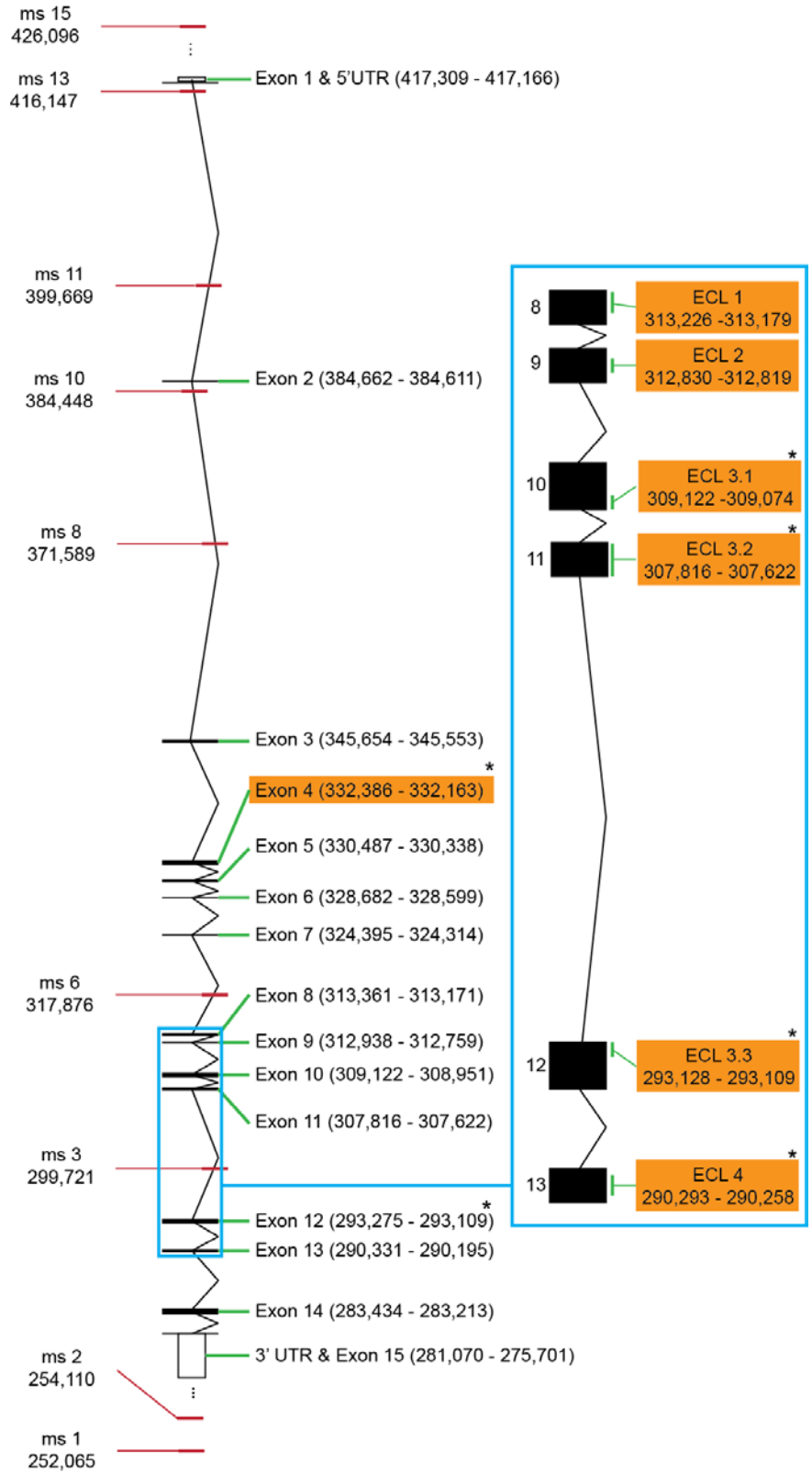


Figure 2.2: Genomic overview of the analyzed candidate gene on chromosome 1 (chr1:155). The exons are shown by black and the microsatellite loci by red boxes. Their respective genomic positions are indicated next to the boxes. On the right side an enlarged view of exons 9 to 13 is shown. Sequenced parts of the gene are represented by orange boxes and variable parts are marked by a star.

Population genetic analysis

The phase of *Xpr1* alleles was determined manually with regard to the homozygous individuals. Then, allele frequencies were calculated for all populations and their spatial distribution was visualized on a map made by NaturalEarth. A median network was generated with SplitsTree4 v4.13.1 (Huson and Bryant, 2006), while Tajima's D calculations were done in DNAsp v5 (Librado and Rozas, 2009).

For the microsatellite loci, allele frequencies in populations were calculated with regard to *Xpr1* allele and plotted in histograms. Those were then visually inspected for signatures of positive selection and associations between *Xpr1* and microsatellite alleles.

The aligned sequences of the receptor binding domain in the surface unit of the viral envelope were all inspected for SNPs and a phylogenetic tree was calculated using MrBayes v3.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003).

Results

Haplotypic variation of *Xpr1* in wild mouse populations

To assess the allelic variation of *Xpr1* on the population level in the wild, I analyzed Sanger sequences of different parts of *Xpr1* from several species and population samples.

Between 11 and 12 individuals per population and two individuals per outgroup could be sequenced successfully. The analysis revealed the presence of altogether 14 different *Xpr1* alleles. These alleles occurred in the form of distinct haplotypes which were characterized by eight nonsynonymous SNPs, nine synonymous SNPs and two indels. The two indels occurred in *Mus musculus castaneus* and *musculus* and were located in the forth putative extracellular loop (ECL4). The SNPs were distributed over exon 4, exon 12 and ECL 3, and 4 (Tab. 2.1, Tab. S2.4).

Table 2.1: Identified *Xpr1* haplotypes in wild populations of the house mouse or wild-derived mice from the breeding stock. Grey background indicates a nonsynoymous SNP, blue letters indicate the SNPs located in one of the extracellular loops.

haplotype comp.	sequence: EX4										ECL3.2		ECL3.3	EX12		ECL4																species/population	color code
I	.	G	.	A	.	A	.	T	.		.	.			C	.																	*Mus spicilegus*	
II	.	G	.	A	.	A	.	T	.		.	T	.		.	A																	*Mus spretus*	
III	.	G	.	.	.	.	C	.	.		.	.	.		.	.																	*Mus musculus castaneus*	
IV	.	G	G	A	.	.	.	.	.		C	.	.		A	.	.	-	-	-	-	-	A	.	.	.	.	.	.	.	.	.	Musc-AL	
V	A	G	G	A	.	.	.	.	.		C	.	.		A	.	.	-	-	-	-	-	A	.	.	.	.	.	.	.	.	.	Musc-AL	
VI	.	G	G	A	.	.	.	.	.		C	.	.		.	.	.	-	-	-	-	-	A	.	.	.	.	.	.	.	.	.	Musc-CR	
VII	.	G	G	A	.	.	.	.	.		.	.	A		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Dom-IR, Fra_{MC}	
VIII	.	.	.	.	.	.	.	.	.		.	.	A		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Dom-IR, Fra_{MC}	
IX	G	C	A	G	T	G	T	C	A		A	C	C		T	G	T	G	A	T	T	A	C	A	A	C	G	T	T	T	A	A	Fra_{AN}, Fra_{NA}, Fra_{MC}, Fra_{LO}, Fra_{ES}, Fra_{DB}, Ger_{CB}, Ger_{SL}	
X	.	G	.	A	C	.	.	.	.		.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Fra_{AN}, Fra_{NA}, Fra_{MC}, Fra_{LO}, Fra_{ES}, Fra_{DB}, Ger_{CB}, Ger_{SL}	
XI	.	G	G	A	.	.	.	.	.		.	T	A		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Fra_{MC}	
XII	.	G	G	A	.	.	.	.	.		.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Fra_{MC}	
XIII	.	.	.	.	.	.	.	.	G		.	.	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Fra_{MC}	
other	.	S	R	R	Y	.	.	.	.		.	Y	W		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Fra_{MC}	
amino acid	R/K	G/A		T/A							K/Q		N/K		V/M	A/V		I/-		T/-		T/K/-		F/-		K/-		P/-					
genomic position chr1:155,	332,362	332,302	332,295	332,294	332,291	332,281	332,244	332,229	332,204		307,799	307,671	307,656		293,268	293,255	293,254	293,250	290,265	290,264	290,263	290,256	290,255	290,254	290,253	290,252	290,251	290,250	290,249	290,248,	290,247	290,246	290,245	290,144	290,243	290,242		

The identified *Xpr1* haplotypes varied in abundance and frequency in the populations. Of the eight European *Mus musculus domesticus* -populations, seven exhibited two haplotypes in differing, but mostly intermediate frequencies. Only the population Ger_{CB} was almost fixed for one of the two haplotypes. The population Fra_{MC} showed the highest *Xpr1* diversity with eight different haplotypes. Notably, I could identify a new haplotype in the Iranian (Dom-IR) population which was ancestral to the European *M. m. domesticus* -populations. This haplotype, was almost fixed in Dom-IR, but was also found in one other population in Southern France (Fra_{MC}). Furthermore, it was characterized by a nonsynonymous SNP in ECL3. A comparison between this haplotype, termed $Xpr1^{i}$ and previously identified virus restrictive haplotypes is shown in Tab. 2.2. The two analyzed *M. m. musculus* populations exhibited haplotypes that were different from the *M. m. domesticus* variants. While Musc-AL showed three different haplotypes, Musc-CR was fixed for one (Fig. 2.3).

Table 2.2: Comparison of SNPs in ECL 3 and 4 of *Xpr1*. The first five alleles were identified in previous studies and known to be virus restrictive (Kozak et al. 2010). The sixth allele (*Xpr1*^{i}) is the one identified in an Iranian and Southern French population in this study.

Allele	SNP												
	ECL3					ECL4							
Xpr1^{sxv}	K	N	S	T	V	I	A	T	T	F	K	P	D
Xpr1^{n}	E	.	.	.	.	.	.	-	.	.	.	.	N
Xpr1^{m}	.	.	.	.	.	-	.	.	K	.	.	.	.
Xpr1^{c}	.	.	.	.	.	.	.	-	-	-	-	-	.
Xpr1^{p}	K	.	P	Y	K	.	V	.	.	.	.	.	.
Xpr1^{i}	.	K	.	.	.	.	.	.	.	.	.	.	.
Residue	500	503	505	507	508	579	581	582	583	584	585	586	590

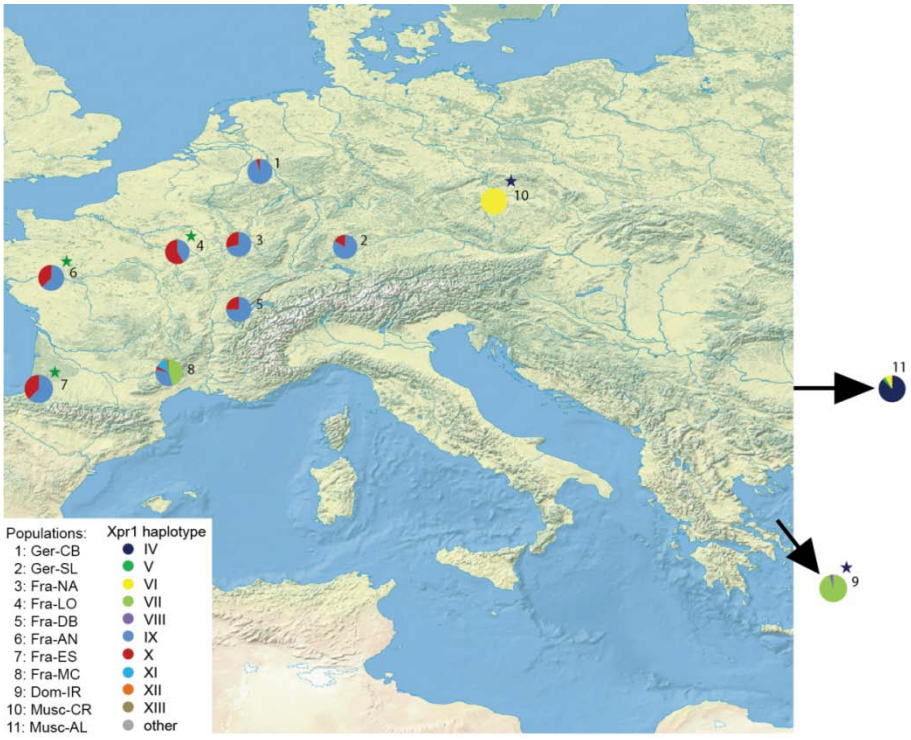


Figure 2.3: Map of the haplotype frequencies of *Xpr1* in wild populations of the house mouse. The color code corresponds to the one in the haplotype overview table (Tab. 4).

I also used the sequencing data to search for patterns of positive selection in the populations. Complementary to that, I analyzed several microsatellite loci in the region of *Xpr1* to identify selective sweeps. The distribution of *Xpr1* haplotypes across populations demonstrated that there is no clear differentiation of *Xpr1* according to populations. The samples rather differed by *Xpr1* haplotype than by population of origin (Fig. 2.4). Analysis of Tajima's D indicated significant balancing selection on two haplotypes in the three *M. m. domesticus*-populations Fra_{LO}

(Tajima's D = 2.249), Fra_{AN} (Tajima's D = 2.079) and Fra_{DB} (Tajima's D = 2.079). Separate analysis of the sequenced parts of *Xpr1* revealed that this signal is caused by exon 4 which exhibits most of the SNPs. With the microsatellite data, I was able to identify a selective sweep in the ancestral Iranian population. Furthermore, there was apparently introgression of this novel *Xpr1* haplotype between Iran and Southern France (Fra_{MC}) as it was mostly associated to the same microsatellite alleles in both populations. Another positive selection event seemed to have happened in Musc-CR where a different haplotype was found to be fixed (Fig. 2.3, Fig. S2.1).

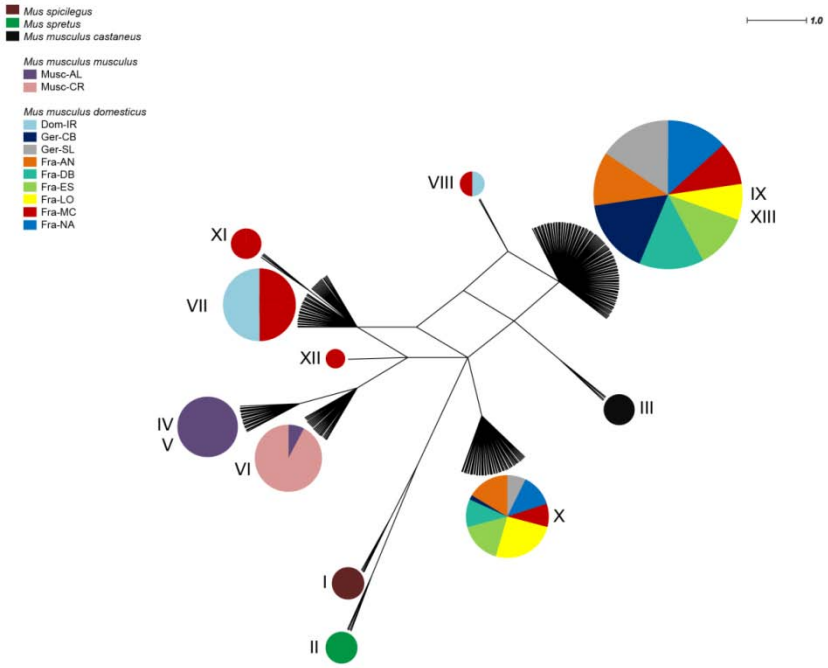


Figure 2.4: Haplotype network of *Xpr1* exon sequences generated with SplitsTree. Each haplotype is indicated by Roman numerals and pie charts indicate which populations carry the respective haplotype.

Variation of P-MLVs in wild-derived mice of different populations

In this part of the study, I aimed to assess possible adaptive effects of the identified allelic variation in *Xpr1* with regard to susceptibility to murine leukemia viruses. In order to do this, I analyzed the variation in the receptor-binding domain (RBD) of

polytropic murine leukemia viruses in wild-derived mice from three distinct populations that exhibited different *Xpr1* haplotypes.

Altogether, I could amplify transcribed P-MLVs in 12 samples originating from Fra_{MC} and six samples each from Ger_{CB} and Dom-IR. I detected 41 sites in the amplified P-MLVs that were variable within or between populations (Tab. S2.5, Fig. S2.2). Phylogenetic analysis of the unphased sequences revealed no separation of samples according to *Xpr1* receptor, but rather according to the population of origin. The samples from Iranian mice were clearly separated from German and French samples. The latter are not clearly separated regardless of the *Xpr1* haplotypes of the respective individuals (Fig. 2.5).

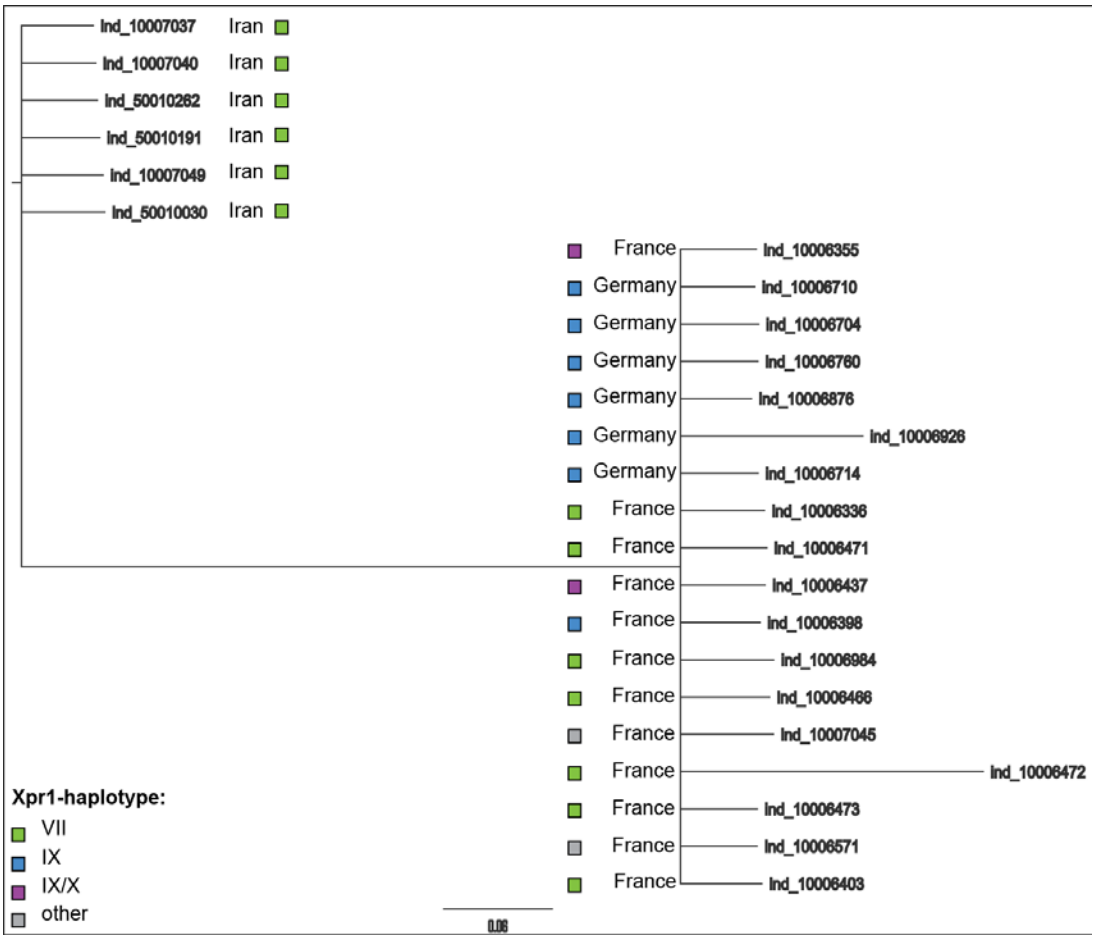


Figure 2.5: Bayesian tree of P-MLV RBD Env sequences of wild-derived mice generated with MrBayes. Population of origin and *Xpr1* haplotype of the respective individuals are indicated next to the samples.

After analyzing the variation of P-MLVs, I tested whether the amplified viruses were infectious in wild-derived mice from our stock or simply inherited proviruses. To be able to test this, I set up crosses between mice from Iran and mice from Germany. The results of this experiment demonstrated that the amplified P-MLV variants were

a combination of the parental variants. Thus, I probably detected no infectious viruses with our approach, but proviruses that are transcribed in the cells (Tab. S2.2).

Discussion

To gain insights into potentially functional variation among *Xpr1* alleles on the level of wild populations of *Mus musculus*, I analyzed DNA samples from 11 geographically distinct regions. Further, I looked for possible effects of this variation on MLVs which mostly depend on *Xpr1* to infect cells. Here our findings are discussed in context of knowledge about previously identified restrictive *Xpr1* alleles, implications due to the applied markers, the population history of our samples and the role of MLV-induced disease in wild mice.

Various *Xpr1* alleles in wild *M. m. domesticus* and *M. m. musulus* populations

I analyzed two of the extracellular loops of *Xpr1* (ECL 3 and 4) which are known to be involved in virus interaction and carry several in this context important residues (Van Hoeven and Miller, 2005; Marin et al., 1999; Tailor et al., 1999; Yan et al., 2007, 2010). Additionally, exon 4 was sequenced because it was found to be variable among inbred strains (Keane et al., 2011). To better assess the variation in *Xpr1* not only across populations, but also across several species, I also sequenced the same regions in two samples of *M. m. castaneus*, *M. spretus* and *M. spicilegus* as outgroups. All included gene regions were variable among our samples and several SNPs and Indels resulted in distinct *Xpr1* haplotypes which varied in abundance and frequency among European *M. musculus* populations.

To date, *Xpr1* alleles were mostly described in a *Mus* species or *Mus musculus* subspecies context and variation of alleles on the population level had not been studied (reviewed in Kozak, 2010; Kozak 2013). All identified *Xpr1* alleles were characterized by several SNPs in ECL3 and 4 and the ones of some lab strains, *M. m. musculus* and *M. m. castaneus* carried unique deletions in ECL4. These deletions occur in most of the to date identified MLV-restrictive *Xpr1* alleles and have been shown to contribute to resistance phenotypes (Marin et al., 1999; Yan et al., 2010). One exception is the $Xpr1^{p}$ variant which was found in *Mus pahari* and represents a full-length receptor, mediating resistance to P-MLVs (Yan et al., 2007). So far analyzed wild-caught *M. m. domesticus* mice from the Americas and Europe were described as having all a full-length *Xpr1* allele which

was called *Xpr1*^{sxv} and is apparently permissive to all XP-MLVs tested (Kozak, 2011; Yan et al., 2009, 2010). Consistent with the previous studies, I found the characteristic deletions in the *M. m. musculus* and *M. m castaneus Xpr1* variants and a full-length receptor in all analyzed *M. m. domesticus* samples. However, with our finer scaled population-based approach, I also found several different *Xpr1* alleles within *M. m. domesticus* and *M. m. musculus* populations that were characterized by various SNPs in exon 4, ECL3 and ECL4.

Most of the SNPs occurred in exon 4 and I cannot infer any functional effects of those SNPs, because they are not directly involved in the interaction between receptor and RBD of the virus. Nevertheless, I cannot exclude the possibility that these SNPs contribute to allele-dependent variation in receptor function and adaptive processes. Additionally, these SNPs can contribute to a comprehensive understanding of the evolution of *Xpr1* and the influence of local selective pressures. In fact, European *Xpr1* haplotypes which mostly differed in exon 4, showed very interesting geographic variation patterns and signatures of balancing selection in some (Fra_{ES}, Fra_{An}, Fra_{LO}) and of positive selection in other populations (Dom-IR, Musc-CR) according to sequence and microsatellite data. This suggests involvement in adaptation processes although differences in the known virus-interaction domain are absent and I do not know what functional effect this variation has. In contrast to the lack of knowledge about implications of the variation in exon 4, variable sites in ECL3 and 4 have been shown to be determinants for virus entry or to modulate virus interaction (Van Hoeven and Miller, 2005; Marin et al., 1999; Yan et al., 2007, 2009, 2010). I identified a nonsynoymous SNP in ECL3 at residue 503 which led to an exchange between the amino acids K and N and was variable within European *M. m. domesticus* samples. This site has not been tested yet for its effect on virus interaction, but interestingly it was shown to be under positive selection in a previous study (Yan et al., 2010) and marks a possible site for N-linked glycosylation. This protein modification is known to modulate virus entry in specific virus-cell combinations and many viruses use glycans on cell-surface glycoproteins as attachment factor (Kozak, 2011; Olofsson and Bergström, 2005; Yan et al., 2008). In the *Xpr1* haplotype which carried the amino acid K at this position, the potential glycosylation site was missing which is interestingly also the case for P-MLV restrictive *Xpr1*^{p} allele of *Mus pahari* (Yan et al., 2007, 2010). Furthermore, in our study this haplotype was found to be under positive selection in the *M. m. domesticus* population Dom-IR and there seemed to be introgression of this haplotype between Dom-IR and Fra_{MC} according to microsatellite data.

Altogether, these findings suggest a possibly important role in virus interaction of this newly identified *M. m. domesticus* allele of *Xpr1* which I termed *Xpr1*^{i}.

Implications due to applied markers and population history

The application of SNPs and microsatellites as markers to detect signatures of natural selection was successful in this study and had been used many times before to detect and characterize candidate regions (Johnsen et al., 2009; Teschke et al., 2008). However, it has to be considered that evolutionary processes can be influenced by population dynamics, geographic distribution of populations and selection. Therefore, also the identified signatures can be influenced by other processes than selection, and it is an important challenge to disentangle the different factors (de Meaux and Mitchell-Olds, 2003). The population structure of the *M. m. domesticus* populations studied here was analyzed in previous studies which helps us to address this issue (Ihle, 2006; Linnenbrink, 2012). The variation pattern of the mitochondrial D-loop region and genome-wide distributed microsatellite markers from the previous studies do not match the allele frequency distribution pattern of *Xpr1*. The population Fra_{MC}, for example, is according to the neutral markers very homogenous compared to other populations, but it exhibits the most *Xpr1* haplotypes. The populations Ger_{CB} and Fra_{NA} as well as Fra_{LO} and Fra_{DB} have a similar genetic background based on microsatellites but different *Xpr1* allele frequency patterns. The same is the case based on D-loop sequences for Fra_{MC} and Ger_{SL}. On the other hand, the *Xpr1* allele frequencies in Fra_{NA} and Fra_{ES} are the same, but the genetic background is different. Altogether, these findings make population structure a less likely explanation for the *Xpr1* haplotype distribution and support the hypothesis that local selective pressures act on this gene.

Besides the population history, also implications due to the marker system have to be considered for the interpretation of observed patterns. Microsatellites can be applied to detect recent positive selection events which represent so called hard selective sweeps. These signatures are caused when individual alleles of neutral markers are dragged towards high frequency by positive selection on a neighboring functional allele and the genetic diversity in the whole region is reduced. However, this only occurs if the functional allele under positive selection is very rare or new in a population and thus linked to only one or a few alleles in the neighboring microsatellite loci. If the advantageous allele was present at intermediate frequencies for a long time before it was subjected to positive selection, it would be linked to various alleles and there would be no clear sweep signature (Charlesworth, 1992; Cutter and Payseur, 2013; Schlötterer, 2003). Furthermore,

microsatellites have a variable but generally fast mutation rate and soon accumulate new mutations which will clear older sweep signatures (reviewed in Ellegren, 2004). Altogether, this means that the signatures of positive selection in our study are recent and caused by new or very rare alleles which occurred in the ancestral Iranian *M. m. domesticus* and the Czech *M. m. musculus* population, respectively.

The role of MLV-induced disease in *Xpr1* evolution

The support for the hypothesis of ongoing adaptive evolution of *Xpr1* across the *Mus* phylogeny and also in wild populations from this and other studies (Staubach et al., 2012; Yan et al., 2010), raises the question which factors are responsible for these processes. The fact that restrictive receptor variants coincide with the acquisition of XP-MLV endogenous retroviruses (ERV) highlights the role of XP-MLVs in receptor evolution. However, data on wild *M. m. domesticus* individuals sampled at scattered locations suggests that these mice only carry inactive P-MLV ERVs and have maintained the full length, permissive *Xpr1* allele (Kozak, 2013). P-MLVs can get involved in disease if they are rescued by ecotropic MLVs which are known to cause disease in wild and lab mice, but use a different receptor (Jung et al., 2003). The rescue involves recombination events which results in the formation of mink cell focus-forming MLVs (MCF). Those viruses often carry a P-MLV RBD and are oncogenic in some lab mice (Hartley et al., 1977; Stoye et al., 1991). Endogenous E-MLVs have not been found in *M. m. domesticus* so far (Kozak, 2013) which means that the formation of MCF-MLVs would depend on re-occurring exogenous infections. These exogenous E-MLV infections have been shown to play a role in some wild mouse communities as for example in mice sampled in California at Lake Casitas (Gardner, 1993; Gardner et al., 1973, 1976). Yet, no recombination between endogenous MLV-DNA and the infectious E-MLVs could be detected so far in wild mice (Barbacid et al., 1979; Rassart et al., 1986). Furthermore, wild mice have been shown to be quite resistant to disease induced by MLVs and if they develop disease, general immunity and fertility seem not to be affected (Gardner et al., 1976; Kozak, 2013). Thus, in contrast to results which suggest that MLVs are an important factor in *Xpr1* evolution, the few results concerning the pathogenicity of XPR1-dependent MLVs in wild mice do not support this (reviewed in Kozak, 2013).

In accordance to this, I did not find infectious viruses with a P-MLV-related RBD. The sequenced RBDs of captive mice from Fra_{MC}, Ger_{CB} and Dom-IR was similar to

the corresponding region of the MCF247 isolate which belongs to the recombinant MCF-MLVs and is leukemogenic in inbred mice (Kelly et al., 1983). The crossing experiment with wild-derived mice from Ger_{CB} and Dom-IR demonstrated that the partially sequenced viruses were not transmitted as infectious particles but as genomic proviruses. Furthermore, the phylogenetic analysis revealed that the separation of the RBD matches the mouse population history. The Iranian mice are ancestral to the European ones, and arrived in the Middle East about 14,000 years ago. The European mice arrived here and split about 3,000 years ago (Cucchi et al., 2005). This is reflected in the RBD-tree with a clear separation of Iranian from German and French mice and incomplete splitting of the latter, regardless of *Xpr1* allele. Overall, these findings suggest that the analyzed viral sequences have remained inactive in the mouse genome for a long time.

In the end, knowledge about the factors which shape the evolution of *Xpr1* in wild mouse populations and the role of XPR1-dependent MLVs in these processes remains scarce. The results presented here show that *Xpr1* is subjected to differing local selective pressures in wild mouse populations which implies that there are functional and adaptively relevant differences among *Xpr1* alleles. MLV-induced disease does not seem to be a sufficient explanation for the complex variation of *Xpr1* alleles. It has also been reported that virus receptors are often down regulated by expression of virus Env glycoproteins, probably to avoid superinfection of cells (Dudley et al., 2011; Wensel et al., 2003) which could have an effect on native receptor function and evolution of it. Interestingly, results from a previous study suggested that *Xpr1* was significantly higher transcribed in brain and testis in mice from the population Fra_{MC} compared to mice from Ger_{CB} (Bryk et al., 2013). This and the recently identified native XPR1 function in phosphate export provide an improved basis for functional analyses. It might be interesting to analyze the different receptor alleles in context with expression levels and phosphate export to gain a more comprehensive picture of XPR1 function and the factors which shape its evolution.

Chapter 2

Characterization of functional allelic variation in Dmrt1 in wild populations of the house mouse

Abstract

Alterations in the activity of transcription factors is most likely a major source of adaptation and detailed analysis of the nature of adaptive changes in these proteins is of great interest in evolutionary biology. The transcription factor *Dmrt1* is a highly conserved member of the sex-determination cascade in vertebrates and for example required for testis differentiation in mice. Notably, in a previous systematic microsatellite screen, a potential positive selection event in this gene was detected in a single wild mouse population from Southern France. In this study, I took advantage of the large amount of data available about *Dmrt1* function to further investigate the apparent adaptive changes in this gene. I aimed to identify the allele which was targeted by positive selection, to assess its geographic distribution across several distinct wild mouse populations and to study its potentially functional effects. Through sequencing analysis of the coding region I identified a nonsynonymous SNP which caused an exchange of the amino acids Asparagine (N) and Serine (S). Interestingly, the N allele showed signs of positive selection in 5 French populations according to microsatellite profiles of six loci within a 200 kb window around *Dmrt1*. That this SNP characterizes the allele targeted by positive selection was also supported by the fact that so far no evidence for differential alternative splicing or differential expression of this gene could be found among populations. A microarray-based screen for differential expression of genes in brain and testis across different postnatal stages, revealed that *Dmrt1* target genes were enriched among the most differently transcribed genes between mice carrying N or S across different postnatal stages. Identified functional links between these differently transcribed target genes were in central nervous system development and spermatogenesis. Further analysis in the future will be required to investigate this functional aspect in more detail and to find out through which effects and mechanisms the presumably advantageous nature of the N allele is mediated.

Introduction

Understanding the evolutionary change of transcription factors (TFs) is of particular interest for evolutionary geneticists as alterations in their activity and regulatory specificity are potentially a major source for phenotypic diversity and adaptation (Bustamante et al., 2005; De et al., 2008; Lopez-Bigas et al., 2008). They constitute proteins that bind to the DNA in a sequence-specific manner (Wunderlich and Mirny, 2009) and enhance or repress gene expression (Latchman, 1997). Many of these factors play a crucial role in the development of organisms and tissue differentiation (de-Leon and Davidson, 2007; Vaquerizas et al., 2009).

Dmrt1 is the most highly conserved member of the sex determination network and has homologues even in flies and worms (Raymond et al., 1999, 2000). It belongs to a transcription factor family with a cysteine-rich zinc DNA binding motif, called DM domain (Erdman and Burtis, 1993; Zhu et al., 2000). The target gene specificity of this TF is probably facilitated by homo- and heterodimerization on the DNA and a set of target genes has recently been identified in mice (Murphy et al., 2007, 2010). Like other members of the DM domain TF family, DMRT1 binds to the minor groove of the DNA at sites potentially overlapping with those of major groove-binding proteins. Thus, it can repress or activate transcription by physical interference with the binding of other TFs or cooperation with them which has been observed for MAB-3 and DSX (Kopp et al., 2000; Lints and Emmons, 2002; Ross et al., 2005; Wang et al., 2011; Williams et al., 2008).

In mice *Dmrt1* shows a complex temporal and tissue-specific expression pattern. It was detected in testis and ovary between embryonic day E9.5 and E15.5 (Hong et al., 2007; Kim et al., 2007; Krentz et al., 2009, 2011; Lei et al., 2007) and in the brain during embryonic development (Ahituv et al., 2007). After birth it reappears in testis and seems to be expressed there throughout postnatal development and adulthood (Hong et al., 2007; Kim et al., 2007; Krentz et al., 2009, 2011; Lei et al., 2007). Male *Dmrt1* knockout mice are sterile (Raymond et al., 2000), while female knockout mice are fertile, but form a reduced number of follicles due to severely decreased Stra8 expression and an abnormal meiotic prophase (Krentz et al., 2011). Extensive functional studies, mostly in the context of developing testis tissue, have demonstrated that *Dmrt1* contributes to sex differentiation rather than determination (Herpin and Schartl, 2011; Hong et al., 2007). Specifically, DMRT1 seemed to be involved in germ cell proliferation, pluripotency and control of mitosis versus meiosis decisions in adult

spermatogonia as well as differentiation of Sertoli cells in neonates (Kim et al., 2007; Krentz et al., 2009; Matson et al., 2010; Raymond et al., 2000). Additionally, recent studies have demonstrated that *Dmrt1* is required to actively maintain male testis identity by suppressing the transcription of female-specific factors like FOXL2 (Matson et al., 2011; Raymond et al., 2000).

Notably, *Dmrt1* was also identified as a candidate target for positive selection in a wild house mouse population from Southern Europe called Fra_{MC} (Büntge, 2010; Staubach et al., 2012). This suggested that this gene is actively involved in adaptation processes and that functionally different alleles of the gene are present.

Altogether, the vast amount of knowledge about *Dmrt1* function and the signature of positive selection in wild mouse populations provide an exceptional good basis to study the nature and effect of potentially functional alleles. In this study, I identified such alleles of *Dmrt1* in wild mice and assessed their geographic variation across several genetically distinct populations together with possibly associated positive selection events. Further, I aimed to describe potential functional effects of the identified alleles on target gene regulation, by breeding mice that carried the different alleles, or were heterozygous for them.

Material and Methods

Mice and Sampling

Population samples

132 DNA samples from 11 distinct wild mouse populations representing two subspecies (*Mus musuclus musculus* and *M. m. domesticus*) were analyzed for the coding sequence variation of *Dmrt1*. The samples were mostly collected by Ihle et al. (2006) and Linnenbrink et al. (2013). Sampling was performed according to a common sampling strategy (described in Ihle et al. (2006)) and the samples used for this study represented distinct sampling sites to ensure that individuals were unrelated (Tab. S1, Fig. 2.1).

Wild-derived mouse samples from breeding stock

Additional to the wild population samples of *M. m. domesticus* and *M. m. musculus,* I collected ear punches of two *Mus spretus*, *Mus spicilegus*, *Apodemus uralensis* and 12 *M. m. castaneus* individuals from the breeding stock for sequencing and *Dmrt1* allele comparison.

After the identification of adaptively relevant *Dmrt1* alleles in wild populations, I also analyzed *Dmrt1* allele-dependent transcription of *Dmrt1* and potential target genes across different developmental stages. Wild-derived mice from the outcrossed breeding stock in Plön were used to generate mouse lines which were homozygous for the respective *Dmrt1* allele in its native population background. Two lines originated from the population Ger_{CB} (N and S) and one from the population Fra_{MC} (MC-N). Apart from *Dmrt1* allele and population of origin, mice from these lines were genetically heterogeneous i.e. originated from six to eight different families and no inbreeding was performed. The mice were kept under the same conditions and for breeding purposes *Dmrt1* alleles were identified by Sanger sequencing samples obtained from ear punches. For the final transcription profile analysis, male mice carrying the *Dmrt1* alleles N, S, SN (heterozygous Ger_{CB}) and MC-N were sacrificed one, five, seven and nine days after birth (P01, P05, P07, P09). Brain and testis tissue was sampled, flash frozen and stored at -80°C for later RNA extraction.

Extraction of nucleic acids

DNA from the wild population samples was extracted by Ihle et al. (2006) or Linnenbrink et al. (2013), respectively. Extraction of DNA from mice from the breeding stock followed a standard salt extraction procedure. DNA stocks were stored at -20°C and diluted to 5 ng/µL prior to PCR.

Total RNA from brain and testis tissue was extracted by means of the NucleoSpin 96 RNA kit (Macherey-Nagel) following the protocol by the manufacturer. The samples were stored at -80°C. RNA quality and content was assessed by repeated NanoDrop (Peqlab) and RNA Nano Chip (Agilent Technologies) measurements.

Molecular analysis of DNA and RNA

Sanger sequencing and sequence analysis

To identify coding sequence variants (here referred to as alleles) of *Dmrt1* occurring in wild mouse populations, I first sequenced its complete coding sequence (Fig. 3.1) in two individuals from each European *M. m. domesticus* population. Afterwards, I focused on exon 2 and sequenced 12 individuals from all available populations and 2 individuals from the above mentioned other species. The applied primer pairs (Metabion) were designed with the software “Primer3Plus” and are listed in Tab. S3.1.

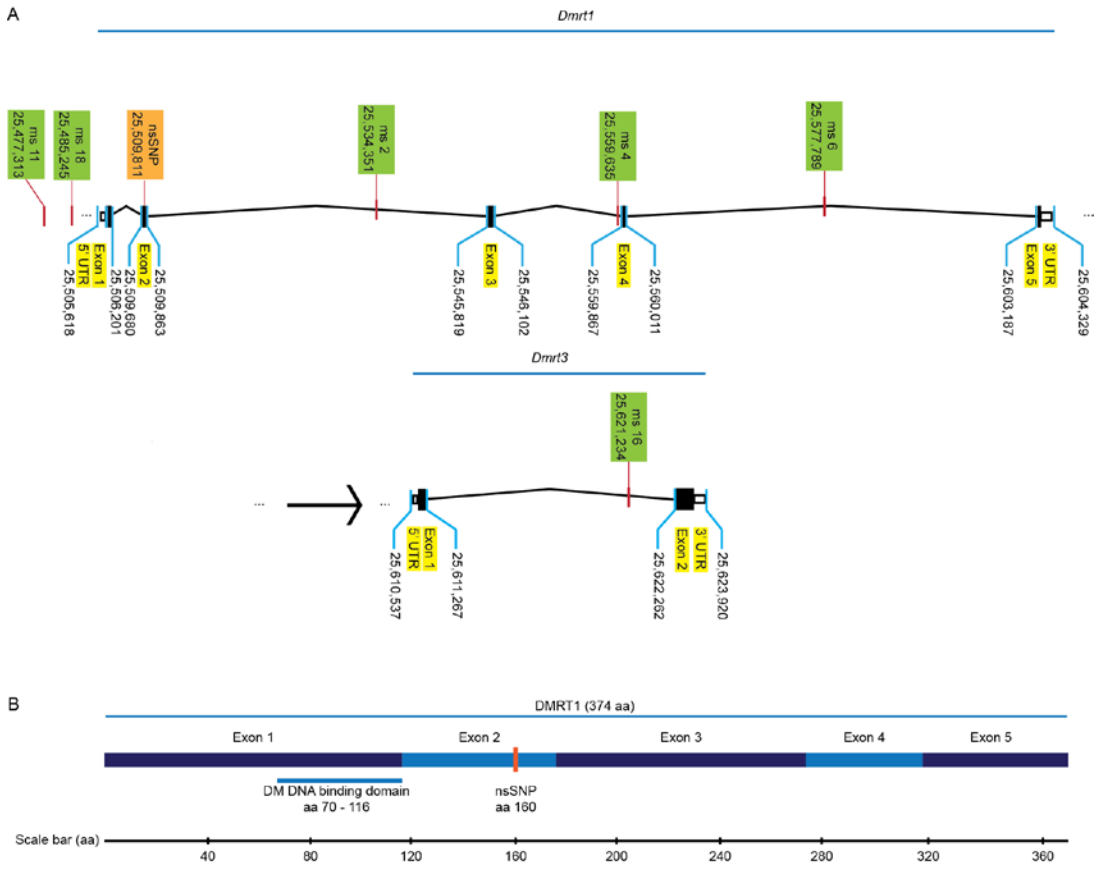


Figure 3.1: Overview of the analyzed candidate gene.
(A) shows the genomic overview of *Dmrt1* on chromosome 19 and its surroundings including the closest neighboring gene *Dmrt3*. The exons and their positions are shown below the genes, analyzed microsatellite loci are indicated by green boxes and the nonsynonymous SNP (nsSNP) by the orange box. The chromosomal positions of the markers are indicated below the label.
(B) shows a sketch of the DMRT1 protein including the DM DNA binding domain as the only functionally characterized region. The nsSNP is indicated by the orange bar. The amino acid (aa) positions of the DM domain and the SNP in the protein are indicated below the labels and by the scale bar.

PCR reactions were carried out in 10 µL final volume with a multiplex PCR kit (Qiagen) under the following cycling conditions: 95°C for 15 minutes followed by 35 cycles of 95°C for 30 s, 60°C for 1.30 min, 72°C for 30 s and 10 min at 70°C elongation time. Exo-Sap purification (USB Corp.) was conducted with the following incubation: 37°C for 20 min and 80°C for 20 min. Cycle sequencing reaction was performed by means of the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Reaction parameters were 96°C for 1 min followed by 29 cycles of 96°C for 10 s, 55°C for 15 s and 60°C for 4 min. Sequencing products were purified with the BigDye XTerminator Purification Kit (Applied Biosystems). Sequences were generated by a 3730 DNA Analyzer (Applied Biosystems). Independent base calling and analysis was done by means of CodonCodeAligner v4.0.2 (CodonCode Corp.).

Microsatellite genotyping

Six microsatellite loci within a region of 200 kb around *Dmrt1* were typed for the European *M. m. domesticus* populations. This window contained another gene of the *Dmrt* family, named *Dmrt3* (Fig. 3.1). The microsatellite loci were identified applying the program “tandem repeats finder” (Benson, 1999) and primers were designed with help of “Primer3Plus” (Untergasser et al., 2007). The forward primers respectively carried a Hex or Fam label at their 5’ end and several primer pairs (Metabion) were pooled for each reaction. The pools contained primer combinations which yielded distinct product size ranges (Tab. S3.2) and only high quality loci were used for further analyses. PCRs were done with the help of a multiplex PCR kit (Qiagen) in 5 µL final volumes with 5 ng DNA as template. Cycle conditions were as follows: 95°C for 15 min followed by 28 cycles at 95°C for 30 s, 60°C for 1.30 min, 72°C for 1.30 min with a final extension at 72°C for 10 min. PCR products were diluted 1:20 in water and 1 µL was denatured in 10 µL Hidi formamide and 0.1 µL 500 Rox size standard (Applied Biosystems). The denaturation step was performed at 90°C for 2 min followed by 20°C for 5 min. Product sizes were determined by a 3730 DNA Analyzer (Applied Biosystems), and allele calling was done by means of the GeneMapper v4.0 software (Applied Biosystems).

Sequence and microsatellite analysis

Frequencies of *Dmrt1* genotypes were calculated for all populations and spatial patterns visualized on a map made by NaturalEarth. For the microsatellite loci, allele frequencies were calculated with regard to *Dmrt1* genotype and population and plotted in histograms. These histograms were visually inspected for potential signatures of positive selection and associations between *Dmrt1* genotypes and microsatellite alleles.

An intrinsic protein disorder prediction was performed by means of the DisEMBL 1.5 (Linding et al., 2003, available from: http://dis.embl.de/cgiDict.py) to assess whether identified polymorphisms lay in regions with a defined or a flexible, disordered secondary structure. For this analysis the whole amino acid sequence of *Dmrt1* was used with the default settings of DisEMBL.

Microarray design and sample preparation

To analyze *Dmrt1* allele-dependent changes in the transcription profile of target genes and co-transcribed genes, I designed custom oligonucleotide microarrays using the Agilent e-array software (https://earray.chem.agilent.com/earray/). A list of

7479 protein coding genes was used as input to generate specific 60-mer oligonucleotide probes. These genes had been found to be either directly regulated by *Dmrt1* (Murphy et al., 2010) or to be co-transcribed during the developmental stages of interest (Su et al., 2004; Zhang et al., 2004). The design also included replication of oligonucleotide probes 8 times for each target and random distribution of these replicates across each of the ordered 60 k expression microarrays, following the procedure recommended in (Pozhitkov et al., 2014).

The total RNA samples were pooled with respect to organ, developmental stage and *Dmrt1* allele. For each of these conditions, samples from 5 individuals were pooled in equal amounts and then labeled using the Agilent Low Input Quick Amp Labeling kit (Agilent). Before loading the arrays with labeled cRNA as described in the Agilent protocol for One-Color Microarray-Based Gene Expression Analysis (v. 6.6, June 2013), a calibration pool containing equal amounts of labeled cRNA from all tested samples was prepared. Then, additional to the sample pools seven arrays were loaded with dilutions of the calibration pool of 0.125, 0.25, 0.5, 1, 2, 4, 4.975 times the recommended amount of cRNA (600 ng). Microarrays were scanned with an Agilent DNA Microarray Scanner type C and feature extraction was performed with the Agilent feature extraction software (v. 10.7.3.1).

Microarray data calibration and analysis

The unprocessed “gMedianSignal” of the obtained microarray data was used for the calibration of the microarrays as described in Pozhitkov et al. (2014). During this calibration procedure, raw intensity values are corrected for probe responsiveness and transformed into measures of relative content of transcript, further referred to as transcript content or level. The calibration was conducted by means of scripts by Alexander Pozhitkov (available from http://web.evolbio.mpg.de/~alexander.pozhitkov/microarray123/). With help of these scripts signal intensities of replicated probes were averaged for each array. Afterwards, the signals were corrected for the probe responsiveness based on parameters that were calculated from calibration arrays for each probe. According to this, poorly responding probes were removed (at the recommended cutoff level of $R^2 < 0.98$). Furthermore, values which lay outside of the calibration range (dynamic range) were treated as missing data. After the calibration, a between array normalization was performed by dividing all values on an array by their 75 percentile as recommended in the “One-Color Microarray-Based Gene Expression Analysis (Low Input Quick Amp Labeling. Protocol (v. 6.5, May 2010) by Agilent).

After calibration and data filtering, I aimed to find signal differences due to different *Dmrt1* alleles. Thus, I focused on differences that occurred between arrays for the lines N and S. The 1% of genes most differently transcribed between S and N was analyzed for *Dmrt1* target gene enrichment relative to genes detected at this stage in the respective organ. This was done by means of Fisher's exact tests calculated in R (R Core Team, 2013). The Bonferroni correction was applied to obtain p-values for multiple testing. Furthermore, I assessed to which extent heterozygote samples showed intermediate transcription levels.

Results

Variation of *Dmrt1* alleles in wild mouse populations

Sequencing of the complete coding sequence of *Dmrt1* of two wild *M. m. domesticus* individuals, each from eight European locations, revealed the presence of 3 synonymous and 1 nonsynonymous SNPs in the second exon at residue 160. The other exons did not show any polymorphism.

The single nonsynonymous SNP caused an exchange of the amino acids Asparagine (N) and Serine (S) and was located at residue 160 outside of the DNA binding domain. The N created a potential motif for N-linked glycosylation. According to the disorder prediction, the region showed an elevated disorder probability relative to the random expectation level. The synonymous SNPs occurred only in single individuals from the populations Fra_{AN} and Ger_{CB} and were not considered for further analyses.

The comparison of the identified alleles across different rodent species revealed that N seemed to be only present in *M. m. musculus* and *M. m. domesticus* whereas S was carried by *Rattus norvegicus* according to NCBI (NP_446158) and all tested *Apodemus uralensis*, *Mus spicilegus* and *Mus spretus* (Fig. 3.2).

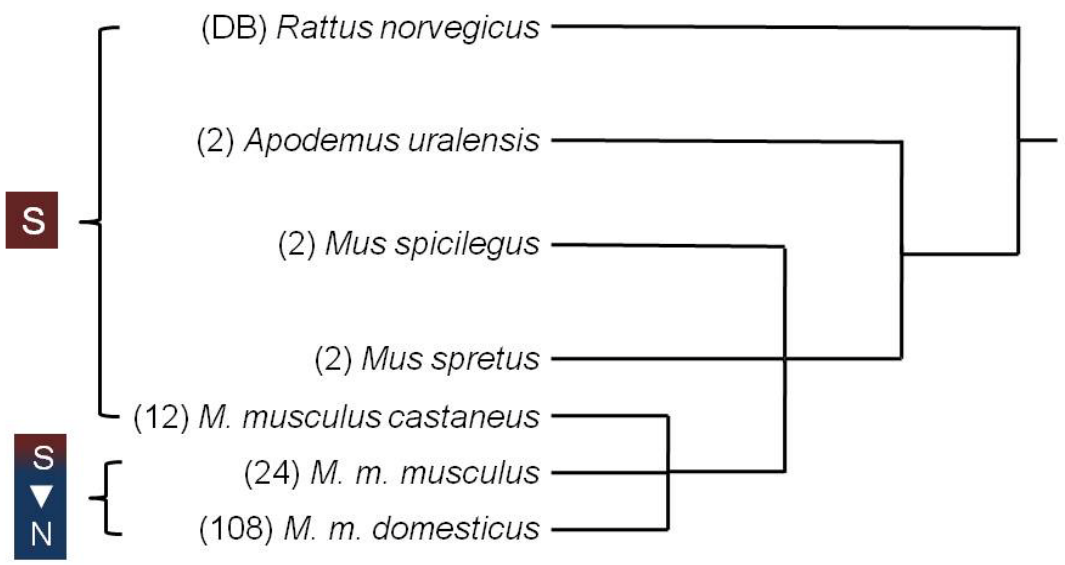


Figure 3.2: Phylogenetic relationships of the species whose *Dmrt1* coding sequence was analyzed. Letters on the left indicate the functional alleles of *Dmrt1* characterized by a single nonsynonymous SNP. S=Serine, N=Asparagine. Numbers in brackets show the number of analyzed individuals, DB= reference genome.

Sequencing of the variable exon in 12 *M. m. musculus* and *M. m. domesticus* individuals per population showed that the S allele is dominant in *M. m. musculus* where the N allele was only found in heterozygote individuals. In the Iranian population, which is ancestral to the European *M. m. domesticus* populations, both alleles were present in homozygous and heterozygous state. The other *M. m. domesticus* populations in Europe exhibited only N homozygously and interestingly this allele was fixed in 4 populations, namely Fra_{MC}, Fra_{DB}, Fra_{ES} and Fra_{LO}. In the other 4 populations heterozygous individuals were present at varying frequencies (Fig. 3.3). Notably, the microsatellite allele frequency data indicated recent positive selection on the N allele in the populations Fra_{MC}, Fra_{AN}, Fra_{DB}, Fra_{ES} and Fra_{LO} (Fig. 3.3, Fig. S3.1). This potential signature of positive selection was consistently indicated by all loci except for ms 16 which was located in the intragenic region of *Dmrt3* (Fig. 3.1, Fig. S3.1).

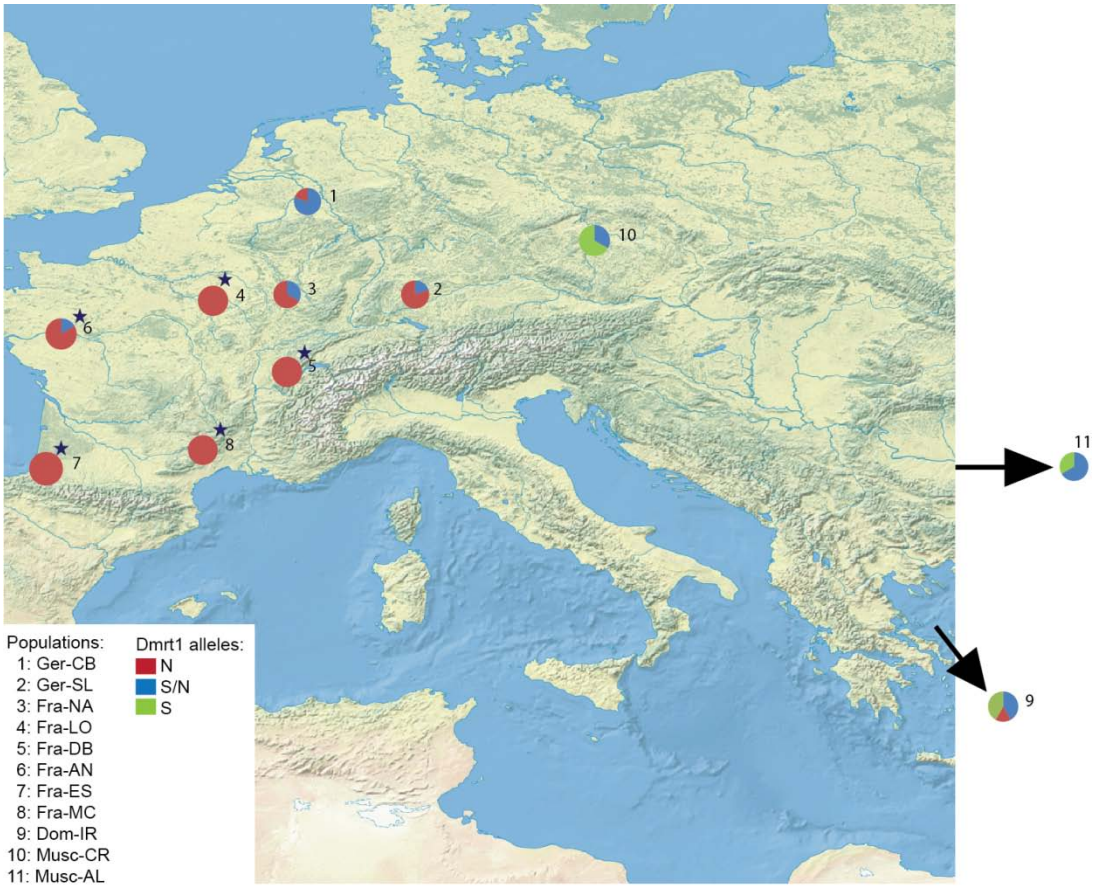


Figure 3.3: Frequencies of the *Dmrt1* genotypes corresponding to the identified nonsynonymous SNP in wild house mouse populations. Musc-CR and Musc-AL belong to the subspecies *Mus musculus musculus*, the other populations to *M. m. domesticus*. Blue stars indicate positive selection signatures.

Functional effects of *Dmrt1* alleles in terms of differential target gene transcription

To assess possible functional effects of the different *Dmrt1* alleles on transcriptional regulation of other genes, we aimed to analyze the transcriptional activity of 7479 protein-coding genes in wild mouse lines carrying either N or S or both alleles (SN).

Of the 7479 probes, 7262 showed an appropriate dose-response behavior and were considered for further analysis. To measure deviations between N and S, I used difference (N minus S) instead of fold change, as high fold changes in our dataset were to a large extent due to low transcription levels with higher uncertainty. High differences on the other hand were mostly due to intermediate or high transcription levels (Fig. 3.4).

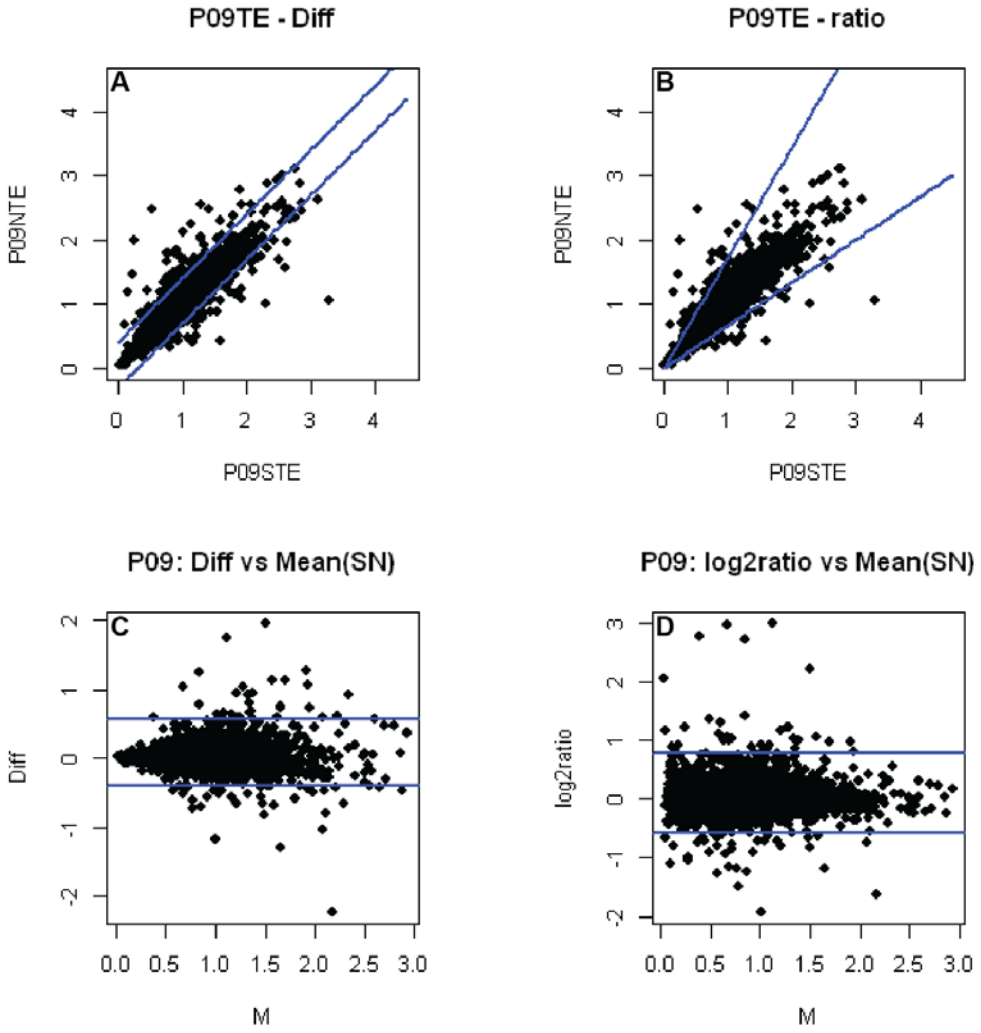


Figure 3.4: Exemplary comparison of difference measures for the microarray data analysis.
In the upper panel calibrated values obtained from animals carrying the *Dmrt1* N allele were plotted against S. The respective functions of the difference thresholds were plotted in blue. In (A), the difference measure corresponded to the actual difference (N-S), while in (B) it corresponded to the log-transformed logarithm (log2(ratio)).
In the lower panel, the difference measures were plotted against the mean of the calibrated values in N and S. The blue lines correspond again to the 1% threshold that was calculated in (A) as difference and in (B) as log-transformed fold change.

All data points above or below the blue lines correspond to the 1% most differently transcribed genes between mice carrying the N or S allele, respectively.

Dmrt1 transcripts were detected in testis tissue at all tested developmental stages (P01, P05, P07, P09) and in the brain one day after birth (P01). After P01, *Dmrt1* was not detected in the brain anymore. Deviations in transcript levels of *Dmrt1* ranged from 0.079 to 0.409 between N and S, 0.017 to 0.235 between MCN and N and 0.059 to 0.328 between MCN and S. Thus, the gene did not belong to the most differently transcribed ones at any tested stage or for any tested allele combination (Tab. S3.3).

I checked for *Dmrt1* target gene enrichment, *Dmrt1*-dependent behavior of heterozygote (SN) individuals and overrepresentation of GO terms among the 1% most differently transcribed genes. This fraction of 1% comprised a set of 72 genes per stage and organ. Significant enrichment of *Dmrt1* target genes among these 1% most differently transcribed genes, was found in brain at P01 and in testis at P05 and P09 when *Dmrt1* was present. At day one and seven (P01, P07) there was no significant enrichment of *Dmrt1* targets but by trend their representation was higher than in the brain samples after day 1. The fraction of *Dmrt1* target genes in SN samples which showed intermediate transcript levels compared to N and S ranged from 43% to 85% across the different stages (Tab. 3.1). Between 12% and 48% of the *Dmrt1* target genes showed higher transcript levels in SN samples compared to N and S, while between 9% and 29% exhibited lower levels. In brain at P01, two GO terms were significantly enriched among the most differently transcribed genes relative to the detected genes at this time and location. These GO terms were central nervous system development and system development (Tab. 3.2). For the other stages and tissues no GO term enrichment was found. In the following, *Dmrt1* target genes that belonged to the 1% most differently transcribed genes and showed in addition intermediate transcription on SN arrays are briefly described for each stage and organ. Complete lists of all most differently transcribed genes between N and S for each stage and tissue can be found in the supplement (Tab. S3.4).

Table 3.1: Overview of test results from microarray analysis. In the first column the arrays are labeled according to stage and organ (BR=brain, TE=testis). The term “candidate” refers to the genes that belonged to the 1% most differently transcribed genes between mice carrying the *Dmrt1*-N or -S allele respectively. “Targets” are known *Dmrt1* target genes which were identified by Murphy et al. 2010. Column 2 to 5 show counts of the genes that served as matrix input for Fisher’s exact test. The corrected significance threshold corresponds to the Bonferroni-adjusted 0.05 threshold.

	Dmrt1-target enrichment							*Dmrt1*-dependence pattern
array	candidate targets	all non-candidate targets detected	non target candidates	all non-candidate non-targets detected	Fisher's exact test (p value)	Corrected significance threshold	signif.	heterozygotes intermediate among candidate targets %
P01BR	23	1224	49	5952	0.002365	0.006	*	52
P01TE	18	1228	54	5853	0.116	0.006		72
P05BR	17	1206	55	5881	0.155	0.006		47
P05TE	25	1218	47	5845	0.0004233	0.006	*	80
P07BR	14	1215	58	5877	0.6366	0.006		85
P07TE	21	1222	51	5841	0.01201	0.006		43
P09BR	19	1208	53	5871	0.04153	0.006		47
P09TE	28	1217	44	5855	1.32E-05	0.006	*	61

.**Table 3.2:** GO terms enriched among the 1% most differently transcribed genes in brain at P01, including the genes which contributed to the enrichment.

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0007417	central nervous system development	5.03E-07	4.70E-03	19.23 (6857,31,69,6)	Cyp26c1 B3gnt5 Lig4 Otx2 Zic1 Zic2 Lig4
GO:0048731	system development	1.70E-06	7.96E-03	4.89 (6857,264,69,13)	Otx2

P01

Among most differently transcribed *Dmrt1* target genes in the brain were *Gnas*, *Bcat1*, *Zic1* and *Bfsp2* which showed a higher transcript abundance in mice carrying N. The genes *Dpysl3*, *Eya1*, *Inhbc*, *Diras1*, *Mmp24*, *Nfe2l3*, *Sap130* and *Pcdhac1* showed a higher transcript abundance in mice carrying S.

In testis the genes *Htr7*, *Bcat2*, *4930429B21Rik*, *H2-T22*, *Tlr9*, *Raet1d* and *Krtap31-1* exhibited a higher transcript abundance on N arrays. In contrast, transcripts of *Gckr*, *Slc9a*, *9530002B09Rik*, *Actl11*, *Pde4a* and *Spp1* seemed to be more abundant in animals carrying S.

P05

In brain, the most differently transcribed *Dmrt1* targets that showed intermediate transcription in mice carrying SN and higher abundance with the N allele were *Osm*, *Kcna6*, *Msr1*, *Clca3* and *Mapk4*. Higher transcript abundance in mice carrying S was exhibited by *Cdk19*, *Tcp11* and *Tmco4*. In testis, the genes *Tmco4*, *Olfr958*, *Ermn*, *A430078G23Rik*, *H2-T22*, *Cst9*, *Bcat2*, *Prokr1*, *Ifna13* and *Rab19* showed higher transcript abundance in samples carrying N. In contrast, transcripts of *Xrcc6bp1*, *Wtap*, *Sfrp2*, *Chrna1*, *Sohlh1*, *Heatr5a*, *Cyct*, *Stra8*, *Fndc3c1* and *AA467197* were more abundant in mice carrying S.

P07

In the brain samples exhibiting the allele N, *Il9*, *AI597479*, *Tbx4* and *Ralgps1* exhibited higher transcript abundance. In mice carrying S, *Bfsp*, *Pcdhb9*, *Got1*, *Cwc22*, *Cartpt*, *Tas2r102*, *Ankrd28* and *Olfr958* were transcribed more abundantly.

In testis of mice carrying N, transcripts of *Slc25a44* and *Emb* were more abundant. *Gpbp1l1*, *Aard*, *Lonrf3*, *Hist1h1a*, *Col4a3*, *Trim52* and *Stra8* showed higher transcript abundance in samples carrying S.

P09

Among the most differently transcribed *Dmrt1* targets in brain were *Qk*, *9530002B09Rik*, *Pear1*, *Gp1ba*, *4930547C10Rik* and *Aqp6* which were more abundant in mice carrying N. In samples exhibiting the allele S, among the most abundant transcripts were *Wisp2*, *Drd1a* and *Tmem161*.

In testis of mice carrying the allele N, transcripts of *Myh1*, *Sfrp2*, *F2rl1*, *Siglec1*, *Ceacam20*, *Chrna1*, *Clcn1* and *Heatr5a* were more abundant. In mice exhibiting S, more abundant transcripts belonged to *Mapk4*, *Efcab11*, *Tmem161b*, *Gbp11*, *Mapkapk3*, *Spink12*, *Hist1h1a*, *Hist1h1t* and *Cst9*.

I found little overlap in the 1% of most differently transcribed genes for both organs between the different stages (Fig. 3.5) and the shared genes are listed in the supplement (Tab. S3.5).

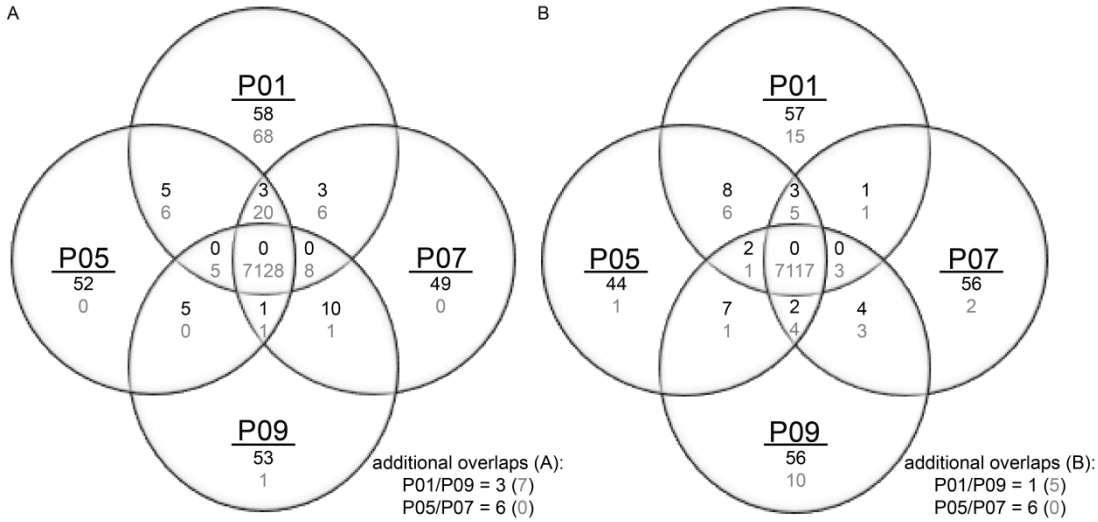


Figure 3.5: Venn-Diagrams showing the overlap of detected genes in brain (A) and testis (B) across the tested postnatal stages. The black numbers represent the 1% most differently transcribed genes between N and S. The grey numbers represent the overlaps of all detected genes.

Discussion

The previous detection of *Dmrt1* as candidate gene for a positive selection event in the wild mouse population Fra_{MC} (Büntge, 2010) suggested that there are different

functional alleles of this transcription factor. In this study, I successfully identified potential functional, adaptively relevant coding sequence variation in this gene, and signatures of positive selection in several additional wild mouse populations. Finally, I was able to provide first insights into functional and potentially adaptive effects of the different *Dmrt1* alleles on target gene regulation.

Coding sequence variation of *Dmrt1* and multiple positive selection events

Functionally different alleles can be generated by diverse mechanisms like protein-coding changes, allele-specific alternative splicing which is known to contribute to variation in protein function (Hull et al., 2007; Kwan et al., 2007; Lin et al., 2010; Nembaware et al., 2004) or differential expression of genes caused by regulatory changes (Carroll et al., 2008; Hoekstra and Coyne, 2007; Khaitovich et al., 2005; Staubach et al., 2010). Thus, there are many different possible ways for positive selection to act on allelic variation.

Dmrt1 has previously been reported to be alternatively spliced during development and in adults (Lu et al., 2007). However, in the here studied adult mice from the populations Fra_{MC} and Ger_{CB} no alternative splicing could be detected in RNA-Seq data of 8 individuals per population (Rafik Neme, personal communication). Possible significant differential expression of *Dmrt1* between Fra_{MC} and Ger_{CB} was also not found in adult mice (Bryk et al., 2013). Additionally, it was included in this study here, but neither ranked among the most differently transcribed genes when comparing the identified coding sequence variants nor when comparing the populations regardless of the coding sequence variants. Hence, it appears unlikely that splicing variants or expression differences have caused the putative sweep signature.

I was able to identify a single nonsynonymous SNP in the *M. m. musculus* and *M. m. domesticus* populations which caused an amino acid change from Serine (S) to Asparagine (N). The apparent prevalence of S in the analyzed rodents suggested that this variant is the ancestral and N the derived state. Furthermore, I observed an interesting geographical pattern of allele frequency distribution with a frequency shift between S and N when comparing the two subspecies. In the Eastern European populations only S was present in the homozygous state and N only in heterozygous individuals. This pattern was reversed in the Western European populations, though an Iranian population ancestral to those exhibited both alleles in the homozygous state and heterozygous genotypes. This allele distribution pattern and the intriguing fact that the N allele was associated to signatures of

positive selection in neighboring microsatellite loci in several populations indicates that there is in fact a functional difference between these two coding sequence alleles.

I detected signatures of positive selection by applying microsatellite markers which have been used many times to successfully identify recent positive selection events and to characterize the respective regions in more detail (Johnsen et al., 2009; Teschke et al., 2008). These studies infer positive selection events mainly by applying statistics like the lnRH-calculation (Kauer et al., 2003) which relies on comparing the variability of one microsatellite in two populations with a large genome-wide reference set of neutrally evolving microsatellites. A positive signal is only obtained if the variability is reduced in one population compared to the other one and if this reduction deviates significantly from the neutral expectations. This approach accounts for (1) the marked differences in microsatellite mutation rates which makes it difficult to compare different loci and (2) for demographic effects that can cause signatures similar to positive selection events. In this study, I relied on inferring signatures of positive selection from the microsatellite allele frequency distribution patterns of 6 loci located within a 200 kb window including *Dmrt1* and the neighboring gene *Dmrt3*. This approach was already used by Teschke et al. (2008), complementary to the lnRH-statistics, who demonstrated that the frequency distributions consistently reflected the signatures detected by the statistics. Considering these previous results, I expected to find one microsatellite allele per locus in high frequency due to its linkage with the functional allele which was presumably targeted by positive selection. The occurrence of several additional alleles in low frequency and mostly gathered around the high frequency allele can also be expected as microsatellites have a high mutation rate and fast accumulate new mutations in a stepwise manner (reviewed in Ellegren, 2004). This high mutation rate is also the reason why microsatellite markers are very useful to detect recent positive selection events on new or rare alleles which cause signatures called “hard selective sweeps” (reviewed in Cutter and Payseur, 2013). These signatures come about, if positive selection acts on a functional allele which is linked to few other alleles in the neighboring markers. Positive selection would drive the functional variant to high frequency as well as the few other linked alleles which results in a well detectable loss of genetic diversity in this region. The N allele is not a new allele in the Western European house mouse populations as it also segregates in the ancestral Iranian population as well as in the *M. m. musculus* populations. Consequently, the fact that I detected a signature of positive selection

by microsatellites, gives evidence that it was recently introduced to the respective populations in low frequency, before it got targeted by positive selection and spread quickly to relative high frequency.

I observed such a pattern of positive selection in the five microsatellite loci in the French populations except for Fra_{NA}. The latter population as well as Ger_{CB} and Ger_{SL} exhibited different allele distribution patterns with several alleles being represented in more balanced frequencies compared to the other populations. Thus, apparently no positive selection event on *Dmrt1* happened in these populations which dragged the neighboring microsatellite alleles into high frequency. This is supported by the fact that these populations also exhibited a relatively high amount of heterozygous individuals and that the N allele is not close to fixation as could be expected under positive selection.

Apart from positive selection evolutionary processes and their genomic signatures are also influenced by other factors than selection like population dynamics and geographical distribution of populations (de Meaux and Mitchell-Olds, 2003). For example demographic events like population bottlenecks can result in reduced genetic diversity which could be misinterpreted as signature of positive selection when only looking at one particular genomic region (reviewed in Hughes, 2008) for review). An approach to distinguish these effects is to compare the genome-wide to the local variability, as demographic events have genome-wide effects while positive selection acts locally (Stajich and Hahn, 2005). This comparison is implemented in the lnRH-statistics (Kauer et al., 2003) which was used in a previous study that identified *Dmrt1* as a candidate target for positive selection in Fra_{MC} (Büntge, 2010).

In addition, previous studies on here studied populations have addressed questions concerning their population history and structure. The population history was very well documented based on fossil records (Cucchi et al., 2005), the genome-wide variability was assessed (Baines and Harr, 2007; Ihle et al., 2006) and the population structure had been investigated by means of D-loop sequences and genome-wide distributed microsatellites (Bonhomme et al., 2011; Linnenbrink, 2012). Altogether, this data provides a good basis to discuss which forces apart from positive selection might have shaped the observed allelic variation of *Dmrt1*.

Cucchi et al. (2005) documented a massive invasion of Europe by house mice and a split into several populations about 3,000 years ago based on fossil records.

Strong subsequent bottlenecks at least in the Central European house mice are unlikely as the European populations showed a slightly reduced, but still relatively high genome-wide diversity when comparing ancestral and derived populations (Baines and Harr, 2007; Teschke et al., 2008). Furthermore, the population structure inferred from D-loop sequences and genome-wide distributed microsatellite data (Linnenbrink, 2012), did also not match the genotype variation pattern observed for *Dmrt1* in this study. There are populations which exhibit a similar genetic background, but different *Dmrt1* genotype compositions, like Fra_{MC} and Ger_{SL} for the D-loop and Fra_{LO} and Fra_{NA} for the microsatellite data (Linnenbrink, 2012). On the other hand, other populations have a different genetic background, but similar *Dmrt1* genotype distributions, like Fra_{ES} and Fra_{MC} for D-loop sequences and microsatellites (Linnenbrink, 2012).

Altogether, this indicates that population-specific positive selection rather than population structure or severe demographic events have shaped the variation pattern of *Dmrt1* in wild European house mouse populations.

Potential functional effects of *Dmrt1* alleles in terms of transcriptional regulation

Given the evidence of positive selection in the N allele, I was interested in possible functional differences between N and S alleles. Thus, I focused on testing whether known or potential target genes of DMRT1 were differentially transcribed in a *Dmrt1* allele-specific manner in mice of different ages sharing the same population background.

In order to do this, I mainly focused on the analysis of wild-derived mice from Ger_{CB} exhibiting N, S or SN *Dmrt1* genotypes. I conducted microarrays and quantified the transcription of 7262 genes of which 1256 were previously identified targets of the transcription factor in nine day old mice (Murphy et al., 2010). An underlying assumption of microarray-based expression analysis is that there is a one-to-one relationship between detected probe response and transcript abundance. However, previous studies have demonstrated that the binding behavior of the surface bound probes can vary considerably such that the signal intensities do not reflect the actual transcript abundance (Pozhitkov et al., 2007). That is why I here applied a recently developed calibration method which enabled me to account for the individual probe responsiveness i.e. to identify probes producing a reliable signal and to correct for the probe-specific binding behavior (Pozhitkov et al., 2014). By using this approach I obtained reliable transcription activity data that I then used to get first insights into potential *Dmrt1* allele-dependent regulatory changes.

Our approach with pooled individuals was meant to screen several developmental stages for differentially regulated target genes. As a consequence of pooling, it is limited in the sense that I could not test for statistically significant differences in transcript levels. However, assuming that especially direct DMRT1 targets would be influenced by different alleles, I first tested for an enrichment of targets among the 1% most differently transcribed genes between N and S. This served as an indirect test of the influence of the transcription factor variants, which I notably found in the tissues where *Dmrt1* was active (including several testis samples and one brain sample). In testis the enrichment of targets was by trend generally higher than in the brain, except for P01, with two significant values for P05 and P09 after very stringent correction for multiple testing. This repeated target gene enrichment at different stages when *Dmrt1* transcripts were present, and its absence when none were detectable, strongly supported the hypothesis that the two coding sequence alleles of this gene are functionally different.

As mentioned above, functional effects on target gene transcription due to differences in transcript abundance of *Dmrt1* itself, did not seem to be likely at the tested stages. Also the fact that *Dmrt1* transcription was detected in the brain at P01 only in mice carrying the N allele, was probably rather a result of the post calibration filtering than of differential transcription. The *Dmrt1* transcript abundance at this day was treated as unknown for all alleles except for N as the detected intensity values were outside of the dynamic range for the calibration and thus not to be determined reliably. However, they seemed to be only slightly lower than the value measured for the N allele i.e. there was no big difference in the transcript abundance between the different alleles. Thus, I continued to focus on the possible functional role of the identified coding sequence alleles.

It has long been under discussion whether transcription factor evolution by coding sequence change is likely and how such changes could influence the transcription factor function. Coding changes were thought to be generally deleterious as they would result in the alteration of the regulation of too many downstream genes and most studies so far focused on changes in cis-regulatory elements (Carroll, 2000; Tautz, 2000). However, recent experimental evidence has changed this view by demonstrating that coding changes in TFs can also only alter the transcription of a subset of target genes without disrupting the entire downstream regulation (reviewed in Hsia and McGinnis, 2003). One famous example for this is the transcription factor FOXP2 which is highly conserved and relevant to the development of language in humans (Lai et al., 2001). After the divergence of the

chimpanzee and human lineage, a two amino acid sequence variation was fixed and maintained in the human lineage which might have contributed to the acquisition of speech and language (Enard et al., 2002). The two SNPs are located outside of the DNA binding domain. An exchange of Threonine and Asparagine apparently caused a minor change in the secondary structure, while an exchange of Asparagine and Serine generated a potential target site for phosphorylation (Enard et al., 2002).

Notably, the here detected substitution of S by N creates a potential site for N-linked glycosylation. Generally, glycosylation is a common posttranslational modification that is known to influence for example protein folding and stability as well as signaling processes (Alberts et al., 2008). Studies on glycosylation of transcription factors have shown that it can have effects on DNA binding, co-activator binding and transcriptional activation (Ahmad et al., 2006; Chan et al., 2010; Gewinner et al., 2004; Jackson and Tjian, 1988). Besides the generation of the possible modification motif, the only apparent difference between N and S seems to be the size. The latter belongs to the smallest amino acids while N carries a larger side chain which could influence the secondary structure of the protein (Alberts et al., 2008; Haubold and Wiehe, 2006). The biochemical conditions in the concerned region are probably not altered as both amino acids belong to the uncharged polar group.

Assessment of the potential effect of the observed exchange on gene function is complicated by the fact that the altered DMRT1 domain is not functionally characterized. However, I found that the SNP was situated in a region of increased disorder probability. This feature might also explain why this region has not been further characterized so far, as such stretches often lead to difficulties in protein expression, purification and crystallization (Linding et al., 2003). Disordered regions have been shown to be overrepresented in TFs and seem to be often involved in DNA binding and protein-protein interactions (Liu et al., 2006). They seem to be functionally very important as they provide an intrinsic plasticity that enables a single protein to recognize and bind many biological targets with high specificity (Dunker et al., 2001; Dyson and Wright, 2002, 2005; Gunasekaran et al., 2003; Meador et al., 1992; Wright and Dyson, 1999) and have the potential to form large interaction surfaces (Dunker et al., 2001; Gunasekaran et al., 2003; Meador et al., 1992). It is conceivable that an exchange of amino acids of different sizes or even the generation of a glycosylation site might have an influence on the flexibility of this region and thus on the interactions with other proteins.

Nevertheless, the question how exactly the functional difference between the N and S allele is mediated remains still open.

After concluding that the two identified *Dmrt1* alleles are in fact functionally different and that such a difference due to the observed coding sequence change is principally possible, I was interested in the actual functional effects of the two alleles. Thus, to get an impression of potential functional links between genes whose transcription was most affected by differing *Dmrt1* alleles, I tested for GO term enrichment among the most differently transcribed genes. Though the set of genes chosen to run this test was very limited (72 genes per stage), I nonetheless found an overrepresentation of two overlapping GO terms at P01 in the brain. These terms were “system development” and more specifically “central nervous system development”. For other stages no significant enrichment was found, which could be due to the small set of genes used as input. Interestingly, in testis at P09 there was by trend an enrichment of differently transcribed genes specifically involved in spermatogenesis.

There was little overlap between the most differently transcribed genes as well as between by trend enriched GO terms between all analyzed stages. Possible reasons for this could be again the very small fraction of genes I looked at and the tested stages which altogether represented a time frame of nine days. During this time, DMRT1 apparently contributes critically to germ cell differentiation, Sertoli cell maturation, mitosis initiation and spatial organization of Sertoli cells as well as germ cells. I chose stages for the analysis at which important developmental changes occur and which are affected when DMRT1 is knocked out (Kim et al., 2007; Lei et al., 2007; Murphy et al., 2010). At P01 *Dmrt1* transcripts reappear the first time after its expression was switched off in the developing testis during embryogenesis. It has been shown that the mitosis initiation in germ cells is closely correlated with DMRT1 acquisition at this stage (Kim et al., 2007; Lei et al., 2007). P05 represents the initial period of spermatogenesis (Kim et al., 2007). At P07 germ cells start to migrate and DMRT1 seems to be required in germ cells for migration to the tubule periphery as well as in Sertoli cells for the proper development of germ cells (Kim et al., 2007; Lei et al., 2007). Testis at P09 were sampled for the identification of DMRT1 target genes and at this stage DMRT1 seems to be involved in the ring shaped organization of Sertoli cells (Kim et al., 2007; Murphy et al., 2010). Altogether, DMRT1 involvement in these various processes over time, probably associated with the interaction of numerous proteins might explain the little overlap of differentially regulated targets across stages.

Altogether, the data from this study suggests that the N and S allele are functionally different and that this divergence can influence certain developmental processes. However, so far I do not know through which mechanisms this functional difference is mediated. Furthermore, the comparisons only refer to allele function within the same population background. The included samples from the population Fra_{MC} where N was fixed presumably due to positive selection highlighted that the functional, allele-specific effects depend on the (population)genetic background. Thus, I cannot directly draw conclusions about the advantageous nature of N that led to its fixation in the French populations. Future analyses, focused on the background-dependent function of the different *Dmrt1* alleles will be necessary to highlight the potentially beneficial effects of the N allele in the French population.

Bibliography

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L.A., and Rubin, E.M. (2007). Deletion of ultraconserved elements yields viable mice. PLoS Biol *5*, e234.

Ahmad, I., Hoessli, D.C., Walker-Nasir, E., Rafik, S.M., Shakoori, A.R., and Nasir-ud-Din (2006). Oct-2 DNA binding transcription factor: functional consequences of phosphorylation and glycosylation. Nucleic Acids Res. *34*, 175–184.

Alberts, B., Wilson, J., and Johnson, A. (2008). Molecular biology of the cell: (New York, NY [u.a.]: Garland Science).

Pozhitkov, A.E., Noble, P.A., Bryk, J., and Tautz, D (2013). A new procedure for microarray experiments to account for experimental noise and the uncertainty of probe response. PloS one 9, e91295.

Andolfatto, P., and Przeworski, M. (2001). Regions of lower crossing over harbor more rare variants in African populations of Drosophila melanogaster. Genetics *158*, 657–665.

Andrés, A.M. (2001). Balancing selection in the human genome. In eLS, (John Wiley & Sons, Ltd),.

Baines, J.F., and Harr, B. (2007). Reduced X-linked diversity in derived populations of house mice. Genetics *175*, 1911–1921.

Barbacid, M., Robbins, K.C., and Aaronson, S.A. (1979). Wild mouse RNA tumor viruses. A nongenetically transmitted virus group closely related to exogenous leukemia viruses of laboratory mouse strains. J. Exp. Med. *149*, 254–266.

Barton, N.H. (2000). Genetic hitchhiking. Philos. Trans. R. Soc. B Biol. Sci. *355*, 1553–1562.

Battini, J.L., Heard, J.M., and Danos, O. (1992). Receptor choice determinants in the envelope glycoproteins of amphotropic, xenotropic, and polytropic murine leukemia viruses. J. Virol. *66*, 1468–1475.

Battini, J.-L., Rasko, J.E.J., and Miller, A.D. (1999). A human cell-surface receptor for xenotropic and polytropic murine leukemia viruses: Possible role in G protein-coupled signal transduction. Proc. Natl. Acad. Sci. *96*, 1385–1390.

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

Bonhomme, F., Orth, A., Cucchi, T., Rajabi-Maham, H., Catalan, J., Boursot, P., Auffray, J.-C., and Britton-Davidian, J. (2011). Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization. Proc. R. Soc. B Biol. Sci. *278*, 1034–1043.

Boursot, P., Auffray, J.C., Britton-Davidian, J., and Bonhomme, F. (1993). The evolution of house mice. Annu. Rev. Ecol. Syst. *24*, 119–152.

Boursot, P., Din, W., Anand, R., Darviche, D., Dod, B., Von Deimling, F., Talwar, G.P., and Bonhomme, F. (1996). Origin and radiation of the house mouse: mitochondrial DNA phylogeny. J. Evol. Biol. *9*, 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.

Bromham, L. (2008). Reading the story in DNA: A beginner's guide to molecular evolution (Oxford□; New York: Oxford University Press, USA).

Bryk, J., Somel, M., Lorenc, A., and Teschke, M. (2013). Early gene expression divergence between allopatric populations of the house mouse (Mus musculus domesticus). Ecol. Evol. *3*, 558–568.

Büntge, A. (2010). Tracing signatures of positive selection in natural populations of the house mouse.

Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Todd Hubisz, M., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., et al. (2005). Natural selection on protein-coding genes in the human genome. Nature *437*, 1153–1157.

Carroll, S.B. (2000). Endless forms: the evolution of gene regulation and morphological diversity. Cell *101*, 577–580.

Carroll, S.B. (2008). Evo-Devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. Cell *134*, 25–36.

Chan, C.-P., Mak, T.-Y., Chin, K.-T., Ng, I.O.-L., and Jin, D.-Y. (2010). N-linked glycosylation is required for optimal proteolytic activation of membrane-bound transcription factor CREB-H. J. Cell Sci. *123*, 1438–1448.

Charlesworth, B. (1992). New genes sweep clean. Nature *356*, 475–476.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. *10*, 195–205.

Charlesworth, B., Morgan, M.T., and Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. Genetics *134*, 1289–1303.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. Genome Res. *20*, 393–402.

Chesebro, B., and Wehrly, K. (1985). Different murine cell lines manifest unique patterns of interference to superinfection by murine leukemia viruses. Virology *141*, 119–129.

Cloyd, M.W., Hartley, J.W., and Rowe, W.P. (1980). Lymphomagenicity of recombinant mink cell focus-inducing murine leukemia viruses. J. Exp. Med. *151*, 542–552.

Cloyd, M.W., Thompson, M.M., and Hartley, J.W. (1985). Host range of mink cell focus-inducing viruses. Virology *140*, 239–248.

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. Biol. J. Linn. Soc. *84*, 429–445.

Cutter, A.D., and Payseur, B.A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. *14*, 262–274.

De, S., Lopez-Bigas, N., and Teichmann, S.A. (2008). Patterns of evolutionary constraints on genes in humans. BMC Evol. Biol. *8*, 275.

Dudley, J.P., Mertz, J.A., Bhadra, S., Palmarini, M., and Kozak, C.A. (2011). Endogenous retroviruses and cancer. In retroviruses and insights into cancer, J. Dudley, ed. (Springer New York), pp. 119–162.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., et al. (2001). Intrinsically disordered protein. J. Mol. Graph. Model. *19*, 26–59.

Dyson, H.J., and Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. Curr. Opin. Struct. Biol. *12*, 54–60.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. *6*, 197–208.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. *5*, 435–445.

Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S.L., Wiebe, V., Kitano, T., Monaco, A.P., and Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. Nature *418*, 869–872.

Erdman, S.E., and Burtis, K.C. (1993). The Drosophila doublesex proteins share a novel zinc finger related DNA binding domain. EMBO J. *12*, 527–535.

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

Fischinger, P.J., Nomura, S., and Bolognesi, D.P. (1975). A novel murine oncornavirus with dual eco- and xenotropic properties. Proc. Natl. Acad. Sci. U. S. A. *72*, 5150–5155.

Gardner, M.B., Officer, J.E., Rongey, R.W., Charman, H.P., Hartley, J.W., Estes, J.D., and Huebner, R.J. (1973). C-type RNA tumor virus in wild house mice (Mus musculus). Bibl. Haematol. *39*, 335–344.

Gardner, M.B., Klement, V., Rongey, R.R., McConahey, P., Estes, J.D., and Huebner, R.J. (1976). Type C virus expression in lymphoma-paralysis-prone wild mice. J. Natl. Cancer Inst. *57*, 585–590.

Gardner, M. (1993). Genetic-control of retroviral disease in aging wild mice. Genetica *91*, 199–209.

Gewinner, C., Hart, G., Zachara, N., Cole, R., Beisenherz-Huss, C., and Groner, B. (2004). The coactivator of transcription CREB-binding protein interacts preferentially with the glycosylated form of Stat5. J. Biol. Chem. *279*, 3563–3572.

Giovannini, D., Touhami, J., Charnet, P., Sitbon, M., and Battini, J.-L. (2013). Inorganic phosphate export by the retrovirus receptor XPR1 in Metazoans. Cell Rep. *3*, 1866–1873.

Guénet, J.L., and Bonhomme, F. (2003). Wild mice: an ever-increasing contribution to a popular mammalian model. Trends Genet. TIG *19*, 24–31.

Gunasekaran, K., Tsai, C.-J., Kumar, S., Zanuy, D., and Nussinov, R. (2003). Extended disordered proteins: targeting function with less scaffold. Trends Biochem. Sci. *28*, 81–85.

Hamilton, M. (2009). Population genetics (Chichester, UK; Hoboken, NJ: John Wiley & Sons).

Harr, B., Kauer, M., and Schlötterer, C. (2002). Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in Drosophila melanogaster. Proc. Natl. Acad. Sci. *99*, 12949–12954.

Hartl, D.L., and Clark, A.G. (2006). Principles of population genetics, Fourth Edition (Sunderland, Mass: Sinauer Associates, Inc.).

Hartley, J.W., Wolford, N.K., Old, L.J., and Rowe, W.P. (1977). A new class of murine leukemia virus associated with development of spontaneous lymphomas. Proc. Natl. Acad. Sci. *74*, 789–792.

Haubold, B., and Wiehe, T. (2006). Introduction to computational biology: an evolutionary approach (Basel [u.a.]: Birkhäuser).

Hedrick, P.W. (2007). Balancing selection. Curr. Biol. *17*, R230–R231.

Herpin, A., and Schartl, M. (2011). Sex determination: switch and suppress. Curr. Biol. *21*, R656–R659.

Hoekstra, H.E., and Coyne, J.A. (2007). The locus of evolution: evo devo and the genetics of adaptation. Evolution *61*, 995–1016.

Hoekstra, H.E., Hirschmann, R.J., Bundey, R.A., Insel, P.A., and Crossland, J.P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. Science *313*, 101–104.

Van Hoeven, N.S., and Miller, A.D. (2005). Use of different but overlapping determinants in a retrovirus receptor accounts for non-reciprocal interference between xenotropic and polytropic murine leukemia viruses. Retrovirology *2*, 76.

Hong, C.-S., Park, B.-Y., and Saint-Jeannet, J.-P. (2007). The function of Dmrt genes in vertebrate development: It is not just about sex. Dev. Biol. *310*, 1–9.

Hsia, C.C., and McGinnis, W. (2003). Evolution of transcription factor function. Curr. Opin. Genet. Dev. *13*, 199–206.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics *17*, 754–755.

Hughes, A.L. (2008). Near neutrality. Ann. N. Y. Acad. Sci. *1133*, 162–179.

Hull, J., Campino, S., Rowlands, K., Chan, M.-S., Copley, R.R., Taylor, M.S., Rockett, K., Elvidge, G., Keating, B., Knight, J., et al. (2007). Identification of common genetic variation that modulates alternative splicing. PLoS Genet *3*, e99.

Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. *23*, 254–267.

Ihle, S. Ravaoarimanana, I., Thomas, M., Tautz, D. (2006). An analysis of signatures of selective sweeps in natural populations of the house mouse. Mol. Biol. Evol. *23*, 790–797.

Jackson, S.P., and Tjian, R. (1988). O-glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. Cell *55*, 125–133.

Jobling, M.A., Hurles, M., and Tyler-Smith, C. (2004). Human evolutionary genetics: origins, peoples & disease (New York: Garland Science).

Johnsen, J.M., Teschke, M., Pavlidis, P., McGee, B.M., Tautz, D., Ginsburg, D., and Baines, J.F. (2009). Selection on cis-regulatory variation at B4galnt2 and its influence on von Willebrand factor in house mice. Mol. Biol. Evol. *26*, 567–578.

Jung, Y.T., Wu, T., and Kozak, C.A. (2003). Characterization of recombinant nonecotropic murine leukemia viruses from the wild mouse species Mus spretus. J. Virol. *77*, 12773–12781.

Kauer, M.O., Dieringer, D., and Schlötterer, C. (2003). A microsatellite variability screen for positive selection associated with the “Out of Africa” habitat expansion of Drosophila melanogaster. Genetics *165*, 1137–1148.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. Nature *477*, 289–294.

Kelly, M., Holland, C.A., Lung, M.L., Chattopadhyay, S.K., Lowy, D.R., and Hopkins, N.H. (1983). Nucleotide sequence of the 3’ end of MCF 247 murine leukemia virus. J. Virol. *45*, 291–298.

Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M., and Pääbo, S. (2005). Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science *309*, 1850–1854.

Kim, Y., and Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics *160*, 765–777.

Kim, S., Bardwell, V., and Zarkower, D. (2007). Cell type-autonomous and non-autonomous requirements for Dmrt1 in postnatal testis differentiation. Dev. Biol. *307*, 314–327.

Kopp, A., Duncan, I., and Carroll, S.B. (2000). Genetic control and evolution of sexually dimorphic characters in Drosophila. Nature *408*, 553–559.

Kozak, C.A. (1983). Genetic mapping of a mouse chromosomal locus required for mink cell focus-forming virus replication. J. Virol. *48*, 300–303.

Kozak, C.A. (1985). Susceptibility of wild mouse cells to exogenous infection with xenotropic leukemia viruses: control by a single dominant locus on chromosome 1. J. Virol. *55*, 690–695.

Kozak, C.A. (2010). The mouse "xenotropic" gammaretroviruses and their XPR1 receptor. Retrovirology *7*, 101.

Kozak, C.A. (2011). Naturally occurring polymorphisms of the mouse gammaretrovirus receptors CAT-1 and XPR1 alter virus tropism and pathogenicity. Adv. Virol. *2011*.

Kozak, C.A. Evolution of different antiviral strategies in wild mouse populations exposed to different gammaretroviruses. Curr. Opin. Virol.

Krentz, A.D., Murphy, M.W., Kim, S., Cook, M.S., Capel, B., Zhu, R., Matin, A., Sarver, A.L., Parker, K.L., Griswold, M.D., et al. (2009). The DM domain protein DMRT1 is a dose-sensitive regulator of fetal germ cell proliferation and pluripotency. Proc. Natl. Acad. Sci. *106*, 22323 –22328.

Krentz, A.D., Murphy, M.W., Sarver, A.L., Griswold, M.D., Bardwell, V.J., and Zarkower, D. (2011). DMRT1 promotes oogenesis by transcriptional activation of Stra8 in the mammalian fetal ovary. Dev. Biol. *356*, 63–70.

Kwan, T., Benovoy, D., Dias, C., Gurd, S., Serre, D., Zuzan, H., Clark, T.A., Schweitzer, A., Staples, M.K., Wang, H., et al. (2007). Heritability of alternative splicing in the human genome. Genome Res. *17*, 1210–1218.

Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. Nature *413*, 519–523.

Latchman, D.S. (1997). Transcription factors: An overview. Int. J. Biochem. Cell Biol. *29*, 1305–1312.

Lei, N., Hornbaker, K.I., Rice, D.A., Karpova, T., Agbor, V.A., and Heckert, L.L. (2007). Sex-specific differences in mouse DMRT1 expression are both cell Type- and stage-dependent during gonad development. Biol. Reprod. *77*, 466 –475.

de-Leon, S.B.-T., and Davidson, E.H. (2007). Gene regulation: Gene control network in development. Annu. Rev. Biophys. Biomol. Struct. *36*, 191–212.

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics *25*, 1451–1452.

Lin, L., Shen, S., Jiang, P., Sato, S., Davidson, B.L., and Xing, Y. (2010). Evolution of alternative splicing in primate brain transcriptomes. Hum. Mol. Genet. *19*, 2958–2973.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. (2003). Protein disorder prediction: Implications for structural proteomics. Structure *11*, 1453–1459.

Linnenbrink, M. (2012). Population genetic and functional analysis of the B4galnt2 gene in the genus Mus (Rodentia; Muridae). PhD thesis. CAU.

Linnenbrink, M., Wang, J., Hardouin, E.A., Künzel, S., Metzler, D., and Baines, J.F. (2013). The role of biogeography in shaping diversity of the intestinal microbiota in house mice. Mol. Ecol. *22*, 1904–1916.

Lints, R., and Emmons, S.W. (2002). Regulation of sex-specific differentiation and mating behavior in C. elegans by a new member of the DM domain transcription factor family. Genes Dev. *16*, 2390–2402.

Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic disorder in transcription factors. Biochemistry (Mosc.) *45*, 6873–6888.

Lopez-Bigas, N., De, S., and Teichmann, S.A. (2008). Functional protein divergence in the evolution of Homo sapiens. Genome Biol. *9*, R33.

Lu, H., Huang, X., Zhang, L., Guo, Y., Cheng, H., and Zhou, R. (2007). Multiple alternative splicing of mouse Dmrt1 during gonadal differentiation. Biochem. Biophys. Res. Commun. *352*, 630–634.

Lyu, M.S., and Kozak, C.A. (1996). Genetic basis for resistance to polytropic murine leukemia viruses in the wild mouse species Mus castaneus. J. Virol. *70*, 830–833.

Marin, M., Tailor, C.S., Nouri, A., Kozak, S.L., and Kabat, D. (1999). Polymorphisms of the cell surface receptor control mouse susceptibilities to xenotropic and polytropic leukemia viruses. J. Virol. *73*, 9362–9368.

Matson, C.K., Murphy, M.W., Griswold, M.D., Yoshida, S., Bardwell, V.J., and Zarkower, D. (2010). The mammalian doublesex homolog DMRT1 is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. Dev. Cell *19*, 612–624.

Matson, C.K., Murphy, M.W., Sarver, A.L., Griswold, M.D., Bardwell, V.J., and Zarkower, D. (2011). DMRT1 prevents female reprogramming in the postnatal mammalian testis. Nature *advance online publication*.

McVean, G. (2007). The structure of linkage disequilibrium around a selective sweep. Genetics *175*, 1395–1406.

Meador, W.E., Means, A.R., and Quiocho, F.A. (1992). Target enzyme recognition by calmodulin: 2.4 A structure of a calmodulin-peptide complex. Science *257*, 1251–1255.

De Meaux, J., and Mitchell-Olds, T. (2003). Evolution of plant resistance at the molecular level: ecological context of species interactions. Heredity *91*, 345–352.

Mitchell-Olds, T., Willis, J.H., and Goldstein, D.B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat. Rev. Genet. *8*, 845–856.

Murphy, M., Zarkower, D., and Bardwell, V. (2007). Vertebrate DM domain proteins bind similar DNA sequences and can heterodimerize on DNA. BMC Mol. Biol. *8*, 58.

Murphy, M.W., Sarver, A.L., Rice, D., Hatzi, K., Ye, K., Melnick, A., Heckert, L.L., Zarkower, D., and Bardwell, V.J. (2010). Genome-wide analysis of DNA binding and transcriptional regulation by the mammalian doublesex homolog DMRT1 in the juvenile testis. Proc. Natl. Acad. Sci. *107*, 13360 –13365.

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 Genet. *17*, 481–485.

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

Nei, M. (1987). Molecular evolutionary genetics (Columbia University Press).

Nembaware, V., Wolfe, K.H., Bettoni, F., Kelso, J., and Seoighe, C. (2004). Allele-specific transcript isoforms in human. FEBS Lett. *577*, 233–238.

Nielsen, R. (2001). Statistical tests of selective neutrality in the age of genomics. Heredity *86*, 641–647.

Nielsen, R. (2005). Molecular signatures of natural selection. Annu. Rev. Genet. *39*, 197–218.

Oleksyk, T.K., Smith, M.W., and O'Brien, S.J. (2010). Genome-wide scans for footprints of natural selection. Philos. Trans. R. Soc. B Biol. Sci. *365*, 185–205.

Olofsson, S., and Bergström, T. (2005). Glycoconjugate glycans as viral receptors. Ann. Med. *37*, 154–172.

Otto, S.P. (2000). Detecting the form of selection from DNA sequence data. Trends Genet. *16*, 526–529.

Pozhitkov, A.E., Tautz, D., and Noble, P.A. (2007). Oligonucleotide microarrays: widely applied—poorly understood. Brief. Funct. Genomic. Proteomic. *6*, 141–148.

Pozhitkov, A.E., Noble, P.A., Bryk, J., and Tautz, D. (2014). A revised design for microarray experiments to account for experimental noise and uncertainty of probe response. PloS One *9*, e91295.

Przeworski, M. (2002). The signature of positive selection at randomly chosen loci. Genetics *160*, 1179–1189.

Rassart, E., Nelbach, L., and Jolicoeur, P. (1986). Cas-Br-E murine leukemia virus: sequencing of the paralytogenic region of its genome and derivation of specific

probes to study its origin and the structure of its recombinant genomes in leukemic tissues. J. Virol. *60*, 910–919.

Raymond, C.S., Kettlewell, J.R., Hirsch, B., Bardwell, V.J., and Zarkower, D. (1999). Expression of Dmrt1 in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. Dev. Biol. *215*, 208–220.

Raymond, C.S., Murphy, M.W., O'Sullivan, M.G., Bardwell, V.J., and Zarkower, D. (2000). Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. Genes Dev. *14*, 2587–2595.

Rieseberg, L.H., Widmer, A., Arntz, A.M., and Burke, J.M. (2002). Directional selection is the primary cause of phenotypic diversification. Proc. Natl. Acad. Sci. *99*, 12242–12245.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics *19*, 1572–1574.

Ross, J.M., Kalis, A.K., Murphy, M.W., and Zarkower, D. (2005). The DM domain protein MAB-3 promotes sex-specific neurogenesis in C. elegans by regulating bHLH proteins. Dev. Cell *8*, 881–892.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature *419*, 832–837.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913–918.

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.

Slatkin, M. (1995). Hitchhiking and associative overdominance at a microsatellite locus. Mol. Biol. Evol. *12*, 473–480.

Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genet. Res. *23*, 23–35.

Stajich, J.E., and Hahn, M.W. (2005). Disentangling the effects of demography and selection in human history. Mol. Biol. Evol. *22*, 63–73.

Staubach, F., Teschke, M., Voolstra, C.R., Wolf, J.B.W., and Tautz, D. (2010). A test of the neutral model of expression change in natural populations of house mouse subspecies. Evolution *64*, 549–560.

Staubach, F., Lorenc, A., Messer, P.W., Tang, K., Petrov, D.A., and Tautz, D. (2012). Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (Mus musculus). PLoS Genet *8*, e1002891.

Stearns, S.C., and Hoekstra, R.F. (2005). Evolution: an introduction (Oxford [England]; New York: Oxford University Press).

Steiner, C.C., Weber, J.N., and Hoekstra, H.E. (2007). Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biol *5*, e219.

Storz, J.F., Payseur, B.A., and Nachman, M.W. (2004). Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. Mol. Biol. Evol. *21*, 1800–1811.

Storz, J.F., Sabatino, S.J., Hoffmann, F.G., Gering, E.J., Moriyama, H., Ferrand, N., Monteiro, B., and Nachman, M.W. (2007). The molecular basis of high-altitude adaptation in deer mice. PLoS Genet *3*, e45.

Stoye, J.P., Moroni, C., and Coffin, J.M. (1991). Virological events leading to spontaneous AKR thymomas. J. Virol. *65*, 1273–1285.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. U. S. A. *101*, 6062–6067.

Tailor, C.S., Nouri, A., Lee, C.G., Kozak, C., and Kabat, D. (1999). Cloning and characterization of a cell surface receptor for xenotropic and polytropic murine leukemia viruses. Proc. Natl. Acad. Sci. U. S. A. *96*, 927–932.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.

Tautz, D. (2000). Evolution of transcriptional regulation. Curr. Opin. Genet. Dev. *10*, 575–579.

Teschke, M., Mukabayire, O., Wiehe, T., and Tautz, D. (2008). Identification of selective sweeps in closely related populations of the house mouse based on microsatellite scans. Genetics *180*, 1537–1545.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A.M. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. *35*, W71–74.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.

Vaughan, A.E., Mendoza, R., Aranda, R., Battini, J.-L., and Miller, A.D. (2012). Xpr1 is an atypical G-protein-coupled receptor that mediates xenotropic and polytropic murine retrovirus neurotoxicity. J. Virol. *86*, 1661–1669.

Wang, W., Kidd, B.J., Carroll, S.B., and Yoder, J.H. (2011). Sexually dimorphic regulation of the Wingless morphogen controls sex-specific segment number in Drosophila. Proc. Natl. Acad. Sci. *108*, 11139–11144.

Wensel, D.L., Li, W., and Cunningham, J.M. (2003). A virus-virus interaction circumvents the virus receptor requirement for infection by pathogenic retroviruses. J. Virol. *77*, 3460–3469.

Williams, T.M., Selegue, J.E., Werner, T., Gompel, N., Kopp, A., and Carroll, S.B. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. Cell *134*, 610–623.

Woolhouse, M.E.J., Webster, J.P., Domingo, E., Charlesworth, B., and Levin, B.R. (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat. Genet. *32*, 569–577.

Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J. Mol. Biol. *293*, 321–331.

WRIGHT, S. (1951). The genetical structure of populations. Ann. Eugen. *15*, 323–354.

Wunderlich, Z., and Mirny, L.A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. Trends Genet. *25*, 434–440.

Yan, Y., Knoper, R.C., and Kozak, C.A. (2007). Wild mouse variants of envelope genes of xenotropic/polytropic mouse gammaretroviruses and their XPR1 receptors elucidate receptor determinants of virus entry. J. Virol. *81*, 10550–10557.

Yan, Y., Jung, Y.T., Wu, T., and Kozak, C.A. (2008). Role of receptor polymorphism and glycosylation in syncytium induction and host range variation of ecotropic mouse gammaretroviruses. Retrovirology *5*, 2.

Yan, Y., Liu, Q., and Kozak, C.A. (2009). Six host range variants of the xenotropic/polytropic gammaretroviruses define determinants for entry in the XPR1 cell surface receptor. Retrovirology *6*, 87.

Yan, Y., Liu, Q., Wollenberg, K., Martin, C., Buckler-White, A., and Kozak, C.A. (2010). Evolution of functional and sequence variants of the mammalian XPR1 receptor for mouse xenotropic gammaretroviruses and the human-derived retrovirus XMRV. J. Virol. *84*, 11970–11980.

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. Mol. Biol. Evol. *5*, 63–78.

Zhang, W., Morris, Q.D., Chang, R., Shai, O., Bakowski, M.A., Mitsakakis, N., Mohammad, N., Robinson, M.D., Zirngibl, R., Somogyi, E., et al. (2004). The functional landscape of mouse gene expression. J. Biol. *3*, 21.

Zhu, L., Wilken, J., Phillips, N.B., Narendra, U., Chan, G., Stratton, S.M., Kent, S.B., and Weiss, M.A. (2000). Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. Genes Dev. *14*, 1750–1764.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die hier vorliegende Dissertation mit dem Titel:

Functional characterization of adaptively relevant genes in the house mouse (*Mus musculus* L.)

in Form und Inhalt selbstständig, mit der Unterstützung meines Betreuers, verfasst habe. Ich habe keine anderen als die angegebenen Hilfsmittel und Quellen verwendet und die Arbeit unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft erstellt.

Diese Arbeit wurde an keiner anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegt, veröffentlicht oder zur Veröffentlichung eingereicht. Dies ist mein bisher erstes und einziges Promotionsverfahren.

Kiel, im April 2014 Natascha Hasenkamp

Appendix

S1 (General):

Table S1: Origin of wild-caught mice.

population ID	sampling location	*Mus musculus* subspecies	sampled by	year
Musc-AL	Kazakhstan	*musculus*	Ihle et al.	2001
Musc-CR	Czech Republic	*musculus*	Ihle et al.	2001
Dom-IR	Iran	*domesticus*	Bonhomme et al.	NA
Ger_{CB}	Cologne-Bonn, Germany	*domesticus*	Linnenbrink et al.	2010
Ger_{SL}	Schömberg/Langenbrand, Germany	*domesticus*	Linnenbrink et al.	2010
Fra_{NA}	Nancy, France	*domesticus*	Linnenbrink et al.	2010
Fra_{LO}	Louan-Villegruis, France	*domesticus*	Linnenbrink et al.	2010
Fra_{DB}	Divonne les Bains, France	*domesticus*	Linnenbrink et al.	2010
Fra_{AN}	Angers, France	*domesticus*	Linnenbrink et al.	2009
Fra_{ES}	Espelette, France	*domesticus*	Linnenbrink et al.	2009
Fra_{MC}	Severac le Château	*domesticus*	Ihle et al., Linnenbrink et al.	2001 2009

S2 (Chapter 1):

Table S2.2: Primers for microsatellite loci in the genomic region of *Xpr1*.

primer ID	sequence	mean product length	pool
ms1-F	ACAGCTGGCAATTTGGAGAT	345	a
ms1-R	AGTCAGCTGCCACAACACAG		
ms2-F	GAGGAGGGAGGCTCTCAAGT	240	b
ms2-R	TCCTTTCTGTCACCAATGGAC		
ms3-F	TGCAACTCCAGTAACGCAAG	251	d
ms3-R	GCCTCCTATATGGCCACTCA		
ms6-F	TGAAGAAAATTGTCATGAACAAGAA	266	c
ms6-R	CCAGAGTCACAGTCACAAGCA		
ms8-F	ATACAGGCGCCAAGTCCATA	158	a
ms8-R	GAGTGCCAGGGTTACATGGT		
ms10-F	TCCATCACAAGGCTGTATGTG	147	b
ms10-R	AAACACCGGGCTATGAAATG		
ms11-F	CAGCTGCCCAGAGGAGTAAG	344	d
ms11-R	TTTCATACCTGCAACCCACA		
ms13-F	CCCTTTCCCTGGTCTTCAAT	459	c
ms13-R	CAAGGCTTTTGGAGGGTTTT		
ms15-F	GAGGTAGAGCGCATGGATTC	448	a
ms15-R	TCCCCCAGCTCTACTCTGAA		

Table S2.3: PCR and sequencing primers (Metabion) with PCR program modifications.

primer ID	sequence	elongation time (s)
ex4-F	GGGCCAAAATGCTTTCTCTT	30
ex4-R	TGATTTCAATCTTTAGAGGATTCAGT	
ECL3.1-F	TCCATAAGGTAGGCTTTGCTG	30
ECL3.1-R	TCTTGGTTTATGCTGGCAATC	
ECL3.2-F	CACACACTGATGGGGAGTTG	30
ECL3.2-R	GCAAAGTCCAGGAAAGCAGA	
ECL3.3-F	TGGGCACTATGAAGAATCCA	30
ECL3.3-R	GAGACCCCAGTCCATCTTGA	
ECL4-F	AACGCTTCTCCATGAGTCTTTG	30
ECL4-R	GATCAGACTTGGTATAAGTGTCT	
P-MLV-F	AAGGGCAGGAGTATCAGTACAACAT	90
P-MLV-R	GGGACGCGGGGCCCTATATTGAG	

Table S2.4: Overview of SNPs that characterize individual *Xpr1*-haplotypes in wild mouse populations. Grey background indicates nonsynonymous SNPs, white letters show SNPs in regions which are involved in virus interaction.

individuals	EX 4									ECL 3.2			ECL 3.3	EX 12			ECL 4																haplotype
	G	C	A	G	T	G	T	C	A	A	C	C	T	G	T	G	A	T	T	A	C	A	A	C	G	T	T	T	A	A	G	C	C	T	
Fra-AN1301-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN1301-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN55-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN55-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN2601-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN2601-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN3101-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN3101-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN3001-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN3001-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN2901-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN2901-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN0801-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN0801-2	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-AN1501-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN1501-2	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-AN2301-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-AN2301-2	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-AN1401-1	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-AN1401-2	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-AN0601-1	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-AN0601-2	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X

individuals	EX 4	ECL 3.2	ECL 3.3	EX 12	ECL 4	haplotype
Fra-AN2701-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-AN2701-2	. G . A C	. . .	. . .	.	. . . . .	X
Ger-CB17-1	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB17-2	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB15-1	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB15-2	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB18-1	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB18-2	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB20-1	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB20-2	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB01-1	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB01-2	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB09-1	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB09-2	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB05-1	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB05-2	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB13-1	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB13-2	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB03-1	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB03-2	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB21-1	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB21-2	. . . . .	. . .	. . .	.	. . . . .	IX
Ger-CB10-1	. G . A C	. . .	. . .	.	. . . . .	X
Ger-CB10-2	. . . . .	. . .	. . .	.	. . . . .	IX

individuals	EX 4	ECL 3.2	ECL 3.3	EX 12	ECL 4	haplotype
Fra-DB221-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB221-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB7-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB7-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB81-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB81-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB151-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB151-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB11-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB11-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB3-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB3-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB201-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB201-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB191-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB191-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-DB22-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB22-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-DB131-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB131-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-DB14-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-DB14-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-DB18-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-DB18-2	. G . A C	. . .	. . .	.	. . . . .	X

individuals	EX 4	ECL 3.2	ECL 3.3	EX 12	ECL 4	haplotype
Fra-ES0201-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-ES0201-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-ES0501-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-ES0501-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-ES0901-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-ES0901-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-ES1101-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-ES1101-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES1301-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-ES1301-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES1501-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-ES1501-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES1201-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES1201-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES1401-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES1401-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES0801-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES0801-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES0701-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES0701-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES0601-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES0601-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES0401-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-ES0401-2	. . . .	. . .	. . .	.	. . . . .	IX

individuals	EX 4	ECL 3.2	ECL 3.3	EX 12	ECL 4	haplotype
Fra-LO101-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO101-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO13-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO13-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO21-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO21-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO31-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO31-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO141-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO141-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-LO81-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO81-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-LO91-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO91-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-LO51-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO51-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-LO41-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO41-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-LO11-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-LO11-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-LO61-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-LO61-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-LO71-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-LO71-2	. . . .	. . .	. . .	.	. . . . .	IX

individuals	EX 4									ECL 3.2			ECL 3.3	EX 12			ECL 4	haplotype
Fra-MC_27-1	.	G	G	A	.	.	.	.	.	.	.	T	A	.	.	.	.	XI
Fra-MC_27-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-MC_2-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_2-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-MC_3-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_3-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-MC_18-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_18-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_13-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_13-2	.	G	G	A	.	.	.	.	.	.	.	T	A	.	.	.	.	XI
Fra-MC_21-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_21-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_25-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_25-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_22-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_22-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	VII
Fra-MC_7-1	.	G	G	A	.	.	.	.	.	.	.	T	A	.	.	.	.	XI
Fra-MC_7-2	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-MC_10-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-MC_10-2	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	XIII
Fra-MC_4-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-MC_4-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-MC_6-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-MC_6-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX

individuals	EX 4										ECL 3.2			ECL 3.3	EX 12			ECL 4	haplotype
Fra-MC_2101-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_2101-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_2301-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_2301-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_1501-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_1501-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_1201-1	.	G	G	A	.	.	.	.	.	.	.	T	A	.	.	.	.	.	XI
Fra-MC_1201-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_0301-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_0301-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_1301-1	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-MC_1301-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_2201-1	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-MC_2201-2	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_0201-1	.	G	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	XII
Fra-MC_0201-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-MC_2001-1	.	G	G	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VII
Fra-MC_2001-2	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	VIII
Fra-MC_2501-1	.	S	R	R	Y	.	.	.	.	.	.	Y	W	.	.	.	.	.	other
Fra-MC_2501-2	.	S	R	R	Y	.	.	.	.	.	.	Y	W	.	.	.	.	.	other
Fra-MC_0601-1	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-MC_0601-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX
Fra-MC_1101-1	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	X
Fra-MC_1101-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IX

individuals	EX 4	ECL 3.2	ECL 3.3	EX 12	ECL 4	haplotype
Fra-NA141-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA141-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA31-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA31-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA61-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA61-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA21-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA21-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA131-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA131-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA101-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA101-2	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA111-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA111-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-NA121-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA121-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-NA41-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA41-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-NA71-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA71-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-NA92-1	. . . .	. . .	. . .	.	. . . . .	IX
Fra-NA92-2	. G . A C	. . .	. . .	.	. . . . .	X
Fra-NA81-1	. G . A C	. . .	. . .	.	. . . . .	X
Fra-NA81-2	. G . A C	. . .	. . .	.	. . . . .	X

individuals	EX 4	ECL 3.2	ECL 3.3	EX 12	ECL 4	haplotype
Ger-SL11-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL11-2	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL3-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL3-2	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL91-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL91-2	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL181-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL181-2	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL51-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL51-2	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL41-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL41-2	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL171-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL171-2	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL151-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL151-2	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL131-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL131-2	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL141-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL141-2	. G . A C	. . .	. . .	.	. . . . .	X
Ger-SL83-1	. . . .	. . .	. . .	.	. . . . .	IX
Ger-SL83-2	. G . A C	. . .	. . .	.	. . . . .	X
Ger-SL16-1	. G . A C	. . .	. . .	.	. . . . .	X
Ger-SL16-2	. G . A C	. . .	. . .	.	. . . . .	X

individuals	EX 4	ECL 3.2	ECL 3.3	EX 12	ECL 4	haplotype
Dom-IR21-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR21-2	. . . .	. . .	A . .	.	. . . . .	VIII
Dom-IR63-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR63-2	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR64-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR64-2	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR61-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR61-2	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR52-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR52-2	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR51-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR51-2	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR31-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR31-2	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR29-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR29-2	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR27-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR27-2	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR25-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR25-2	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR23-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR23-2	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR22-1	. G G A	. . .	A . .	.	. . . . .	VII
Dom-IR22-2	. G G A	. . .	A . .	.	. . . . .	VII

individuals	EX 4	ECL 3.2	ECL 3.3	EX 12	ECL 4	haplotype
Musc-CR1-1	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR1-2	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR18-1	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR18-2	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR10-1	. G G A	C . .	. . .	.	- - - . . . . A	VI
Musc-CR10-2	. G G A	C . .	. . .	.	- - - . . . . A	VI
Musc-CR12-1	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR12-2	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR13-1	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR13-2	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR14-1	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR14-2	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR16-1	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR16-2	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR17-1	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR17-2	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR2-1	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR2-2	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR3-1	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR3-2	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR6-1	. G G A	C . .	. . .	.	- - - . . . . A	VI
Musc-CR6-2	. G G A	C . .	. . .	.	- - - . . . . A	VI
Musc-CR5-1	. G G A	c . .	. . .	.	- - - . . . . a	VI
Musc-CR5-2	. G G A	c . .	. . .	.	- - - . . . . a	VI

individuals	EX 4	ECL 3.2	ECL 3.3	EX 12	ECL 4	haplotype
Musc-AL11-1	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL11-2	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL1-1	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL1-2	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL6-1	. G G A	C . .	.	A . .	- - - . . . . A	IV
Musc-AL6-2	. G G A	C . .	.	A . .	- - - . . . . A	IV
Musc-AL15-1	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL15-2	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL13-1	A G G A	c . .	.	A . .	- - - . . . . a	V
Musc-AL13-2	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL2-1	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL2-2	. G G A	c . .	.	. . .	- - - . . . . a	VI
Musc-AL3-1	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL3-2	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL4-1	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL4-2	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL9-1	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL9-2	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL14-1	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL14-2	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL10-1	. G G A	c . .	.	A . .	- - - . . . . a	IV
Musc-AL10-2	. G G A	c . .	.	. . .	- - - . . . . a	VI
Musc-AL7-1	. G G A	C . .	.	A . .	- - - . . . . A	IV
Musc-AL7-2	. G G A	C . .	.	A . .	- - - . . . . A	IV

individuals	EX 4	ECL 3.2	ECL 3.3	EX 12	ECL 4	haplotype
Cast-TAI2967-1	. G . . . . C . .	. . .	. . .	.	. . . - - - - - - - - - - - - - - - -	III
Cast-TAI2967-2	. G . . . . C . .	. . .	. . .	.	. . . - - - - - - - - - - - - - - - -	III
Cast-TAI2971-1	. G . . . . C . .	. . .	. . .	.	. . . - - - - - - - - - - - - - - - -	III
Cast-TAI2971-2	. G . . . . C . .	. . .	. . .	.	. . . - - - - - - - - - - - - - - - -	III
Scl100-1	. G . A . A . T .	. . .	. .	C .	. . . . .	I
Scl100-2	. G . A . A . T .	. . .	. .	C .	. . . . .	I
Scl103-1	. G . A . A . T .	. . .	. .	C .	. . . . .	I
Scl103-2	. G . A . A . T .	. . .	. .	C .	. . . . .	I
SP2723-1	. G . A . A . T .	. T .	. . .	A	. . . . .	II
SP2723-2	. G . A . A . T .	. T .	. . .	A	. . . . .	II
SP2729-1	. G . A . A . T .	. T .	. . .	A	. . . . .	II
SP2729-2	. G . A . A . T .	. T .	. . .	A	. . . . .	II

Table S2.5: Overview of SNPs which occurred in the sequences provirus RBDs. Grey background indicates nonsynonymous SNPs, blue letters mark SNPs that lie in one of the hypervariable regions.

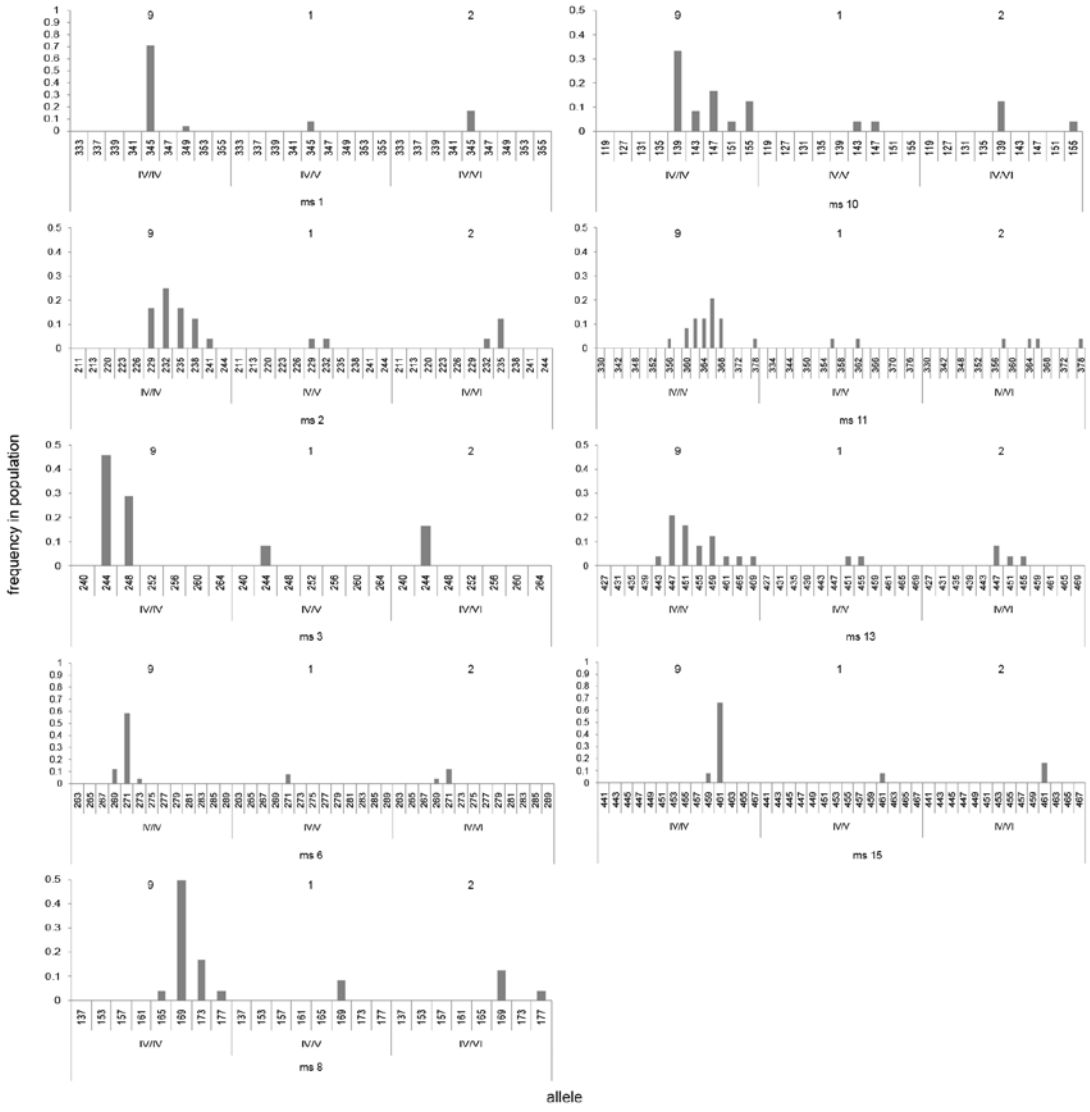
Sample	SNPs																																							
Position	16	36	39	59	73	95	100	112	114	115	122	124	151	154	155	169	200	204	207	259	316	356	369	385	399	400	411	415	416	421	422	429	457	458	475	490	516	527	534	535	537
Dom-IR7049	T	G	R	G	T	A	C	T	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	G	T	C	G	A	C	C	C	G	C	A	G	C	G	G	Y	R	G	C
Dom-IR10191	T	G	G	G	T	A	C	T	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	G	T	C	G	A	C	C	C	G	C	A	G	C	G	G	Y	G	G	C
Dom-IR10262	T	G	G	G	T	A	C	T	A	A	T	T	C	G	G	C	G	G	T	G	G	G	G	T	C	G	A	C	C	C	G	C	A	G	C	G	G	C	G	G	C
Dom-IR7037	T	G	G	G	T	A	C	T	A	A	T	T	C	G	G	C	K	G	Y	G	R	G	G	T	C	G	A	C	C	C	G	C	A	G	C	G	G	C	G	G	C
Dom-IR7040	T	G	G	G	T	A	C	T	A	A	T	T	C	R	G	C	K	G	T	G	G	G	R	Y	C	G	A	C	C	C	G	C	A	R	C	G	G	C	G	G	C
Dom-IR10030	T	G	R	G	T	A	C	T	A	A	T	T	C	G	G	C	G	G	T	G	G	G	G	Y	C	G	A	C	C	C	G	C	A	G	C	G	G	Y	G	G	C
Ger-CB6926	T	R	G	G	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	C	S	G	R	R	T	C	G	R	C	M	Y	G	M	R	G	Y	G	G	T	G	R	C
Ger-CB6710	T	G	G	G	Y	A	M	Y	W	R	Y	T	C	G	G	Y	G	G	C	C	G	G	G	T	C	G	G	C	C	T	G	C	A	G	T	G	G	T	G	A	C
Ger-CB6704	T	G	G	G	Y	A	M	Y	W	R	T	Y	C	G	G	Y	G	G	C	S	G	G	G	T	C	G	G	C	C	T	G	M	A	R	T	R	G	T	G	A	C
Ger-CB6760	T	G	G	G	T	R	C	C	A	A	T	T	C	G	G	C	G	R	C	C	G	G	G	T	C	G	G	C	C	T	G	M	A	R	T	G	G	T	G	A	C
Ger-CB6714	T	G	G	G	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	C	S	R	G	G	T	C	G	R	C	M	Y	G	C	R	G	Y	G	G	T	G	R	C
Ger-CB6876	T	G	G	G	Y	A	M	Y	W	R	T	Y	C	G	G	C	G	R	C	S	G	R	R	T	C	G	R	C	M	Y	G	C	R	R	Y	G	G	T	G	A	C
Fra-MC6355	T	G	G	G	T	A	C	C	A	A	T	T	C	G	G	C	G	G	C	C	G	G	G	T	C	G	G	C	C	T	G	C	A	G	T	G	G	T	G	A	C
Fra-MC6336	Y	G	G	G	Y	A	M	Y	W	R	T	Y	Y	G	G	Y	G	G	C	S	G	G	G	T	Y	G	G	C	C	T	G	C	A	G	T	G	G	T	G	R	Y
Fra-MC6437	Y	G	G	G	Y	A	M	Y	W	R	T	Y	Y	G	G	Y	G	G	C	S	G	G	G	T	Y	R	G	C	C	T	G	C	A	G	T	G	G	T	G	R	Y
Fra-MC6471	Y	G	G	G	Y	A	M	Y	W	R	T	Y	Y	G	G	Y	G	G	C	S	G	G	G	T	Y	G	G	C	C	T	G	C	R	G	T	G	G	T	G	R	Y
Fra-MC6984	Y	G	G	G	Y	A	M	Y	W	R	T	Y	Y	G	G	Y	G	G	C	S	G	G	G	T	Y	G	G	C	C	T	G	C	A	G	T	G	G	T	G	R	Y
Fra-MC6398	Y	G	G	G	Y	A	M	Y	W	R	T	Y	Y	G	G	Y	G	G	C	G	G	G	G	T	Y	G	R	C	M	Y	G	C	R	G	T	G	G	T	G	R	Y
Fra-MC6472	Y	G	G	G	Y	A	M	Y	W	R	T	Y	Y	G	G	Y	G	G	C	S	G	G	G	T	Y	R	R	C	C	Y	G	C	R	G	Y	G	G	T	G	R	Y
Fra-MC6403	C	G	G	G	Y	A	M	Y	W	R	T	Y	T	G	G	Y	G	G	C	S	G	G	G	T	Y	G	G	C	C	T	G	C	R	G	T	G	G	T	G	R	Y
Fra-MC6571	Y	G	G	G	Y	A	M	Y	W	R	T	Y	Y	G	G	Y	G	G	C	S	G	G	G	T	Y	R	G	C	C	T	G	C	A	G	T	G	G	T	G	R	Y
Fra-MC6466	Y	G	G	G	Y	A	M	T	W	R	T	Y	Y	G	G	Y	G	G	C	G	G	G	G	T	Y	G	R	C	C	Y	G	C	R	G	T	G	G	T	G	R	Y

population samples (stock)

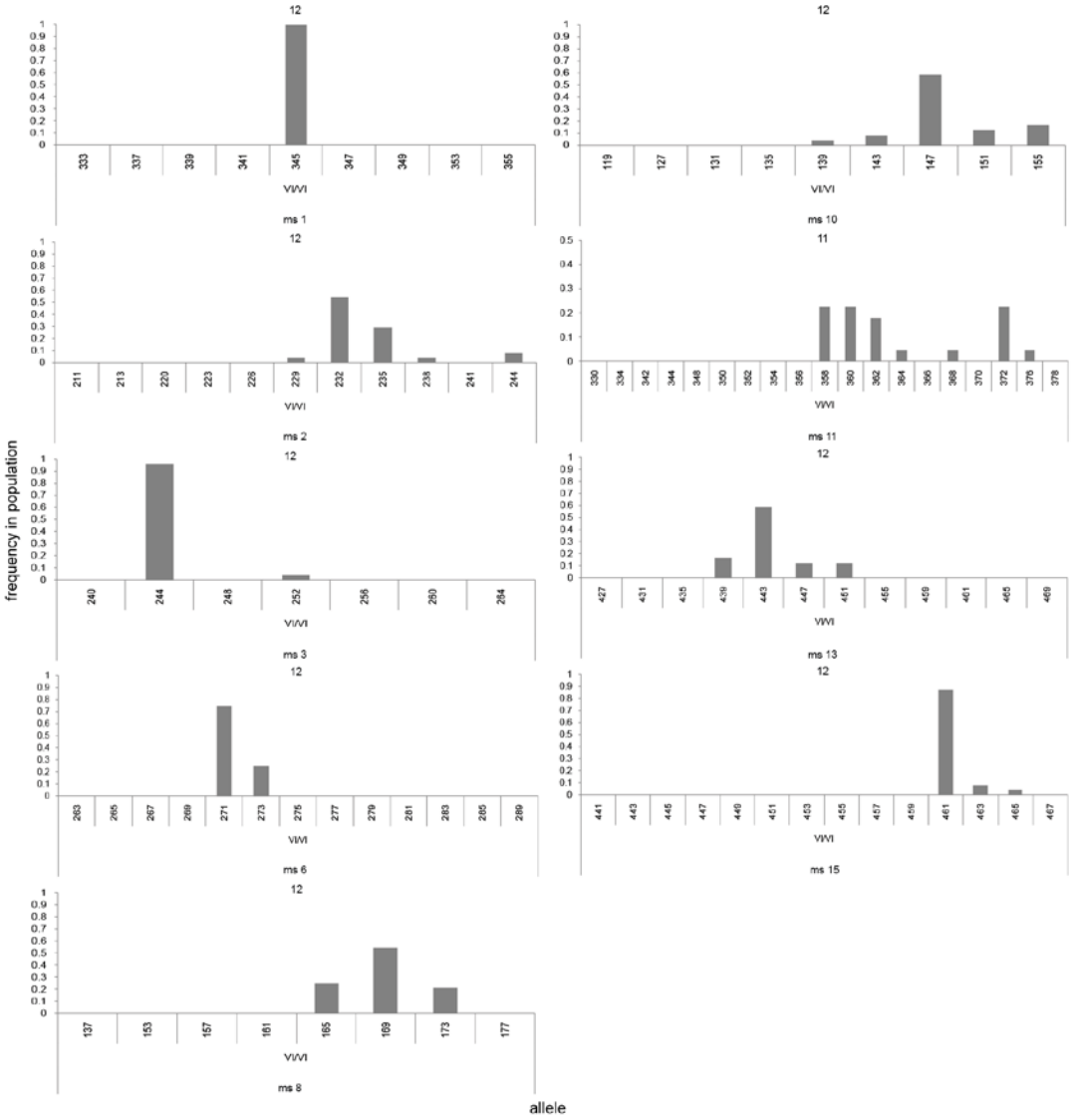
Fra-MC6473	Y	G	G	G	Y	A	M	Y	W	R	T	Y	Y	G	G	Y	G	G	C	S	G	G	G	T	Y	R	G	C	C	T	G	C	A	G	T	G	G	T	G	R	Y	
Fra-MC7045	Y	G	G	G	Y	A	M	T	W	R	T	Y	Y	G	G	Y	G	G	C	G	G	G	G	T	Y	R	R	C	C	Y	G	C	R	G	Y	G	G	T	G	G	Y	
5IRF1	T	G	G	G	T	A	C	T	A	A	T	T	C	R	R	C	G	G	T	G	G	G	R	Y	C	G	A	C	C	C	G	C	A	R	C	G	R	C	G	G	C	MLV101 parents (1=before mating, 2=after mating)
13IRF2	T	G	G	G	T	A	C	T	A	A	T	T	C	R	R	C	G	G	T	G	G	G	R	Y	C	G	A	C	C	C	G	C	A	R	C	G	R	C	G	G	C	
11CBM1	T	G	G	G	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	C	S	G	G	G	T	C	G	R	C	M	Y	G	C	R	G	Y	G	G	T	G	R	C	
19CBM2	T	G	G	G	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	C	S	G	G	G	T	C	G	R	C	M	Y	G	C	R	G	Y	G	G	T	G	R	C	
25CBxIR	T	G	G	G	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	G	T	C	G	A	C	C	C	G	C	A	G	C	G	G	Y	G	G	C	MLV101 litter
26CBxIR	T	G	G	G	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	R	Y	C	G	A	C	C	C	G	C	A	G	C	G	G	Y	G	G	C	
27CBxIR	T	G	G	G	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	R	Y	C	G	A	C	C	C	G	C	A	R	C	G	G	Y	G	G	C	
28CBxIR	T	G	G	G	T	A	C	T	A	A	T	T	C	R	R	C	G	G	Y	G	G	G	R	Y	C	G	A	C	C	C	G	C	A	R	C	G	G	Y	G	G	C	
29CBxIR	T	G	G	G	T	A	C	T	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	G	T	C	G	A	C	C	Y	G	C	A	G	C	G	G	Y	G	G	C	
30CBxIR	T	G	R	G	T	A	C	T	A	A	T	T	C	G	R	C	G	G	Y	G	G	G	G	T	C	G	A	C	C	C	G	C	A	G	C	G	G	Y	G	G	C	
6IRM1	T	G	R	G	T	A	C	T	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	G	T	C	G	A	Y	C	C	G	C	A	G	C	G	G	Y	G	G	C	MLV102 parents (1=before mating, 2=after mating)
14IRM2	T	G	R	G	T	A	C	T	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	G	T	C	G	A	Y	C	C	G	C	A	G	C	G	G	Y	G	G	C	
8CBF1	T	G	G	R	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	C	S	R	G	G	T	C	G	R	C	M	Y	R	C	R	G	Y	G	G	T	G	R	C	
16CBF2	T	G	G	R	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	C	S	R	G	G	T	C	G	R	C	M	Y	R	C	R	G	Y	G	G	T	G	R	C	
20IRxCB	T	G	G	G	T	A	C	T	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	G	T	C	G	A	C	C	C	G	C	A	G	C	G	G	Y	G	G	C	MLV102 litter
21IRxCB	T	G	G	G	T	A	C	T	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	G	T	C	G	A	C	C	C	G	C	A	G	C	G	G	Y	G	R	C	
22IRxCB	T	G	R	G	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	Y	S	G	G	G	T	C	G	R	C	C	Y	G	C	A	G	Y	G	G	T	G	R	C	
23IRxCB	T	G	G	G	T	A	C	T	A	A	T	T	C	G	G	C	G	G	Y	G	G	G	G	T	C	G	A	C	C	C	G	C	A	G	C	G	G	Y	G	G	C	
24IRxCB	T	G	G	G	T	A	C	Y	A	A	T	T	C	G	G	C	G	G	Y	S	G	G	G	T	C	G	A	C	C	C	G	C	A	G	C	G	G	Y	G	G	C	
Amino acids	-	W/S	R/K	G/R	-	M/V	-	-	K/M	-	F/L	-	-	W/*	D/N	-	A/S	R/K	I/T	-	W/*	G/R	R/Q	-	A/V	-	D/G	-	Q/K	-	A/T	P/Q	-	E/K	-	W/*	R/K	S/P	E/G	-	T/I	

Figure S2.1: Allele frequency spectra of microsatellite loci around *Xpr1.* The frequency of each microsatellite allele in a population is shown for each occurring *Xpr1* allele which is indicated below the x-axis in Roman numbers. Numbers above the histograms indicate the number of successfully genotyped samples with the respective *Xpr1* allele. The microsatellite locus is indicated below the each histogram.

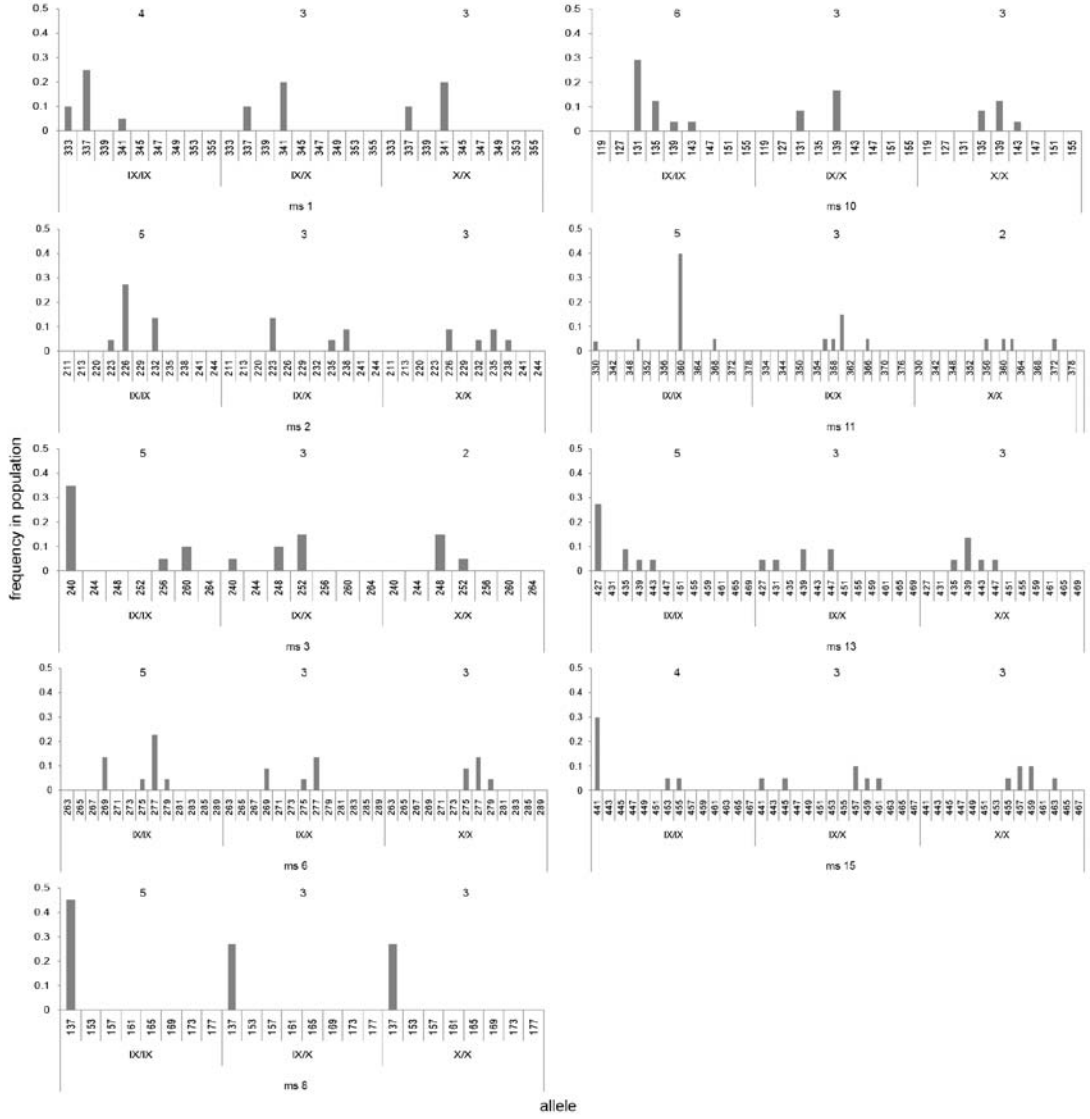
Musc-AL



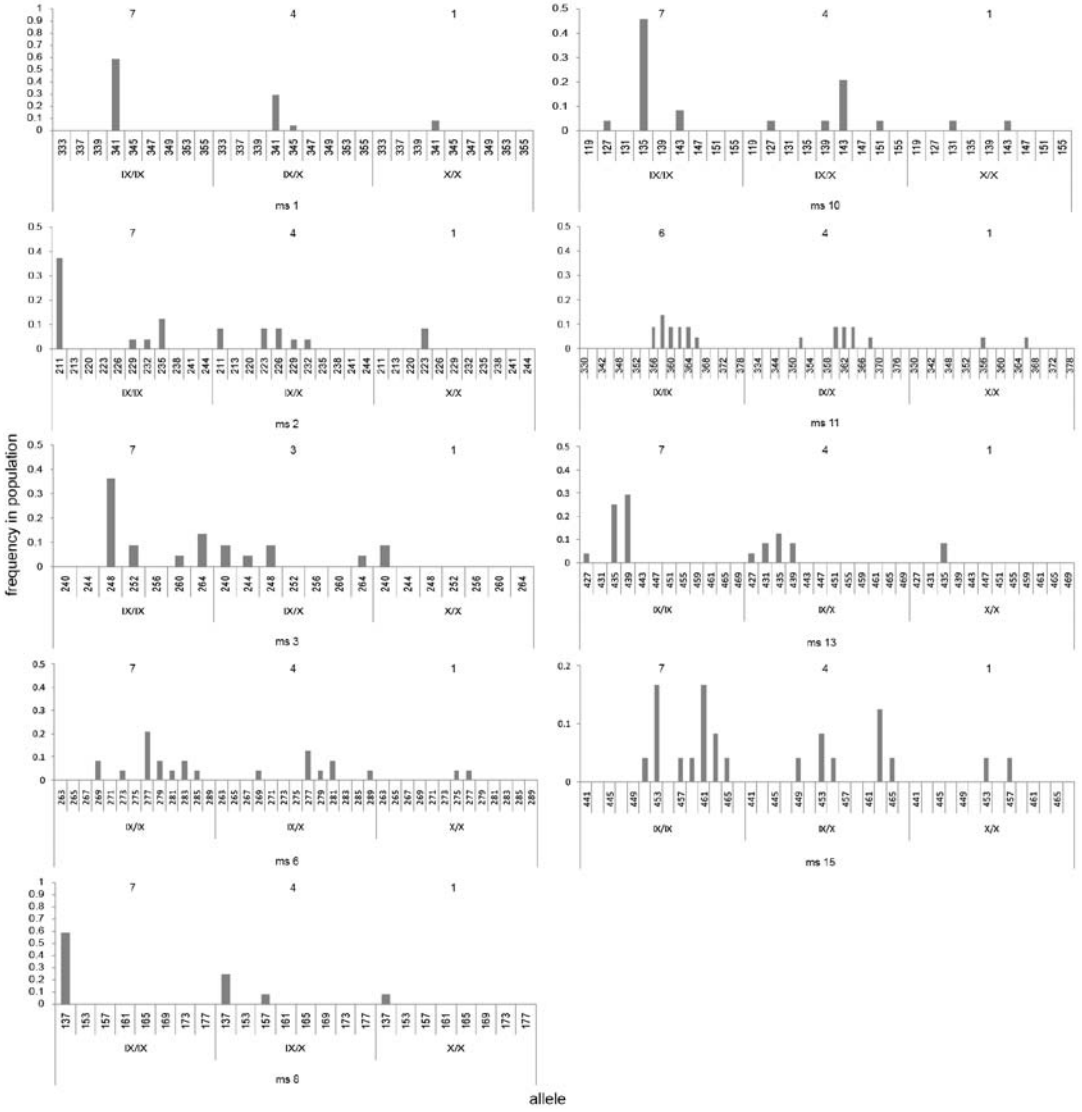
Musc-CR


frequency in population
allele
ms 1
ms 10
ms 2
ms 11
ms 3
ms 13
ms 6
ms 15
ms 8

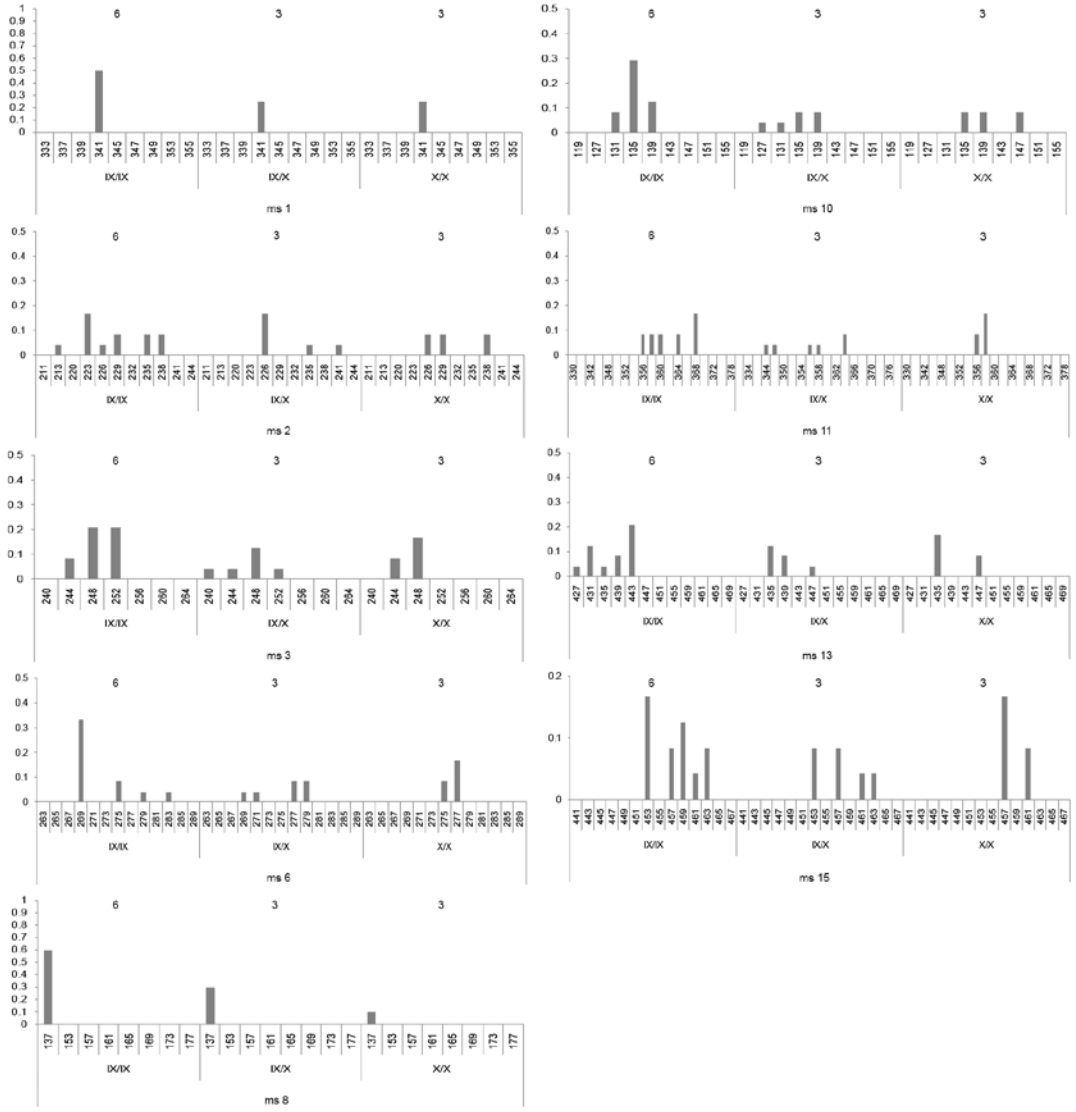
Fra-AN


frequency in population
allele
ms 1
ms 10
ms 2
ms 11
ms 3
ms 13
ms 6
ms 15
ms 8
IX/IX
IX/X
X/X

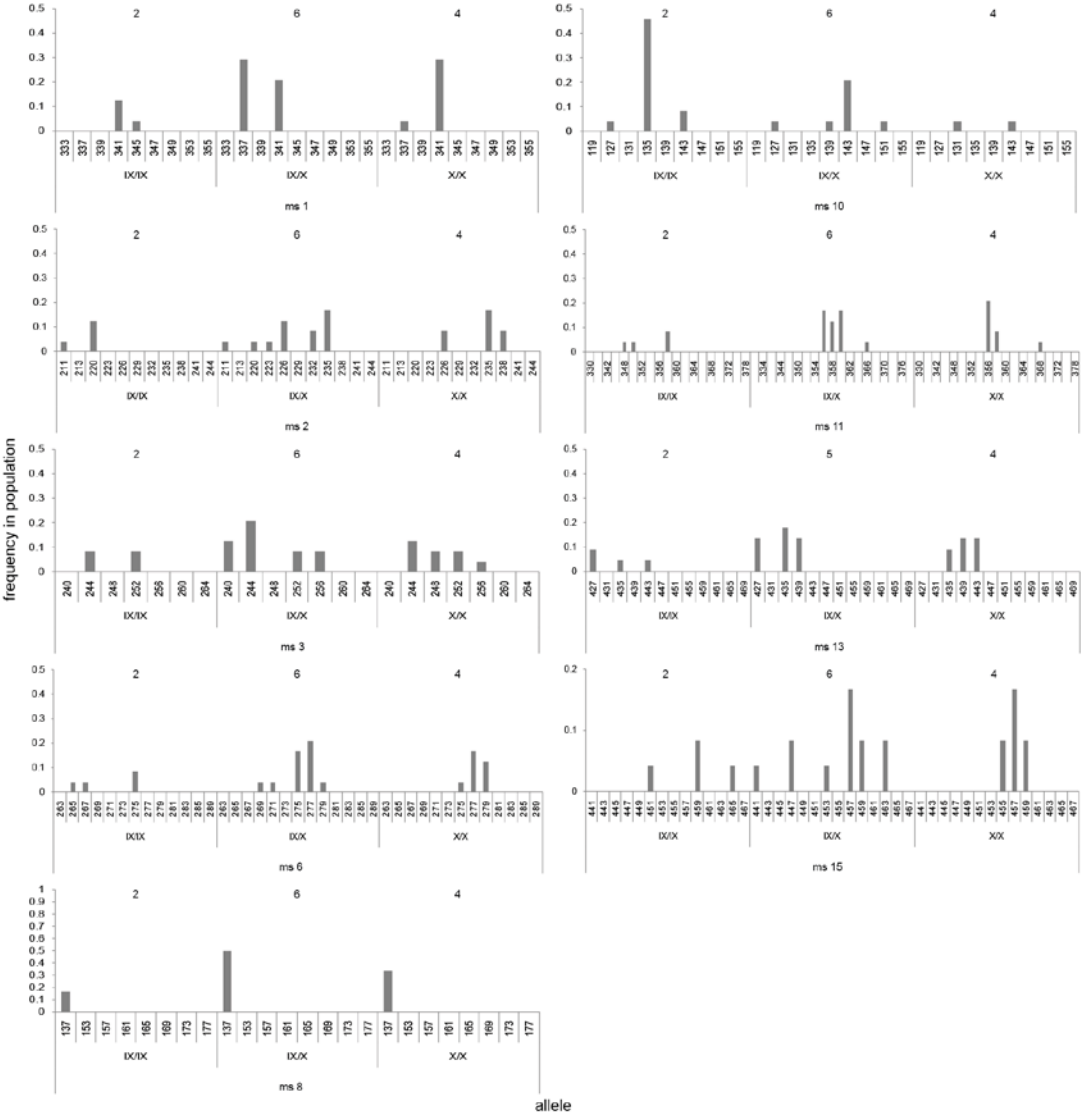
Fra-DB


frequency in population
allele
ms 1
ms 10
ms 2
ms 11
ms 3
ms 13
ms 6
ms 15
ms 8
IX/IX
IX/X
X/X

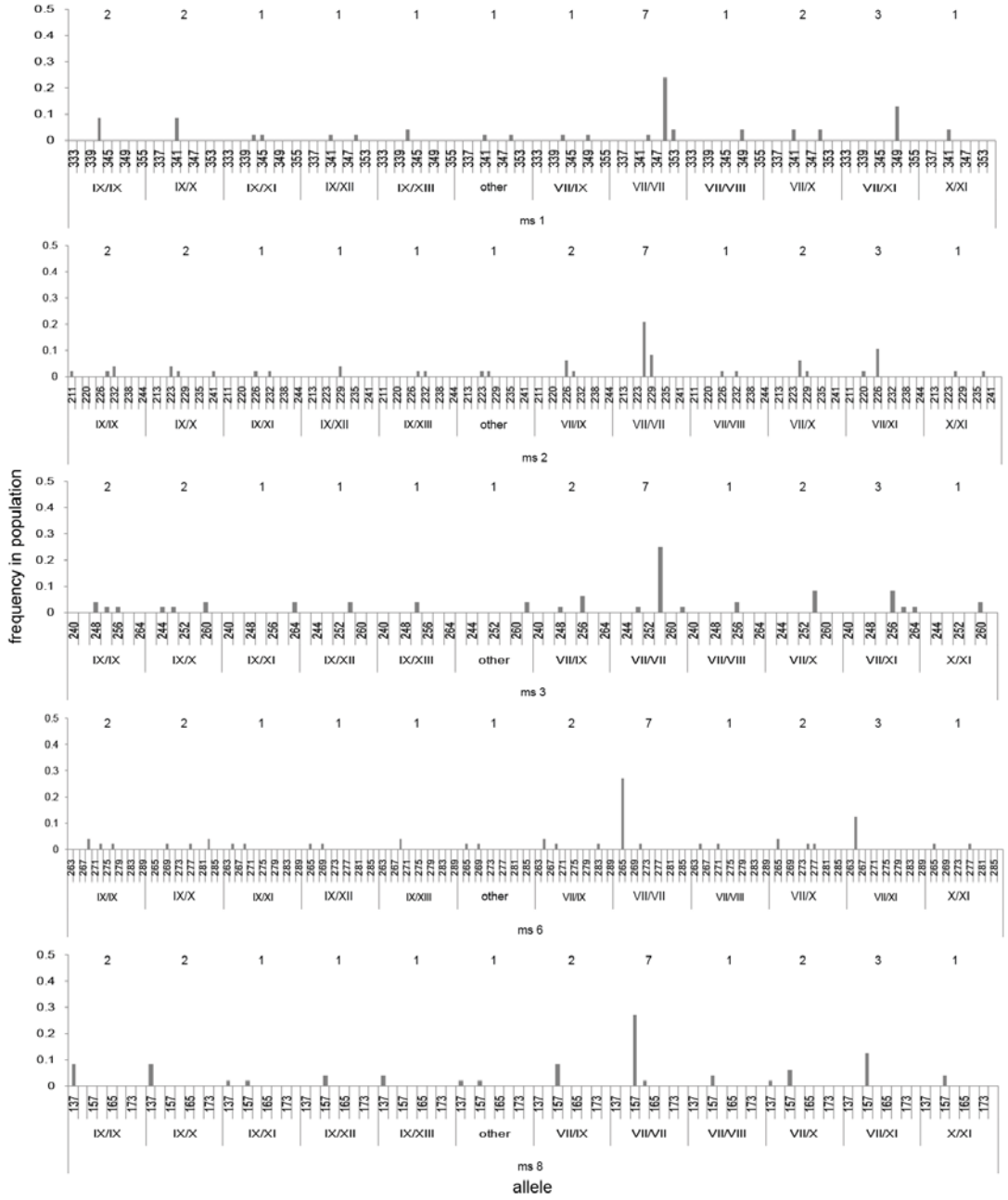
Fra-ES



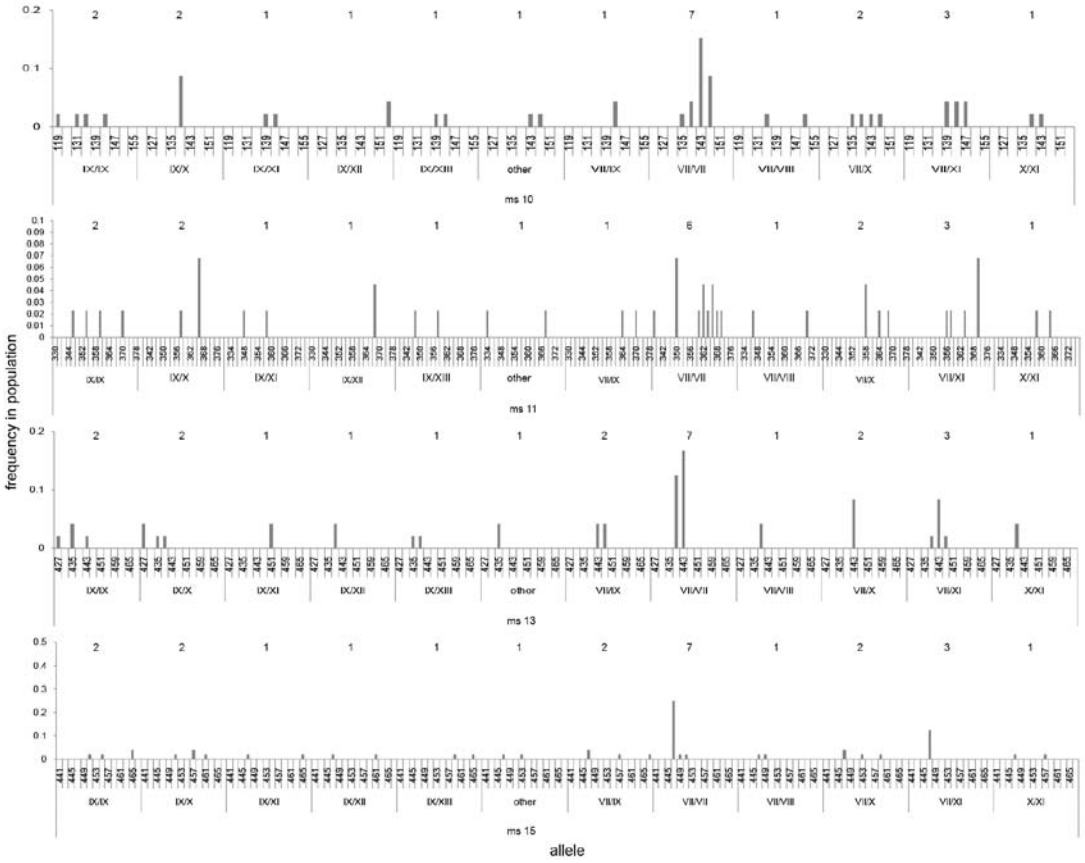
Fra-LO


frequency in population
allele
ms 1
ms 10
ms 2
ms 11
ms 3
ms 13
ms 6
ms 15
ms 8
IX/IX
IX/X
X/X

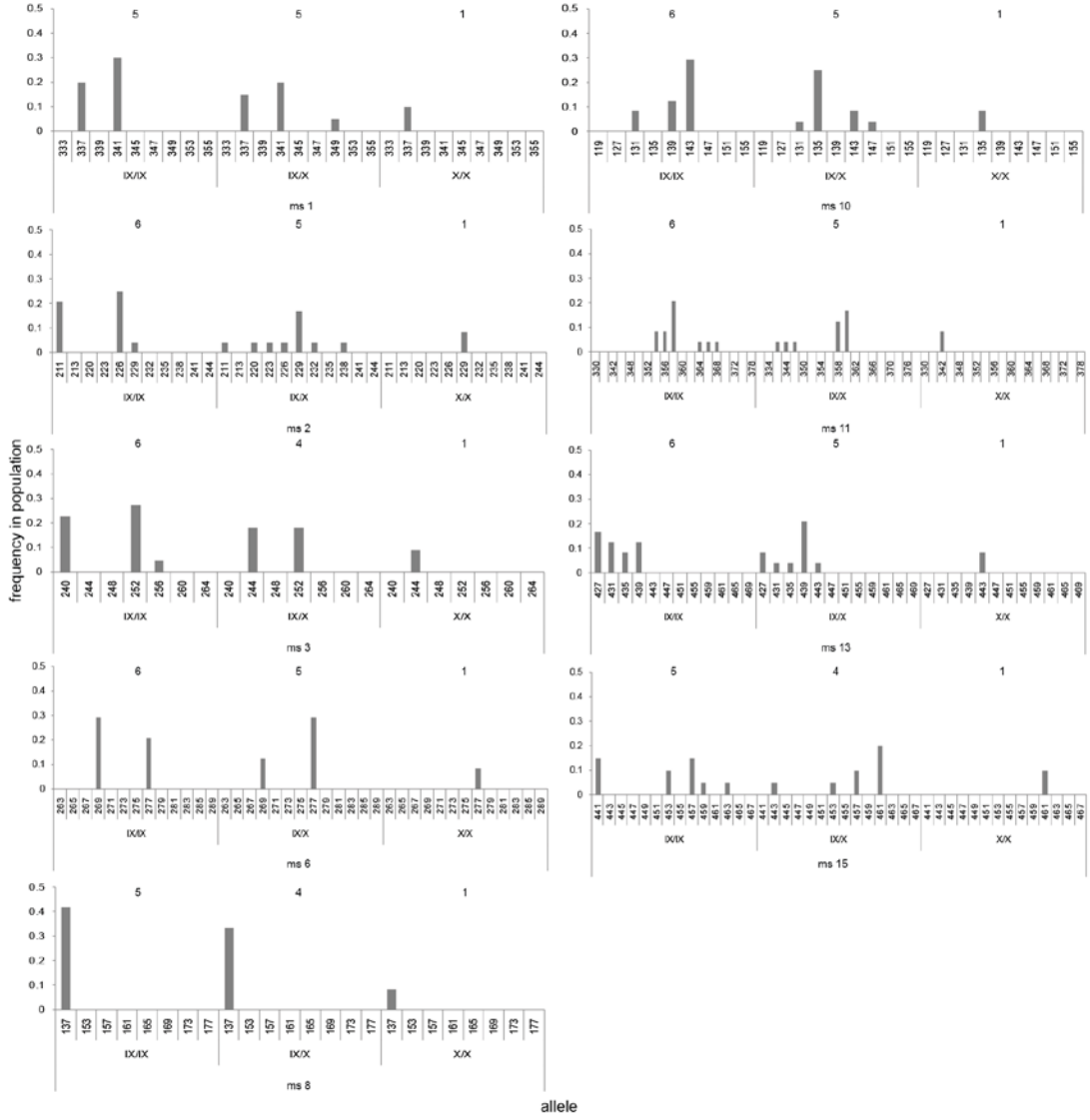
Fra-MC


frequency in population
allele
ms 1
ms 2
ms 3
ms 6
ms 8
IX/IX
IX/X
IX/XI
IX/XII
IX/XIII
other
VII/IX
VII/VII
VII/VIII
VII/X
VII/XI
X/XI

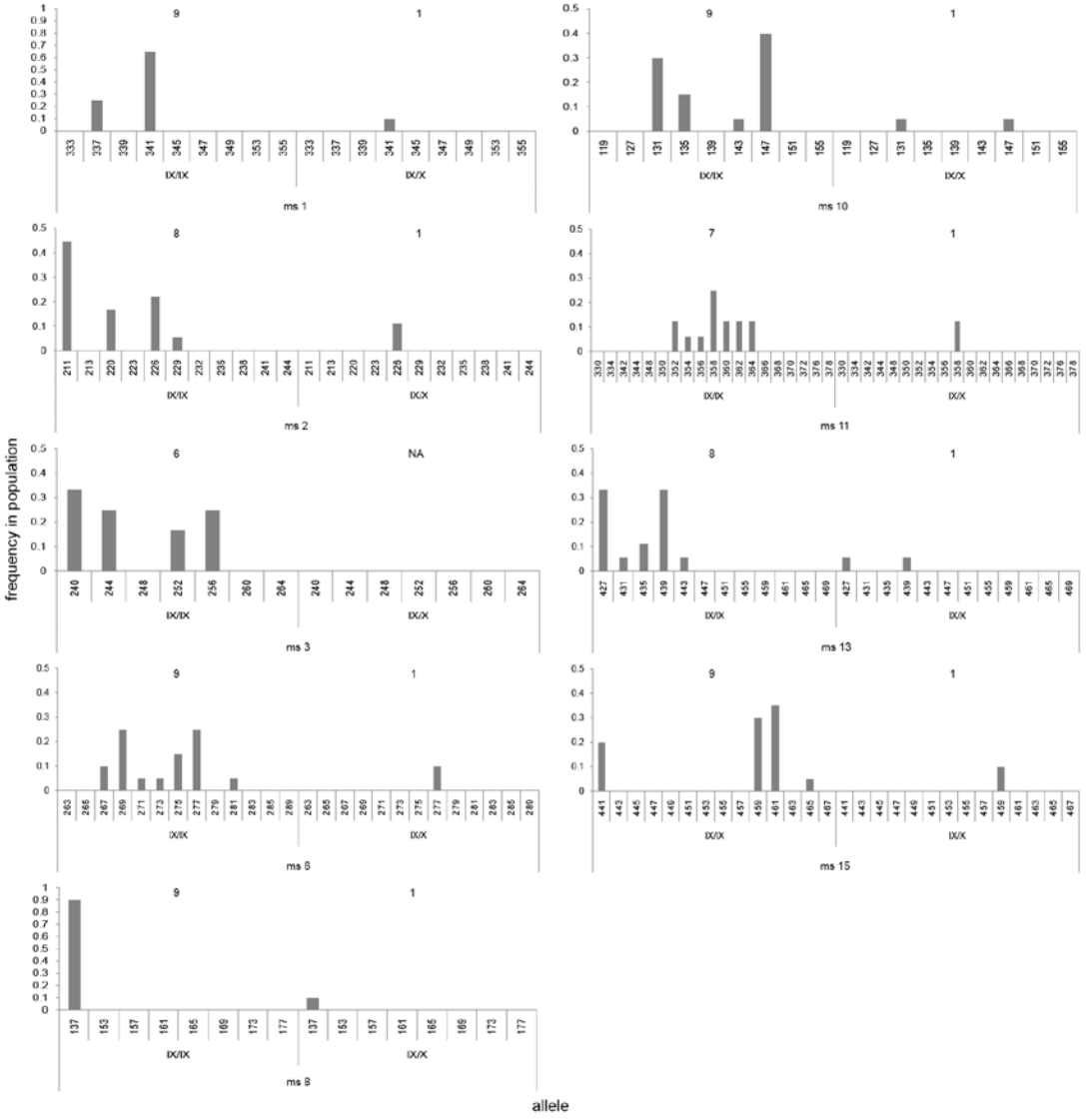
Fra-MC


frequency in population
allele
ms 10
ms 11
ms 13
ms 15
IX/IX
IX/X
IX/XI
IX/XII
IX/XIII
other
VII/IX
VII/VII
VII/VIII
VII/X
VII/XI
X/XI

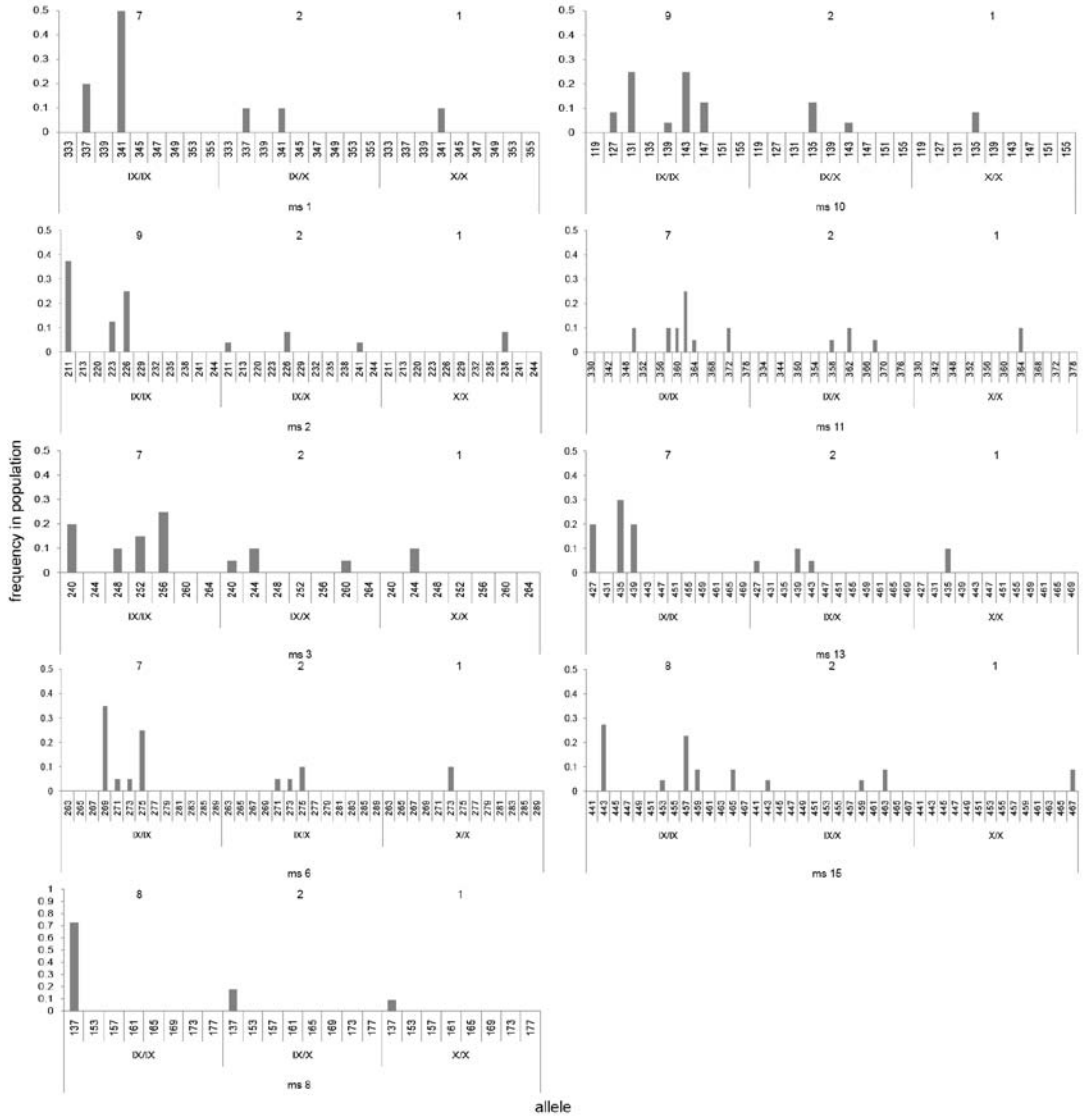
Fra-NA


frequency in population
allele
ms 1
ms 10
ms 2
ms 11
ms 3
ms 13
ms 6
ms 15
ms 8
IX/IX
IX/X
X/X

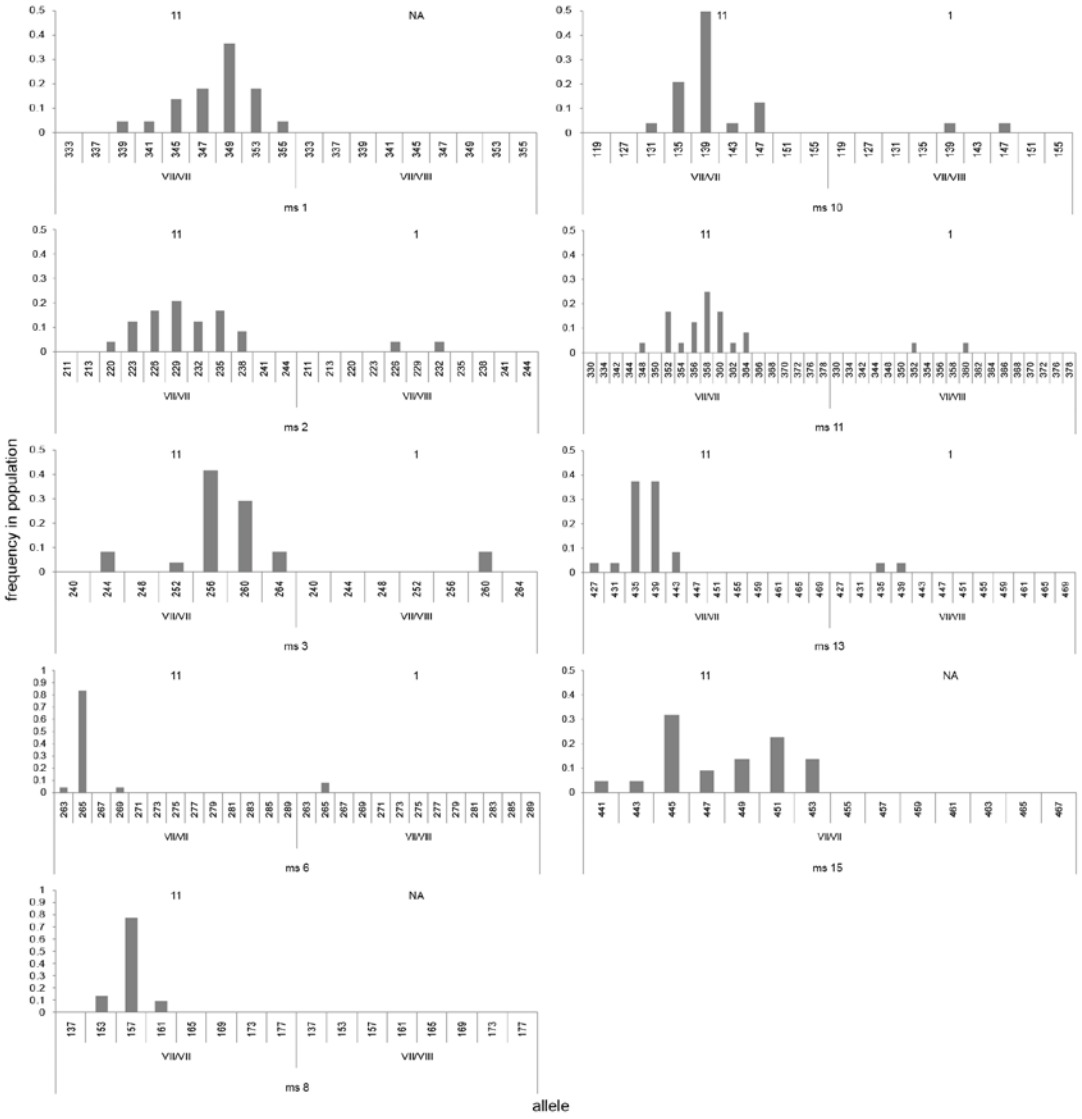
Ger-CB


frequency in population
allele
ms 1
ms 10
ms 2
ms 11
ms 3
ms 13
ms 8
ms 15
ms 9
IX/IX
IX/X

Ger-SL


frequency in population
allele
ms 1
ms 10
ms 2
ms 11
ms 3
ms 13
ms 6
ms 15
ms 8
IX/IX
IX/X
X/X

Dom-IR


frequency in population
allele
ms 1
ms 10
ms 2
ms 11
ms 3
ms 13
ms 6
ms 15
ms 8
VII/VII
VII/VIII

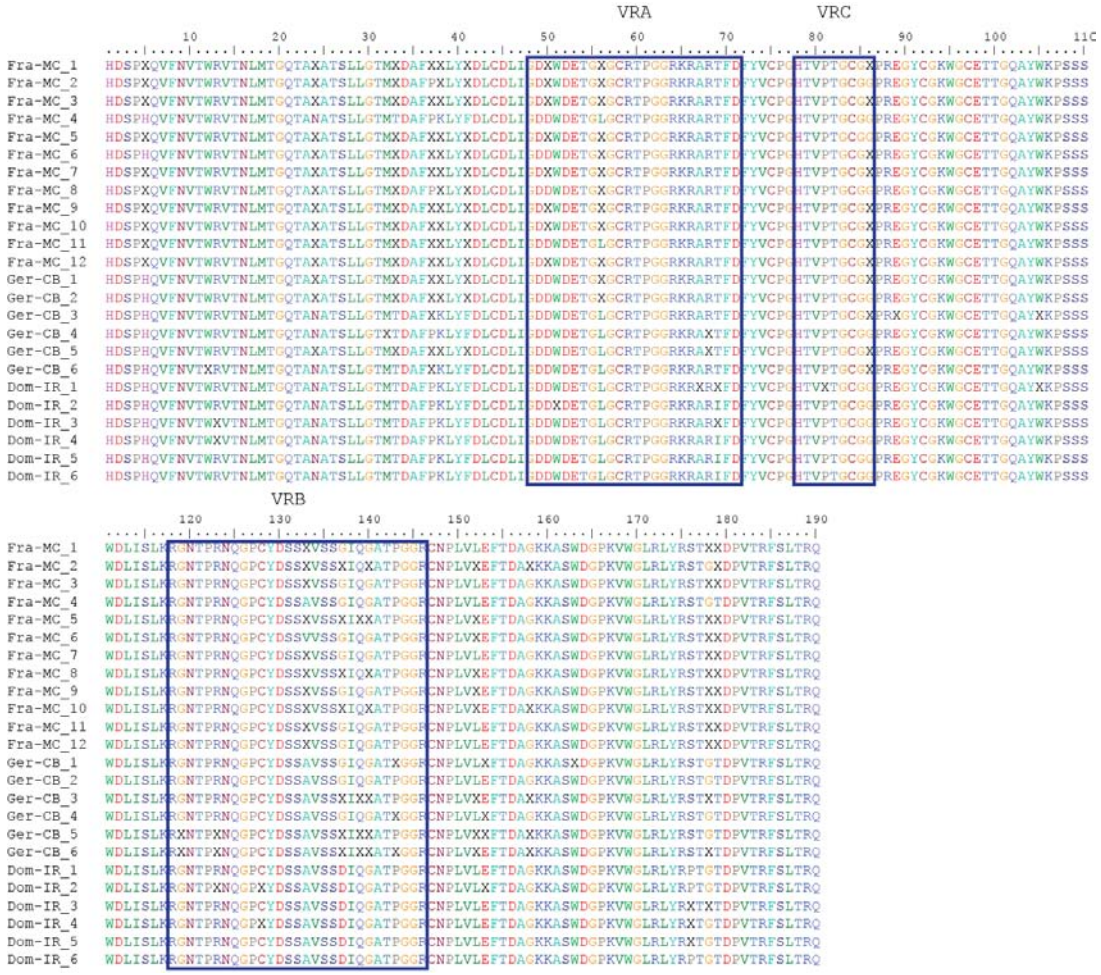


Figure S2.2: Amino acid sequence alignment of the RBD of the sequenced proviruses in captive, wild-derived mice from three different populations (Fra-MC, Ger-CB. Dom-IR). The hypervariable regions VRA and VRB are marked by blue rectangles.

S3 (Chapter 2):

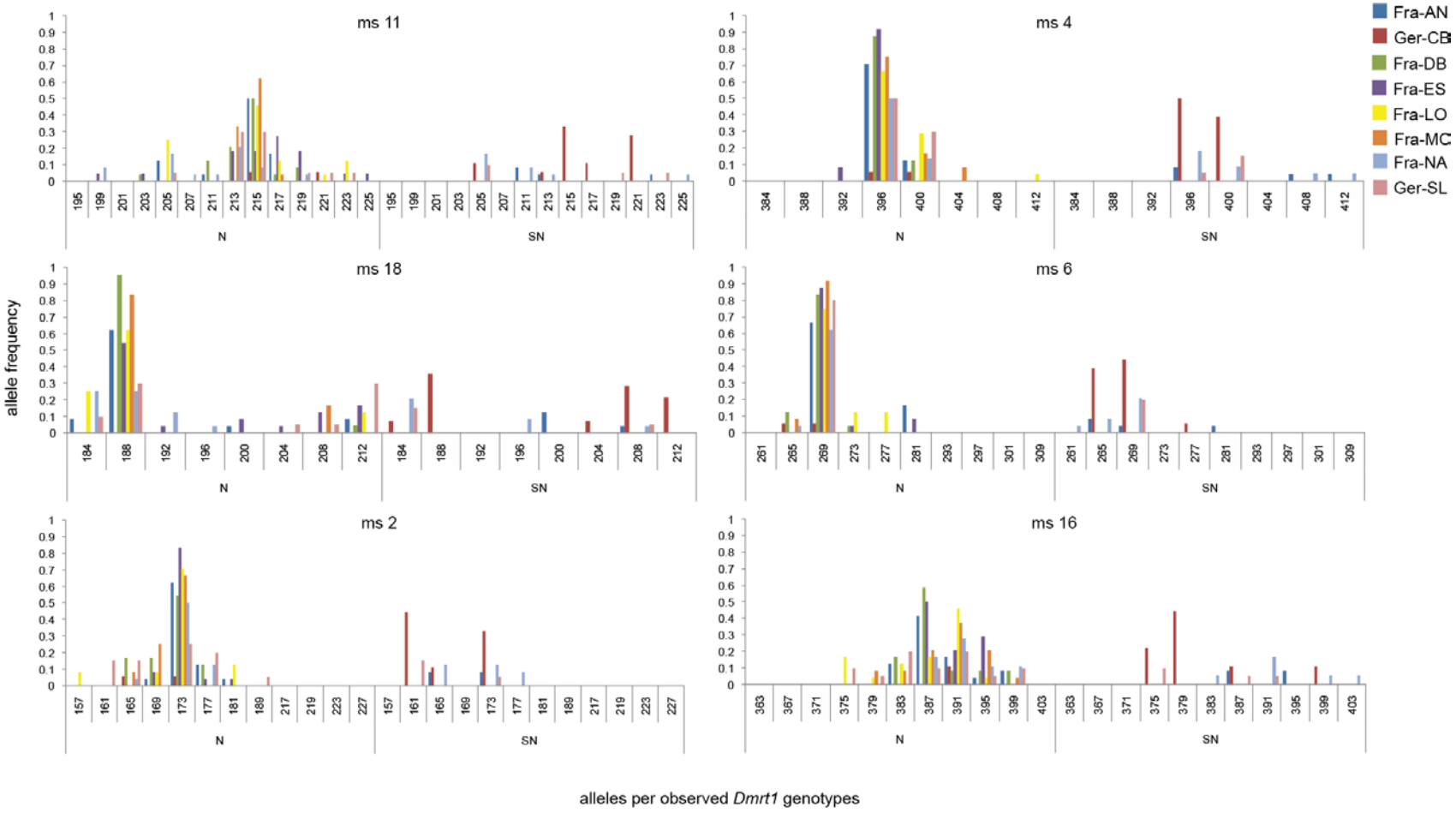

ms 11
ms 4
ms 18
ms 6
ms 2
ms 16
N
SN
allele frequency
alleles per observed *Dmrt1* genotypes
Fra-AN
Ger-CB
Fra-DB
Fra-ES
Fra-LO
Fra-MC
Fra-NA
Ger-SL

Figure S3.1: Microsatellite allele frequency distribution of all loci and for occurring *Dmrt1* genotypes in the European *M. m. domesticus* populations.
The loci are shown in the order in which they are located on the chromosome and each population is represented by a different color. The populations Fra_{AN}, Fra_{DB}, Fra_{ES}, Fra_{LO} and Fra_{MC} exhibited a pattern indicative for a selective sweep with one major allele in high frequency in all loci except for ms 16. This locus is located in the intragenic region of *Dmrt3* which is the closest neighbor to *Dmrt1*. The populations Fra_{NA}, Ger_{SL} and Ger_{CB} did not show signs of a selective sweep as they exhibited several alleles in intermediate frequencies for all loci.

Table S3.1: PCR and sequencing primers for *Dmrt1* coding sequence analysis.

primer ID	sequence
Ex1_F	TATCCAAAAGTCCTGGGCTA
Ex1_R	GGGATTCAAGCAGACGGAAC
Ex2_F	TGAGTGTGCTCTGCTGGAGT
Ex2_R	CACATGCCTACCACCCTCTT
Ex3_F	TGCCCCTTAACCACATCTTT
Ex3_R	TGCATGCGTGTTTCACATATT
Ex4_F	CTGGGGGACGTTTCGTATT
Ex4_R	GCCATCCCAACCCAGACT
Ex5_F	CCAGTGTTCTCGCCCC
Ex5_R	GACGAGTGAGCCGCCGCC

Table S3.2: Primers for microsatellite loci.

primer ID	label	sequence	mean product length	pool
ms2-F	FAM	CCAAAAGGTCTAGACAGGAAGC	173	a
ms2-R		GGCACAGCACTGAGTCCTTC		
ms4-F	FAM	TTCAGATGCACAGTGTCTCCA	399	b
ms4-R		GAAACGTCCCCCAGAGGAT		
ms6-F	HEX	TCTTTTAAGTTGAAGCGGGAAA	273	a
ms6-R		AGTCCCTGGCAGCCATTT		
ms11-F	FAM	CCATTAGCACCTTTTGATGCTCCC	213	c
ms11-R		GGGACCATTCTCACTCAAACCACC		
ms16-F	HEX	GTGCCTTCTGGAAGCTGCTTTGATG	387	d
ms16-R		AATAGTCAGAAGGAGTGGGCAAGG		
ms18-F	HEX	ATGCCCAAGGACTTCACATGAAGG	203	e
ms18-R		GATGGGGAGAAATTACACAATTTCC		

Table S3.3: Transcript levels of *Dmrt1* for different alleles (MCN, N, S, SN), tissues and stages.

Stage	brain				testis			
	MCN	N	S	SN	MCN	N	S	SN
P01	NA	0.093	NA	NA	0.698	0.836	0.757	0.847
P05	NA	NA	NA	NA	1.396	1.609	1.724	1.491
P07	NA	NA	NA	NA	1.579	1.344	1.753	1.910
P09	NA	NA	NA	NA	1.767	1.783	1.991	2.001

Table S3.4.1: List of 1% most differently transcribed genes in brain at P01 between mice carrying the N or S allele of *Dmrt1*.

probe	ensembl_id	gene_id	P01MCNBR	P01NBR	P01SBR	P01SNBR	Diff	*Dmrt1* target	SN intermediate
CUST_6829_PI428259748	ENSMUST00000115535	Zfp944	1.076	3.693	1.220	0.801	2.473	0	0
CUST_4085_PI428259748	ENSMUST00000056449	Arhgap30	0.964	3.214	0.848	0.917	2.366	0	1
CUST_5717_PI428259748	ENSMUST00000075888	Zic2	2.720	3.539	1.703	3.535	1.835	0	1
CUST_5656_PI428259748	ENSMUST00000174379	Dctd	1.011	2.951	1.309	1.114	1.641	0	0
CUST_2586_PI428259748	ENSMUST00000061469	Wfikkn2	0.856	2.321	0.871	0.805	1.450	1	0
CUST_6333_PI428259748	ENSMUST00000110602	Wdr76	0.623	2.050	0.657	0.555	1.393	0	0
CUST_4371_PI428259748	ENSMUST00000109174	Col2a1	2.393	2.913	1.671	2.156	1.242	0	1
CUST_6810_PI428259748	ENSMUST00000073391	Cyp26c1	1.228	2.143	1.136	1.485	1.008	0	1
CUST_1379_PI428259748	ENSMUST00000031378	Stx2	0.380	1.402	0.447	0.450	0.955	0	1
CUST_1471_PI428259748	ENSMUST00000031145	Pdcl2	0.939	1.764	0.818	0.812	0.946	1	0
CUST_4481_PI428259748	ENSMUST00000080754	Pknox2	0.993	2.103	1.194	1.056	0.909	0	0
CUST_2319_PI428259748	ENSMUST00000055688	Phf13	0.603	1.463	0.641	0.567	0.823	1	0
CUST_415_PI428259748	ENSMUST00000010434	AI597479	1.238	1.959	1.163	1.123	0.795	1	0
CUST_2146_PI428259748	ENSMUST00000028794	Siglec1	1.064	1.864	1.094	0.963	0.771	1	0
CUST_609_PI428259748	ENSMUST00000021438	Nova1	2.342	2.780	2.019	3.006	0.761	0	0
CUST_3345_PI428259748	ENSMUST00000044277	Chmp4b	0.805	0.978	0.218	0.820	0.759	0	1
CUST_6342_PI428259748	ENSMUST00000093126	BC107364	1.133	1.853	1.099	1.031	0.754	1	0
CUST_7021_PI428259748	ENSMUST00000093486	Lhx9	1.353	1.908	1.176	1.773	0.732	0	1
CUST_7001_PI428259748	ENSMUST00000155301	Fbxl12	0.590	1.286	0.553	1.056	0.732	0	1
CUST_2835_PI428259748	ENSMUST00000043336	St8sia4	2.096	2.844	2.113	2.509	0.732	0	1
CUST_6579_PI428259748	ENSMUST00000099571	Prmt2	1.816	1.695	0.981	0.797	0.714	0	0
CUST_6806_PI428259748	ENSMUST00000106355	Zfp691	0.656	1.377	0.672	0.851	0.705	0	1
CUST_345_PI428259748	ENSMUST00000006701	Tmem110	0.597	1.227	0.523	0.531	0.704	0	1
CUST_5034_PI428259748	ENSMUST00000112428	Zfp384	0.757	1.337	0.694	0.788	0.643	0	1

probe	ensembl_id	gene_id	P01MCNBR	P01NBR	P01SBR	P01SNBR	Diff	*Dmrt1* target	SN intermediate
CUST_7287_PI428259748	ENSMUST00000066237	Ddc	0.992	1.456	0.818	1.816	0.639	1	0
CUST_1687_PI428259748	ENSMUST00000027020	Gria4	1.418	1.841	1.241	1.980	0.601	0	0
CUST_6736_PI428259748	ENSMUST00000109096	Gnas	1.391	1.054	0.457	0.916	0.597	1	1
CUST_6085_PI428259748	ENSMUST00000119070	Otx2	3.015	3.442	2.850	NA	0.591	0	0
CUST_6799_PI428259748	ENSMUST00000111742	Bcat1	1.607	1.913	1.323	1.895	0.590	1	1
CUST_5978_PI428259748	ENSMUST00000166551	NA	0.577	1.214	0.631	0.752	0.583	0	1
CUST_1748_PI428259748	ENSMUST00000029852	NA	2.445	2.039	1.465	2.158	0.573	0	0
CUST_2324_PI428259748	ENSMUST00000034927	Zic1	NA	3.109	2.538	2.963	0.570	1	1
CUST_7236_PI428259748	ENSMUST00000124310	Bfsp2	0.980	1.284	0.716	1.131	0.568	1	1
CUST_533_PI428259748	ENSMUST00000020308	Ddit4	0.927	1.016	0.448	0.879	0.567	0	1
CUST_7426_PI428259748	ENSMUST00000116408	Hdac7	0.641	1.311	0.751	0.811	0.560	0	1
CUST_2311_PI428259748	ENSMUST00000052377	NA	2.204	1.607	1.059	0.973	0.549	0	0
CUST_6309_PI428259748	ENSMUST00000121805	Dpysl3	2.215	1.642	2.318	2.078	-0.676	1	1
CUST_7035_PI428259748	ENSMUST00000067776	NA	2.045	1.555	2.234	1.769	-0.679	0	1
CUST_6266_PI428259748	ENSMUST00000085519	Anp32a	0.619	0.809	1.494	0.767	-0.685	0	0
CUST_974_PI428259748	ENSMUST00000019117	Hoxb1	0.778	0.779	1.470	1.307	-0.691	0	1
CUST_2200_PI428259748	ENSMUST00000054226	BC049762	0.550	0.643	1.346	0.626	-0.703	1	0
CUST_4581_PI428259748	ENSMUST00000084731	Ipo7	0.912	0.816	1.519	0.781	-0.703	0	0
CUST_5412_PI428259748	ENSMUST00000112097	Rnf10	0.953	0.809	1.520	0.787	-0.711	1	0
CUST_6414_PI428259748	ENSMUST00000170033	Lig4	1.236	1.398	2.110	1.066	-0.712	0	0
CUST_2273_PI428259748	ENSMUST00000060061	Pygo2	1.073	1.099	1.813	0.970	-0.714	0	0
CUST_1716_PI428259748	ENSMUST00000027066	Eya1	1.082	0.902	1.616	1.353	-0.714	1	1
CUST_1659_PI428259748	ENSMUST00000029421	Ptx3	2.517	1.787	2.512	2.233	-0.725	0	1
CUST_6697_PI428259748	ENSMUST00000108589	Dlg4	1.192	0.941	1.667	1.037	-0.726	0	1
CUST_6011_PI428259748	ENSMUST00000074897	Olfr535	1.175	0.995	1.734	1.389	-0.739	0	1

probe	ensembl_id	gene_id	P01MCNBR	P01NBR	P01SBR	P01SNBR	Diff	*Dmrt1* target	SN intermediate
CUST_3519_PI428259748	ENSMUST00000064762	Map1b	1.593	1.803	2.552	1.871	-0.749	0	1
CUST_6504_PI428259748	ENSMUST00000074924	NA	2.153	1.403	2.155	1.583	-0.753	0	1
CUST_1609_PI428259748	ENSMUST00000026472	Inhbc	1.175	0.932	1.700	1.522	-0.768	1	1
CUST_3830_PI428259748	ENSMUST00000059562	Lhfp	0.410	0.913	1.687	0.885	-0.774	0	0
CUST_3515_PI428259748	ENSMUST00000055125	Diras1	1.288	1.303	2.082	1.564	-0.779	1	1
CUST_337_PI428259748	ENSMUST00000006667	Gvin1	0.654	1.051	1.832	1.162	-0.781	0	1
CUST_3116_PI428259748	ENSMUST00000062213	Sfxn3	1.344	1.180	1.974	1.130	-0.795	0	0
CUST_7253_PI428259748	ENSMUST00000164397	B3gnt5	1.887	2.280	3.140	2.618	-0.860	0	1
CUST_5480_PI428259748	ENSMUST00000084677	Gm21093	0.887	0.922	1.801	0.988	-0.879	0	1
CUST_2135_PI428259748	ENSMUST00000025803	Dkk1	1.130	0.872	1.763	1.040	-0.892	0	1
CUST_6990_PI428259748	ENSMUST00000111824	Cwc22	NA	0.736	1.629	0.676	-0.894	1	0
CUST_1963_PI428259748	ENSMUST00000029141	Mmp24	2.349	1.970	2.874	2.098	-0.905	1	1
CUST_6079_PI428259748	ENSMUST00000116362	Gm11678	NA	0.503	1.414	0.754	-0.912	0	1
CUST_4870_PI428259748	ENSMUST00000103101	Pigt	0.813	1.048	2.003	0.895	-0.955	0	0
CUST_5004_PI428259748	ENSMUST00000112104	Mid1	0.932	0.792	1.770	0.743	-0.978	0	0
CUST_6361_PI428259748	ENSMUST00000096121	Gm5458	0.539	1.221	2.240	2.362	-1.019	0	0
CUST_165_PI428259748	ENSMUST00000005103	Nfe2l3	2.070	2.204	3.285	2.291	-1.082	1	1
CUST_6501_PI428259748	ENSMUST00000095456	Mcf2l	2.199	1.929	3.036	2.067	-1.107	0	1
CUST_2920_PI428259748	ENSMUST00000051676	Fam46b	0.546	0.510	1.631	0.738	-1.121	0	1
CUST_5742_PI428259748	ENSMUST00000167140	Mpst	1.389	0.894	2.021	0.664	-1.127	1	0
CUST_1572_PI428259748	ENSMUST00000025109	Sap130	0.753	0.839	2.101	0.867	-1.262	1	1
CUST_6387_PI428259748	ENSMUST00000070797	Pcdhac1	1.345	1.257	2.607	1.619	-1.350	1	1
CUST_4515_PI428259748	ENSMUST00000119336	Ankrd24	0.893	0.909	2.894	0.643	-1.986	0	0

Table S3.4.2: List of 1% most differently transcribed genes in testis at P01 between mice carrying the N or S allele of *Dmrt1*.

probe	ensembl_id	gene_id	P01MCNTE	P01NTE	P01STE	P01SNTE	Diff	Dmrt1 target	SN intermediate
CUST_808_PI428259748	ENSMUST00000020185	Il20ra	0.753	2.073	0.701	0.666	1.372	1	0
CUST_5086_PI428259748	ENSMUST00000119892	Gm15800	0.574	2.019	0.668	0.911	1.351	0	1
CUST_4436_PI428259748	ENSMUST00000166074	Htr7	0.433	1.694	0.444	0.572	1.250	1	1
CUST_6923_PI428259748	ENSMUST00000114471	Cblb	1.034	2.204	1.038	1.022	1.166	0	0
CUST_6979_PI428259748	ENSMUST00000068923	Adpgk	0.607	1.920	0.761	0.925	1.159	0	1
CUST_7229_PI428259748	ENSMUST00000107885	Akt1s1	0.879	2.201	1.057	1.070	1.144	0	1
CUST_3736_PI428259748	ENSMUST00000033098	Bcat2	1.020	1.572	0.462	1.424	1.110	1	1
CUST_6536_PI428259748	ENSMUST00000068068	1700041G16Rik	1.883	2.177	1.096	0.918	1.080	1	0
CUST_3912_PI428259748	ENSMUST00000064795	Egr1	1.188	1.995	0.923	0.901	1.072	0	0
CUST_2372_PI428259748	ENSMUST00000049572	Lipa	0.683	1.515	0.493	0.936	1.022	0	1
CUST_2695_PI428259748	ENSMUST00000034795	NA	0.771	1.694	0.700	0.714	0.993	0	1
CUST_441_PI428259748	ENSMUST00000007301	Myh3	0.700	1.909	0.925	NA	0.984	0	0
CUST_7108_PI428259748	ENSMUST00000086795	Barhl2	0.865	1.884	0.921	0.844	0.963	0	0
CUST_3041_PI428259748	ENSMUST00000057090	Synpo2l	0.751	1.621	0.703	0.891	0.919	0	1
CUST_5952_PI428259748	ENSMUST00000109309	Shank3	0.531	1.419	0.546	0.605	0.873	0	1
CUST_2875_PI428259748	ENSMUST00000057768	4930429B21Rik	0.564	1.521	0.661	0.720	0.860	1	1
CUST_6823_PI428259748	ENSMUST00000156895	Ifi203	2.967	1.746	0.917	2.286	0.830	0	0
CUST_4506_PI428259748	ENSMUST00000078816	9130023H24Rik	0.848	1.818	0.989	0.964	0.829	0	0
CUST_4047_PI428259748	ENSMUST00000038760	Lad1	1.255	2.016	1.189	1.118	0.827	0	0
CUST_4701_PI428259748	ENSMUST00000077960	H2-T22	1.115	1.972	1.148	1.632	0.824	1	1
CUST_4255_PI428259748	ENSMUST00000062241	Tlr9	0.988	1.716	0.950	1.150	0.766	1	1
CUST_4686_PI428259748	ENSMUST00000095795	Raet1d	2.594	2.598	1.841	2.404	0.757	1	1
CUST_6945_PI428259748	ENSMUST00000132795	Tnni1	2.153	2.853	2.099	3.315	0.755	0	0
CUST_115_PI428259748	ENSMUST00000001544	Vwa5a	1.160	2.390	1.653	1.652	0.737	0	0

probe	ensembl_id	gene_id	P01MCNTE	P01NTE	P01STE	P01SNTE	Diff	Dmrt1 target	SN intermediate
CUST_1968_PI428259748	ENSMUST00000025271	Pou5f1	2.810	2.377	1.657	1.690	0.720	0	1
CUST_1763_PI428259748	ENSMUST00000025197	Tap2	0.475	1.147	0.474	0.684	0.672	0	1
CUST_6357_PI428259748	ENSMUST00000093935	Krtap31-1	1.158	1.493	0.823	1.024	0.670	1	1
CUST_4957_PI428259748	ENSMUST00000066134	Rpia	2.020	1.224	0.586	0.549	0.638	0	0
CUST_4730_PI428259748	ENSMUST00000079012	8-Mar	0.662	1.139	0.535	0.661	0.603	0	1
CUST_903_PI428259748	ENSMUST00000021246	Snx11	0.515	1.121	0.538	0.538	0.584	0	1
CUST_3481_PI428259748	ENSMUST00000046689	2810474O19Rik	1.272	1.513	0.933	0.860	0.580	0	0
CUST_5525_PI428259748	ENSMUST00000139725	Mitd1	1.279	1.228	0.651	0.882	0.577	0	1
CUST_5878_PI428259748	ENSMUST00000139560	Tmem48	1.468	1.489	0.915	0.935	0.575	0	1
CUST_5187_PI428259748	ENSMUST00000174147	Amy1	0.552	0.902	0.332	0.329	0.571	1	0
CUST_1420_PI428259748	ENSMUST00000025025	Dusp1	0.806	1.272	0.703	0.710	0.569	0	1
CUST_2109_PI428259748	ENSMUST00000024748	Gtpbp2	1.374	2.614	2.050	2.591	0.563	0	1
CUST_6684_PI428259748	ENSMUST00000066529	Npr3	1.439	1.532	1.988	1.927	-0.456	0	1
CUST_2651_PI428259748	ENSMUST00000040437	Spaca5	1.185	1.022	1.488	1.380	-0.466	0	1
CUST_4657_PI428259748	ENSMUST00000072228	Gckr	1.501	0.983	1.453	1.177	-0.470	1	1
CUST_1902_PI428259748	ENSMUST00000027885	Angptl1	0.700	1.291	1.770	1.970	-0.479	0	0
CUST_1716_PI428259748	ENSMUST00000027066	Eya1	0.840	0.855	1.337	1.345	-0.482	1	0
CUST_6506_PI428259748	ENSMUST00000092162	Lyz1	1.252	1.099	1.587	1.234	-0.488	0	1
CUST_1744_PI428259748	ENSMUST00000028854	Mal	0.590	0.650	1.148	0.669	-0.498	0	1
CUST_2891_PI428259748	ENSMUST00000040104	Hand2	2.485	2.606	3.108	2.219	-0.502	0	0
CUST_5508_PI428259748	ENSMUST00000121043	Hoxa10	1.349	1.270	1.775	2.194	-0.505	0	0
CUST_2135_PI428259748	ENSMUST00000025803	Dkk1	0.659	0.993	1.501	1.372	-0.509	0	1
CUST_6806_PI428259748	ENSMUST00000106355	Zfp691	1.032	0.797	1.306	0.905	-0.509	0	1
CUST_1001_PI428259748	ENSMUST00000022536	Ska3	1.576	1.338	1.857	1.406	-0.519	0	1
CUST_7076_PI428259748	ENSMUST00000099494	Btaf1	0.538	0.578	1.108	0.777	-0.530	0	1

probe	ensembl_id	gene_id	P01MCNTE	P01NTE	P01STE	P01SNTE	Diff	Dmrt1 target	SN intermediate
CUST_3814_PI428259748	ENSMUST00000050952	Stbd1	2.016	1.527	2.072	1.422	-0.545	0	0
CUST_6942_PI428259748	ENSMUST00000098090	Coq7	1.020	1.207	1.766	1.309	-0.559	0	1
CUST_1156_PI428259748	ENSMUST00000023265	Psca	0.582	0.396	0.958	0.426	-0.562	0	1
CUST_974_PI428259748	ENSMUST00000019117	Hoxb1	1.147	0.868	1.441	1.239	-0.574	0	1
CUST_1624_PI428259748	ENSMUST00000027231	Slc9a2	0.531	0.492	1.068	0.684	-0.576	1	1
CUST_1559_PI428259748	ENSMUST00000024032	9530002B09Rik	0.786	0.583	1.175	0.741	-0.592	1	1
CUST_171_PI428259748	ENSMUST00000003574	Cyp4f18	2.153	1.377	1.989	1.545	-0.612	0	1
CUST_1322_PI428259748	ENSMUST00000030041	Ambp	0.685	0.971	1.596	1.677	-0.626	1	0
CUST_2740_PI428259748	ENSMUST00000039559	Thbs1	1.168	1.097	1.726	1.163	-0.629	0	1
CUST_7384_PI428259748	ENSMUST00000161178	Fam109b	2.223	1.590	2.224	1.674	-0.634	0	1
CUST_649_PI428259748	ENSMUST00000021682	Angel1	0.929	0.785	1.491	1.027	-0.706	0	1
CUST_5683_PI428259748	ENSMUST00000100420	Loxl2	2.222	1.744	2.453	2.748	-0.709	0	0
CUST_1649_PI428259748	ENSMUST00000027695	Slc45a3	1.940	1.346	2.060	1.633	-0.714	0	1
CUST_7183_PI428259748	ENSMUST00000111751	Myl2	1.295	0.395	1.142	2.440	-0.746	0	0
CUST_5696_PI428259748	ENSMUST00000085073	Actl11	0.751	0.843	1.602	1.311	-0.758	1	1
CUST_6864_PI428259748	ENSMUST00000140490	Steap3	0.904	0.969	1.732	1.060	-0.763	0	1
CUST_606_PI428259748	ENSMUST00000018765	Mmp8	1.399	1.413	2.180	NA	-0.767	0	0
CUST_2248_PI428259748	ENSMUST00000036618	Stab1	2.493	2.009	2.778	2.436	-0.769	0	1
CUST_4301_PI428259748	ENSMUST00000037685	Spt1	1.028	0.675	1.453	2.691	-0.778	0	0
CUST_2639_PI428259748	ENSMUST00000053871	Ckap4	1.903	1.581	2.373	1.437	-0.792	0	0
CUST_3413_PI428259748	ENSMUST00000064725	Serpinc1	NA	1.951	2.778	2.149	-0.826	0	1
CUST_3143_PI428259748	ENSMUST00000039413	Pde4a	0.374	0.518	1.447	0.616	-0.929	1	1
CUST_1427_PI428259748	ENSMUST00000031243	Spp1	0.325	0.958	2.340	1.802	-1.382	1	1

Table S3.4.3: List of 1% most differently transcribed genes in brain at P05 between mice carrying the N or S allele of *Dmrt1*.

probe	ensembl_id	gene_id	P05MCNBR	P05NBR	P05SBR	P05SNBR	Diff	Dmrt1 target	SN intermediate
CUST_6417_PI428259748	ENSMUST00000168501	Phf8	1.183	3.731	0.850	0.981	2.881	0	1
CUST_6102_PI428259748	ENSMUST00000102921	Myl7	0.710	3.222	0.665	0.688	2.557	0	1
CUST_6848_PI428259748	ENSMUST00000075221	Osm	0.523	3.098	0.565	0.605	2.533	1	1
CUST_1261_PI428259748	ENSMUST00000023601	St6gal1	0.448	2.815	0.407	0.471	2.408	0	1
CUST_90_PI428259748	ENSMUST00000001583	Ell2	0.450	2.558	0.395	0.438	2.163	0	1
CUST_6739_PI428259748	ENSMUST00000150235	Crem	0.921	3.159	1.049	1.076	2.111	0	1
CUST_6541_PI428259748	ENSMUST00000165307	Anp32e	0.296	2.337	0.242	0.225	2.094	1	0
CUST_728_PI428259748	ENSMUST00000019721	Pdk4	1.168	2.966	0.919	1.161	2.047	0	1
CUST_1166_PI428259748	ENSMUST00000027059	Tfap2b	0.444	2.278	0.233	1.775	2.044	0	1
CUST_6822_PI428259748	ENSMUST00000112242	Kcna6	1.290	3.059	1.153	1.324	1.907	1	1
CUST_6167_PI428259748	ENSMUST00000091252	Sec16a	0.594	2.190	0.585	0.612	1.604	0	1
CUST_4336_PI428259748	ENSMUST00000121979	Cep135	0.418	1.990	0.461	0.580	1.528	0	1
CUST_7340_PI428259748	ENSMUST00000176204	Peg10	0.829	2.296	0.827	0.837	1.469	0	1
CUST_46_PI428259748	ENSMUST00000000895	Necab3	1.309	2.490	1.022	1.365	1.468	0	1
CUST_2168_PI428259748	ENSMUST00000026021	Msr1	0.456	1.659	0.256	0.355	1.403	1	1
CUST_2208_PI428259748	ENSMUST00000039994	Cep250	0.656	2.101	0.749	0.675	1.352	0	0
CUST_36_PI428259748	ENSMUST00000000579	Sox9	0.515	1.780	0.517	0.535	1.263	0	1
CUST_6090_PI428259748	ENSMUST00000075422	Mab21l1	1.027	2.623	1.361	2.186	1.262	0	1
CUST_5260_PI428259748	ENSMUST00000105726	NA	0.142	1.826	0.567	0.180	1.259	0	0
CUST_7229_PI428259748	ENSMUST00000107885	Akt1s1	0.844	1.829	0.587	0.905	1.241	0	1
CUST_4106_PI428259748	ENSMUST00000034184	Irx5	0.892	2.156	0.929	1.276	1.227	0	1
CUST_3318_PI428259748	ENSMUST00000046322	Fcrla	0.699	1.802	0.600	0.641	1.201	0	1
CUST_6810_PI428259748	ENSMUST00000073391	Cyp26c1	3.199	1.918	0.829	1.227	1.089	0	1
CUST_5232_PI428259748	ENSMUST00000077548	Cttnbp2nl	0.884	1.837	0.834	0.974	1.003	0	1

probe	ensembl_id	gene_id	P05MCNBR	P05NBR	P05SBR	P05SNBR	Diff	Dmrt1 target	SN intermediate
CUST_6536_PI428259748	ENSMUST00000068068	1700041G16Rik	1.389	2.114	1.125	1.068	0.989	1	0
CUST_5007_PI428259748	ENSMUST00000103157	Gip	1.592	1.951	0.985	1.055	0.966	0	1
CUST_5713_PI428259748	ENSMUST00000072837	Foxk1	0.861	1.791	0.829	0.730	0.962	0	0
CUST_6351_PI428259748	ENSMUST00000117102	Fzd10	0.609	1.542	0.585	0.600	0.957	0	1
CUST_1602_PI428259748	ENSMUST00000026360	Itgb8	0.989	1.610	0.683	0.741	0.927	0	1
CUST_1633_PI428259748	ENSMUST00000029919	Clca3	0.677	1.579	0.685	0.895	0.894	1	1
CUST_3155_PI428259748	ENSMUST00000050887	Prokr1	1.176	1.609	0.724	0.695	0.885	1	0
CUST_4514_PI428259748	ENSMUST00000091851	Mapk4	0.616	1.427	0.562	0.651	0.865	1	1
CUST_2625_PI428259748	ENSMUST00000050544	Has2	0.897	1.390	0.528	0.704	0.862	0	1
CUST_6085_PI428259748	ENSMUST00000119070	Otx2	0.976	2.030	1.199	1.600	0.830	0	1
CUST_6768_PI428259748	ENSMUST00000072271	Agxt2l1	1.296	1.976	1.148	1.173	0.828	0	1
CUST_2692_PI428259748	ENSMUST00000061496	Tcf7l2	0.982	1.997	1.173	0.858	0.824	0	0
CUST_717_PI428259748	ENSMUST00000021333	Foxg1	1.870	1.416	1.799	1.820	-0.383	0	0
CUST_4250_PI428259748	ENSMUST00000044672	Cdk19	1.106	0.744	1.130	1.082	-0.386	1	1
CUST_4506_PI428259748	ENSMUST00000078816	9130023H24Rik	0.963	0.684	1.073	0.944	-0.389	0	1
CUST_5717_PI428259748	ENSMUST00000075888	Zic2	1.086	1.894	2.283	1.754	-0.389	0	0
CUST_5696_PI428259748	ENSMUST00000085073	Actl11	1.150	0.999	1.390	1.430	-0.391	1	0
CUST_5669_PI428259748	ENSMUST00000102655	Pde1a	1.427	1.222	1.616	1.956	-0.395	1	0
CUST_3554_PI428259748	ENSMUST00000045628	R3hdm4	1.331	1.219	1.616	1.284	-0.397	0	1
CUST_4738_PI428259748	ENSMUST00000133595	Jmjd8	1.237	0.822	1.222	1.060	-0.401	0	1
CUST_3154_PI428259748	ENSMUST00000042692	Tcp11	0.959	0.920	1.328	1.026	-0.408	1	1
CUST_1716_PI428259748	ENSMUST00000027066	Eya1	1.227	0.966	1.377	1.824	-0.411	1	0
CUST_5270_PI428259748	ENSMUST00000107374	NA	1.633	1.312	1.724	1.884	-0.411	0	0
CUST_4200_PI428259748	ENSMUST00000061895	Pisd	1.377	0.791	1.210	1.072	-0.419	0	1
CUST_3350_PI428259748	ENSMUST00000060918	Amz1	1.437	1.725	2.150	1.819	-0.425	0	1

probe	ensembl_id	gene_id	P05MCNBR	P05NBR	P05SBR	P05SNBR	Diff	Dmrt1 target	SN intermediate
CUST_3189_PI428259748	ENSMUST00000035061	Ngp	0.982	0.705	1.144	1.171	-0.439	0	0
CUST_1624_PI428259748	ENSMUST00000027231	Slc9a2	1.095	0.875	1.316	1.369	-0.442	1	0
CUST_3747_PI428259748	ENSMUST00000043042	Tmco4	0.438	0.410	0.856	0.496	-0.445	1	1
CUST_5829_PI428259748	ENSMUST00000108681	Gas7	1.025	0.950	1.397	1.126	-0.447	0	1
CUST_5632_PI428259748	ENSMUST00000114251	Ptgds	1.178	1.836	2.288	2.144	-0.452	0	1
CUST_3694_PI428259748	ENSMUST00000038104	Bdp1	0.678	0.461	0.938	1.105	-0.477	0	0
CUST_5108_PI428259748	ENSMUST00000111089	Nefm	1.232	0.849	1.337	1.299	-0.489	0	1
CUST_268_PI428259748	ENSMUST00000014640	Ankrd28	1.481	1.706	2.196	1.695	-0.490	1	0
CUST_6079_PI428259748	ENSMUST00000116362	Gm11678	0.133	0.385	0.909	2.080	-0.525	0	0
CUST_3422_PI428259748	ENSMUST00000044833	Oas3	1.160	1.008	1.544	0.947	-0.536	0	0
CUST_1569_PI428259748	ENSMUST00000027797	Nvl	0.662	0.834	1.398	0.988	-0.564	0	1
CUST_196_PI428259748	ENSMUST00000002925	Timmdc1	0.810	0.555	1.122	1.131	-0.568	0	0
CUST_1659_PI428259748	ENSMUST00000029421	Ptx3	1.454	1.010	1.588	1.616	-0.578	0	0
CUST_2301_PI428259748	ENSMUST00000041331	Scd1	1.088	0.347	0.926	0.613	-0.578	0	1
CUST_6331_PI428259748	ENSMUST00000069922	Mia3	1.175	1.162	1.768	1.739	-0.605	0	1
CUST_6990_PI428259748	ENSMUST00000111824	Cwc22	NA	0.463	1.115	1.719	-0.652	1	0
CUST_2390_PI428259748	ENSMUST00000037057	Zfp40	0.832	0.939	1.610	1.541	-0.671	0	1
CUST_1770_PI428259748	ENSMUST00000024832	Rsph1	0.552	0.620	1.339	0.518	-0.719	0	0
CUST_4301_PI428259748	ENSMUST00000037685	Spt1	1.576	0.815	1.540	1.565	-0.725	0	0
CUST_3005_PI428259748	ENSMUST00000041099	Neurod1	2.068	1.125	1.855	2.244	-0.730	0	0
CUST_6011_PI428259748	ENSMUST00000074897	Olfr535	1.501	0.838	1.604	1.225	-0.766	0	1
CUST_5530_PI428259748	ENSMUST00000107894	Dgke	1.169	1.139	2.046	1.215	-0.907	0	1
CUST_1383_PI428259748	ENSMUST00000030489	Tal1	1.086	1.527	2.447	1.437	-0.921	0	0

Table S3.4.4: List of 1% most differently transcribed genes in testis at P05 between mice carrying the N or S allele of *Dmrt1*.

probe	ensembl_id	gene_id	P05MCNTE	P05NTE	P05STE	P05SNTE	Diff	Dmrt1 target	SN intermediate
CUST_3747_PI428259748	ENSMUST00000043042	Tmco4	1.045	2.435	1.190	1.268	1.246	1	1
CUST_3901_PI428259748	ENSMUST00000062545	Olfr958	1.157	2.245	1.121	1.986	1.124	1	1
CUST_7408_PI428259748	ENSMUST00000090940	Ermn	0.736	1.720	0.638	0.697	1.082	1	1
CUST_2625_PI428259748	ENSMUST00000050544	Has2	1.035	2.110	1.101	1.071	1.009	0	0
CUST_2149_PI428259748	ENSMUST00000028931	Cst8	1.799	1.764	0.828	1.147	0.937	0	1
CUST_6823_PI428259748	ENSMUST00000156895	Ifi203	NA	1.903	0.984	1.756	0.919	0	1
CUST_4287_PI428259748	ENSMUST00000050011	Gbp6	0.777	1.976	1.059	1.018	0.917	1	0
CUST_5598_PI428259748	ENSMUST00000098966	A430078G23Rik	1.148	1.993	1.175	1.417	0.818	1	1
CUST_3985_PI428259748	ENSMUST00000058856	Scd4	0.884	1.953	1.170	1.564	0.783	0	1
CUST_4701_PI428259748	ENSMUST00000077960	H2-T22	1.891	1.799	1.026	1.567	0.773	1	1
CUST_1925_PI428259748	ENSMUST00000028935	Cst9	2.283	1.234	0.469	0.939	0.766	1	1
CUST_1388_PI428259748	ENSMUST00000032198	Usp18	0.488	1.408	0.647	1.445	0.761	1	0
CUST_7154_PI428259748	ENSMUST00000115598	NA	1.135	1.772	1.016	1.193	0.756	0	1
CUST_271_PI428259748	ENSMUST00000009340	Mnda	2.530	1.217	0.468	0.982	0.749	0	1
CUST_3736_PI428259748	ENSMUST00000033098	Bcat2	0.982	1.909	1.166	1.742	0.743	1	1
CUST_6536_PI428259748	ENSMUST00000068068	1700041G16Rik	1.493	1.958	1.231	1.064	0.727	1	0
CUST_6051_PI428259748	ENSMUST00000100097	NA	1.453	1.577	0.865	1.453	0.711	0	1
CUST_601_PI428259748	ENSMUST00000022629	Dpysl2	0.238	0.904	0.196	0.203	0.708	0	1
CUST_2731_PI428259748	ENSMUST00000040271	Cep85	1.114	1.849	1.156	1.149	0.693	0	0
CUST_3155_PI428259748	ENSMUST00000050887	Prokr1	0.981	1.912	1.239	1.358	0.673	1	1
CUST_1508_PI428259748	ENSMUST00000025181	H2-K1	0.251	1.614	0.945	1.267	0.670	0	1
CUST_4968_PI428259748	ENSMUST00000092966	Dynlt1c	0.717	1.081	0.442	0.938	0.639	0	1
CUST_40_PI428259748	ENSMUST00000001092	Zfp276	0.822	1.555	0.946	0.865	0.609	0	0
CUST_1499_PI428259748	ENSMUST00000023610	Adamts1	1.481	1.405	0.806	0.882	0.599	0	1

probe	ensembl_id	gene_id	P05MCNTE	P05NTE	P05STE	P05SNTE	Diff	Dmrt1 target	SN intermediate
CUST_6740_PI428259748	ENSMUST00000118539	Cth	0.966	1.330	0.767	0.915	0.563	0	1
CUST_1602_PI428259748	ENSMUST00000026360	Itgb8	0.567	1.092	0.533	0.544	0.558	0	1
CUST_5692_PI428259748	ENSMUST00000111241	NA	0.708	1.313	0.758	1.130	0.555	0	1
CUST_429_PI428259748	ENSMUST00000017153	Sdc4	0.850	1.926	1.371	1.635	0.554	0	1
CUST_1636_PI428259748	ENSMUST00000031542	Oasl2	0.793	1.690	1.162	1.431	0.528	0	1
CUST_2213_PI428259748	ENSMUST00000045693	Smyd5	0.637	1.134	0.615	0.601	0.519	1	0
CUST_2372_PI428259748	ENSMUST00000049572	Lipa	0.474	1.052	0.534	0.677	0.518	0	1
CUST_7265_PI428259748	ENSMUST00000072734	Ifna13	0.882	1.557	1.039	1.066	0.518	1	1
CUST_1784_PI428259748	ENSMUST00000031986	Rab19	1.412	1.702	1.184	1.515	0.518	1	1
CUST_1313_PI428259748	ENSMUST00000027279	Nabp1	1.538	1.506	0.991	1.216	0.515	0	1
CUST_1320_PI428259748	ENSMUST00000029017	Pck1	0.650	1.347	0.835	1.171	0.512	0	1
CUST_1597_PI428259748	ENSMUST00000025979	Aldh18a1	0.969	1.117	0.615	0.852	0.502	0	1
CUST_3694_PI428259748	ENSMUST00000038104	Bdp1	0.673	0.539	0.976	1.039	-0.437	0	0
CUST_6331_PI428259748	ENSMUST00000069922	Mia3	1.243	0.923	1.366	1.565	-0.443	0	0
CUST_1440_PI428259748	ENSMUST00000026504	Xrcc6bp1	1.468	0.722	1.186	0.805	-0.464	1	1
CUST_6528_PI428259748	ENSMUST00000072097	Hsh2d	0.681	0.735	1.200	1.030	-0.466	0	1
CUST_1968_PI428259748	ENSMUST00000025271	Pou5f1	1.016	0.841	1.307	0.976	-0.466	0	1
CUST_1333_PI428259748	ENSMUST00000023779	Nr4a1	1.129	0.793	1.262	0.796	-0.469	0	1
CUST_4364_PI428259748	ENSMUST00000171999	Gcat	1.113	0.975	1.446	1.184	-0.471	0	1
CUST_5878_PI428259748	ENSMUST00000139560	Tmem48	1.235	0.937	1.412	1.050	-0.474	0	1
CUST_4510_PI428259748	ENSMUST00000159551	Wtap	0.821	0.812	1.294	0.974	-0.481	1	1
CUST_1192_PI428259748	ENSMUST00000029625	Sfrp2	1.338	1.311	1.798	1.757	-0.487	1	1
CUST_7076_PI428259748	ENSMUST00000099494	Btaf1	0.679	0.924	1.412	1.344	-0.488	0	1
CUST_5552_PI428259748	ENSMUST00000142242	Gm12588	0.702	0.897	1.421	1.193	-0.524	0	1
CUST_6293_PI428259748	ENSMUST00000111265	Tspan18	0.452	1.133	1.657	1.859	-0.524	0	0

probe	ensembl_id	gene_id	P05MCNTE	P05NTE	P05STE	P05SNTE	Diff	Dmrt1 target	SN intermediate
CUST_4044_PI428259748	ENSMUST00000064204	Actn2	2.784	0.253	0.778	0.563	-0.526	0	1
CUST_1438_PI428259748	ENSMUST00000028515	Chrna1	2.195	0.656	1.184	0.811	-0.529	1	1
CUST_6703_PI428259748	ENSMUST00000076989	Sohlh1	0.951	1.438	1.973	1.470	-0.534	1	1
CUST_2752_PI428259748	ENSMUST00000040583	Heatr5a	0.900	1.325	1.869	1.502	-0.544	1	1
CUST_7183_PI428259748	ENSMUST00000111751	Myl2	NA	0.284	0.836	1.224	-0.552	0	0
CUST_1529_PI428259748	ENSMUST00000028430	Cyct	1.166	1.503	2.060	1.837	-0.557	1	1
CUST_6320_PI428259748	ENSMUST00000132799	Srpr	0.929	0.916	1.497	1.111	-0.581	0	1
CUST_1872_PI428259748	ENSMUST00000031876	Stra8	0.484	0.900	1.522	1.296	-0.622	1	1
CUST_3446_PI428259748	ENSMUST00000039447	Fndc3c1	1.865	0.832	1.459	1.116	-0.627	1	1
CUST_1926_PI428259748	ENSMUST00000027933	Dtl	1.192	1.199	1.828	1.514	-0.628	0	1
CUST_5108_PI428259748	ENSMUST00000111089	Nefm	0.802	0.803	1.442	1.220	-0.639	0	1
CUST_4748_PI428259748	ENSMUST00000142767	AA467197	3.152	0.844	1.521	1.093	-0.677	1	1
CUST_1798_PI428259748	ENSMUST00000028718	NA	1.175	1.204	1.895	1.306	-0.692	0	1
CUST_2109_PI428259748	ENSMUST00000024748	Gtpbp2	1.977	1.688	2.396	2.780	-0.708	0	0
CUST_6079_PI428259748	ENSMUST00000116362	Gm11678	0.146	0.612	1.342	2.112	-0.731	0	0
CUST_4148_PI428259748	ENSMUST00000049208	Hfe2	2.503	0.447	1.222	0.822	-0.775	0	1
CUST_6990_PI428259748	ENSMUST00000111824	Cwc22	0.135	0.655	1.464	1.769	-0.809	1	0
CUST_5260_PI428259748	ENSMUST00000105726	NA	0.298	1.003	1.846	0.610	-0.843	0	0
CUST_6945_PI428259748	ENSMUST00000132795	Tnni1	NA	0.968	1.912	1.275	-0.945	0	1
CUST_4301_PI428259748	ENSMUST00000037685	Spt1	1.045	1.018	2.009	1.201	-0.991	0	1
CUST_2766_PI428259748	ENSMUST00000032910	Mylpf	NA	0.444	1.583	0.963	-1.139	0	1
CUST_441_PI428259748	ENSMUST00000007301	Myh3	2.507	0.709	2.239	1.669	-1.529	0	1
CUST_7388_PI428259748	ENSMUST00000171303	NA	0.856	0.705	2.271	0.936	-1.566	0	1

Table S3.4.5: List of 1% most differently transcribed genes in brain at P07 between mice carrying the N or S allele of *Dmrt1*.

probe	ensembl_id	gene_id	P07MCNBR	P07NBR	P07SBR	P07SNBR	Diff	Dmrt1 target	SN intermediate
CUST_2924_PI428259748	ENSMUST00000041007	Gjd4	1.276	4.015	1.007	1.318	3.008	0	1
CUST_6810_PI428259748	ENSMUST00000073391	Cyp26c1	2.184	3.370	0.944	1.421	2.426	0	1
CUST_6116_PI428259748	ENSMUST00000110052	Ocel1	1.694	3.368	1.207	1.867	2.161	0	1
CUST_7108_PI428259748	ENSMUST00000086795	Barhl2	1.157	2.855	0.794	0.934	2.061	0	1
CUST_686_PI428259748	ENSMUST00000022019	Il9	0.540	2.056	0.276	0.400	1.780	1	1
CUST_4740_PI428259748	ENSMUST00000086614	Zranb3	0.781	2.366	0.647	0.741	1.719	0	1
CUST_4301_PI428259748	ENSMUST00000037685	Spt1	1.496	2.974	1.258	1.546	1.715	0	1
CUST_3607_PI428259748	ENSMUST00000056178	NA	1.179	2.418	0.810	0.911	1.608	0	1
CUST_3905_PI428259748	ENSMUST00000047379	Ptger4	0.749	2.332	0.890	0.886	1.442	0	0
CUST_2160_PI428259748	ENSMUST00000023660	Ripply3	0.899	2.137	0.741	0.671	1.396	0	0
CUST_5532_PI428259748	ENSMUST00000115222	Zbtb44	0.599	1.874	0.594	0.679	1.280	0	1
CUST_415_PI428259748	ENSMUST00000010434	AI597479	1.208	1.978	0.741	0.881	1.237	1	1
CUST_5028_PI428259748	ENSMUST00000109221	B4galt5	1.287	2.352	1.197	1.121	1.155	0	0
CUST_480_PI428259748	ENSMUST00000009256	Bcl2l13	0.962	1.989	0.894	0.902	1.096	0	1
CUST_5956_PI428259748	ENSMUST00000172019	NA	2.592	2.122	1.062	1.076	1.061	0	1
CUST_6146_PI428259748	ENSMUST00000172267	NA	0.112	1.093	0.110	0.092	0.983	0	0
CUST_7238_PI428259748	ENSMUST00000105849	Luzp1	0.223	1.208	0.301	0.396	0.907	0	1
CUST_6001_PI428259748	ENSMUST00000108047	Tbx4	0.977	1.545	0.653	0.876	0.892	1	1
CUST_6592_PI428259748	ENSMUST00000106410	St3gal3	1.121	1.953	1.061	1.121	0.892	0	1
CUST_7461_PI428259748	ENSMUST00000100866	E130309D14Rik	1.134	1.783	0.918	1.109	0.865	0	1
CUST_2627_PI428259748	ENSMUST00000041105	Sfmbt2	1.215	1.462	0.687	0.929	0.776	0	1
CUST_4182_PI428259748	ENSMUST00000042988	Atg14	0.623	1.406	0.639	0.630	0.766	0	0
CUST_4858_PI428259748	ENSMUST00000113165	Ralgps1	1.346	1.916	1.183	1.417	0.732	1	1
CUST_1779_PI428259748	ENSMUST00000030252	NA	0.811	1.495	0.768	0.928	0.727	0	1

probe	ensembl_id	gene_id	P07MCNBR	P07NBR	P07SBR	P07SNBR	Diff	Dmrt1 target	SN intermediate
CUST_6756_PI428259748	ENSMUST00000091178	Sry	0.740	1.486	0.786	0.647	0.700	0	0
CUST_7269_PI428259748	ENSMUST00000119080	Gjb1	1.090	1.514	0.843	1.215	0.671	0	1
CUST_3718_PI428259748	ENSMUST00000034235	NA	1.325	1.786	1.122	1.310	0.664	0	1
CUST_4444_PI428259748	ENSMUST00000077680	Tfe3	0.642	1.390	0.730	0.702	0.659	0	0
CUST_7072_PI428259748	ENSMUST00000103164	Acsf2	1.015	1.385	0.773	0.829	0.612	0	1
CUST_5353_PI428259748	ENSMUST00000072740	Abr	1.300	1.969	1.371	1.289	0.598	1	0
CUST_6769_PI428259748	ENSMUST00000169502	NA	1.989	1.313	0.726	1.055	0.587	0	1
CUST_6479_PI428259748	ENSMUST00000089289	Arl13b	0.760	1.324	0.751	0.702	0.573	0	0
CUST_4431_PI428259748	ENSMUST00000093490	Clcn3	1.120	2.248	1.686	1.672	0.562	0	0
CUST_190_PI428259748	ENSMUST00000003964	Gys1	0.697	1.238	0.693	0.555	0.545	1	0
CUST_2611_PI428259748	ENSMUST00000054098	Slc35g1	0.386	0.876	0.331	0.301	0.545	0	0
CUST_1166_PI428259748	ENSMUST00000027059	Tfap2b	0.720	1.129	0.608	0.146	0.521	0	0
CUST_2917_PI428259748	ENSMUST00000060603	Vmn1r235	NA	0.660	1.080	0.828	-0.420	0	1
CUST_1380_PI428259748	ENSMUST00000025290	Impact	1.297	1.029	1.453	1.242	-0.423	0	1
CUST_7236_PI428259748	ENSMUST00000124310	Bfsp2	1.940	1.541	1.967	1.829	-0.426	1	1
CUST_1142_PI428259748	ENSMUST00000028123	Gad2	1.523	1.415	1.848	1.677	-0.433	0	1
CUST_196_PI428259748	ENSMUST00000002925	Timmdc1	0.508	0.793	1.227	0.748	-0.434	0	0
CUST_4966_PI428259748	ENSMUST00000161002	Ecel1	1.322	1.285	1.723	1.850	-0.439	0	0
CUST_4233_PI428259748	ENSMUST00000057228	Pcdhb9	1.273	0.810	1.255	1.140	-0.444	1	1
CUST_2900_PI428259748	ENSMUST00000033468	Arhgef6	1.008	0.924	1.372	1.107	-0.448	0	1
CUST_1856_PI428259748	ENSMUST00000026196	Got1	1.577	1.132	1.583	1.403	-0.451	1	1
CUST_3050_PI428259748	ENSMUST00000048229	Myrfl	0.906	0.978	1.432	1.073	-0.455	0	1
CUST_7076_PI428259748	ENSMUST00000099494	Btaf1	0.308	0.318	0.776	0.574	-0.458	0	1
CUST_2216_PI428259748	ENSMUST00000050360	P2ry12	1.784	1.473	1.936	1.732	-0.462	0	1
CUST_6990_PI428259748	ENSMUST00000111824	Cwc22	NA	0.411	0.876	0.450	-0.464	1	1

probe	ensembl_id	gene_id	P07MCNBR	P07NBR	P07SBR	P07SNBR	Diff	Dmrt1 target	SN intermediate
CUST_1059_PI428259748	ENSMUST00000020081	Zwint	1.607	1.629	2.097	1.793	-0.468	0	1
CUST_2630_PI428259748	ENSMUST00000062387	Kcnj9	1.696	2.051	2.527	2.574	-0.476	0	0
CUST_426_PI428259748	ENSMUST00000015987	Rxrg	1.249	1.366	1.844	1.618	-0.478	0	1
CUST_7367_PI428259748	ENSMUST00000105423	NA	1.548	1.700	2.180	1.567	-0.480	0	0
CUST_5752_PI428259748	ENSMUST00000082373	Luzp2	2.488	1.465	1.953	1.277	-0.489	0	0
CUST_499_PI428259748	ENSMUST00000006559	Tpbg	1.127	1.305	1.807	1.446	-0.501	0	1
CUST_6667_PI428259748	ENSMUST00000115300	Kcnq5	2.146	2.319	2.827	2.381	-0.508	0	1
CUST_2682_PI428259748	ENSMUST00000038475	Fa2h	1.481	1.059	1.588	1.428	-0.530	0	1
CUST_1311_PI428259748	ENSMUST00000026541	Caly	1.611	1.703	2.275	2.692	-0.573	0	0
CUST_5656_PI428259748	ENSMUST00000174379	Dctd	1.134	0.974	1.570	1.209	-0.596	0	1
CUST_1770_PI428259748	ENSMUST00000024832	Rsph1	0.820	0.248	0.849	0.807	-0.600	0	1
CUST_3189_PI428259748	ENSMUST00000035061	Ngp	0.869	0.552	1.188	1.029	-0.635	0	1
CUST_675_PI428259748	ENSMUST00000022150	Cartpt	1.525	1.371	2.010	1.654	-0.639	1	1
CUST_6134_PI428259748	ENSMUST00000069268	Tas2r102	1.613	1.229	1.886	1.489	-0.657	1	1
CUST_3088_PI428259748	ENSMUST00000065408	Ly6c1	1.622	1.061	1.749	1.262	-0.688	0	1
CUST_268_PI428259748	ENSMUST00000014640	Ankrd28	1.394	1.501	2.255	1.669	-0.754	1	1
CUST_5245_PI428259748	ENSMUST00000115310	Trp63	1.494	0.889	1.686	2.158	-0.797	0	0
CUST_5552_PI428259748	ENSMUST00000142242	Gm12588	0.682	0.466	1.281	1.032	-0.815	0	1
CUST_1383_PI428259748	ENSMUST00000030489	Tal1	1.180	1.337	2.182	1.578	-0.845	0	1
CUST_3901_PI428259748	ENSMUST00000062545	Olfr958	1.773	1.337	2.323	1.564	-0.986	1	1
CUST_4897_PI428259748	ENSMUST00000078694	Ppp1r1b	1.447	1.713	2.752	2.952	-1.040	0	0
CUST_5513_PI428259748	ENSMUST00000111064	Ntsr2	1.715	1.523	2.719	2.888	-1.196	0	0
CUST_6011_PI428259748	ENSMUST00000074897	Olfr535	1.487	0.944	2.220	1.375	-1.276	0	1

Table S3.4.6: List of 1% most differently transcribed genes in testis at P07 between mice carrying the N or S allele of *Dmrt1*.

probe	ensembl_id	gene_id	P07MCNTE	P07NTE	P07STE	P07SNTE	Diff	Dmrt1 target	SN intermediate
CUST_4085_PI428259748	ENSMUST00000056449	Arhgap30	1.126	3.028	0.818	0.876	2.210	0	1
CUST_6146_PI428259748	ENSMUST00000172267	NA	0.174	1.969	0.165	0.207	1.804	0	1
CUST_1286_PI428259748	ENSMUST00000023758	Asic1	0.129	1.366	0.153	0.138	1.214	0	0
CUST_2528_PI428259748	ENSMUST00000057935	Slc25a44	0.470	1.438	0.513	0.621	0.925	1	1
CUST_1570_PI428259748	ENSMUST00000029076	Car3	NA	2.930	2.014	2.165	0.917	0	1
CUST_899_PI428259748	ENSMUST00000020156	NA	0.999	1.936	1.025	1.194	0.912	0	1
CUST_2670_PI428259748	ENSMUST00000058755	Fzd4	1.709	2.027	1.123	1.254	0.904	0	1
CUST_2208_PI428259748	ENSMUST00000039994	Cep250	0.935	1.722	0.848	0.964	0.874	0	1
CUST_2784_PI428259748	ENSMUST00000047419	Tspo	1.501	2.238	1.387	1.493	0.851	0	1
CUST_3422_PI428259748	ENSMUST00000044833	Oas3	2.029	1.818	0.989	0.832	0.828	0	0
CUST_6315_PI428259748	ENSMUST00000078849	Gpc6	1.037	1.616	0.817	1.114	0.799	0	1
CUST_260_PI428259748	ENSMUST00000016771	Myh9	1.190	2.081	1.302	1.234	0.779	0	0
CUST_6866_PI428259748	ENSMUST00000160262	Armc2	0.785	1.456	0.693	0.670	0.763	1	0
CUST_959_PI428259748	ENSMUST00000022242	Emb	1.077	1.561	0.802	0.825	0.759	1	1
CUST_6252_PI428259748	ENSMUST00000084763	Nlrp14	1.237	1.689	0.966	2.664	0.723	0	0
CUST_5580_PI428259748	ENSMUST00000119026	Tmem132c	0.831	1.367	0.684	0.879	0.682	0	1
CUST_259_PI428259748	ENSMUST00000016143	Wasf3	0.476	1.169	0.493	0.437	0.676	0	0
CUST_3510_PI428259748	ENSMUST00000033608	Sytl4	1.545	2.360	1.685	1.577	0.675	1	0
CUST_2372_PI428259748	ENSMUST00000049572	Lipa	0.467	1.109	0.438	0.640	0.671	0	1
CUST_6942_PI428259748	ENSMUST00000098090	Coq7	1.188	1.855	1.193	1.254	0.662	0	1
CUST_2301_PI428259748	ENSMUST00000041331	Scd1	1.402	1.454	0.842	1.047	0.611	0	1
CUST_5949_PI428259748	ENSMUST00000099137	NA	0.800	1.651	1.041	0.962	0.610	0	0
CUST_2752_PI428259748	ENSMUST00000040583	Heatr5a	0.875	1.749	1.139	0.878	0.610	1	0
CUST_1499_PI428259748	ENSMUST00000023610	Adamts1	1.220	1.117	0.517	0.712	0.600	0	1
CUST_3203_PI428259748	ENSMUST00000045847	Erf	1.106	1.574	0.981	0.949	0.593	1	0
CUST_4894_PI428259748	ENSMUST00000077918	Cfi	0.611	1.818	1.233	1.269	0.585	0	1
CUST_3747_PI428259748	ENSMUST00000043042	Tmco4	1.233	1.933	1.351	1.252	0.582	1	0
CUST_7028_PI428259748	ENSMUST00000102568	Phactr4	1.368	1.649	1.084	1.340	0.565	0	1
CUST_1569_PI428259748	ENSMUST00000027797	Nvl	0.792	1.278	0.717	1.064	0.561	0	1
CUST_258_PI428259748	ENSMUST00000005334	Shbg	1.631	2.255	1.699	1.296	0.556	1	0

probe	ensembl_id	gene_id	P07MCNTE	P07NTE	P07STE	P07SNTE	Diff	Dmrt1 target	SN intermediate
CUST_3985_PI428259748	ENSMUST00000058856	Scd4	0.727	1.504	0.955	1.338	0.550	0	1
CUST_2052_PI428259748	ENSMUST00000023920	Tmem52	1.164	1.609	1.062	1.138	0.548	0	1
CUST_3503_PI428259748	ENSMUST00000034046	Acsl1	2.086	2.122	1.581	1.949	0.541	0	1
CUST_1039_PI428259748	ENSMUST00000018407	Tbx5	0.797	1.420	0.879	0.993	0.541	0	1
CUST_5004_PI428259748	ENSMUST00000112104	Mid1	0.644	0.965	0.426	0.833	0.539	0	1
CUST_3607_PI428259748	ENSMUST00000056178	NA	1.061	1.337	0.806	0.892	0.531	0	1
CUST_271_PI428259748	ENSMUST00000009340	Mnda	1.680	0.201	0.711	0.842	-0.510	0	0
CUST_1371_PI428259748	ENSMUST00000026043	Slk	1.550	1.046	1.558	1.466	-0.512	0	1
CUST_7119_PI428259748	ENSMUST00000084828	Ncapg2	1.244	0.882	1.395	0.776	-0.513	0	0
CUST_4121_PI428259748	ENSMUST00000060798	Unc119b	1.695	1.311	1.826	2.011	-0.515	0	0
CUST_1442_PI428259748	ENSMUST00000030460	Gpbp1l1	1.108	0.956	1.474	1.396	-0.518	1	1
CUST_3553_PI428259748	ENSMUST00000033917	Spata4	0.407	0.462	0.986	0.625	-0.524	0	1
CUST_4539_PI428259748	ENSMUST00000071267	Pcgf5	1.497	1.017	1.542	1.619	-0.525	0	0
CUST_5045_PI428259748	ENSMUST00000169003	Rbpms2	2.287	1.114	1.652	1.181	-0.537	0	1
CUST_1031_PI428259748	ENSMUST00000022522	Tdh	1.494	0.885	1.428	1.245	-0.543	0	1
CUST_2483_PI428259748	ENSMUST00000050129	Cox7b2	1.560	1.318	1.867	1.868	-0.549	1	0
CUST_6100_PI428259748	ENSMUST00000081321	Poteg	1.090	1.429	1.981	1.332	-0.552	1	0
CUST_6761_PI428259748	ENSMUST00000090025	Aard	1.947	2.468	3.020	2.711	-0.552	1	1
CUST_1020_PI428259748	ENSMUST00000022613	Esco2	1.394	1.111	1.664	1.546	-0.553	0	1
CUST_5972_PI428259748	ENSMUST00000097419	Fgfr1op	1.359	0.925	1.486	1.333	-0.561	0	1
CUST_7364_PI428259748	ENSMUST00000106059	S100pbp	1.089	0.881	1.446	1.341	-0.565	0	1
CUST_1770_PI428259748	ENSMUST00000024832	Rsph1	0.494	0.313	0.890	0.699	-0.577	0	1
CUST_2501_PI428259748	ENSMUST00000059836	Kif7	2.166	0.909	1.494	1.010	-0.586	0	1
CUST_5739_PI428259748	ENSMUST00000089056	Tmem255a	0.951	1.103	1.691	1.040	-0.588	1	0
CUST_2056_PI428259748	ENSMUST00000025963	Noc3l	1.490	0.952	1.551	1.400	-0.599	0	1
CUST_388_PI428259748	ENSMUST00000016383	Lonrf3	1.763	1.216	1.825	1.816	-0.608	1	1
CUST_352_PI428259748	ENSMUST00000005607	Asf1b	1.397	1.128	1.739	1.534	-0.611	0	1
CUST_3937_PI428259748	ENSMUST00000055770	Hist1h1a	1.495	1.070	1.715	1.470	-0.644	1	1
CUST_2845_PI428259748	ENSMUST00000065216	Spink2	1.464	0.994	1.658	0.514	-0.664	1	0
CUST_4206_PI428259748	ENSMUST00000033414	Slc6a14	1.317	1.397	2.075	1.257	-0.678	1	0
CUST_5552_PI428259748	ENSMUST00000142242	Gm12588	0.833	0.791	1.514	1.350	-0.723	0	1

probe	ensembl_id	gene_id	P07MCNTE	P07NTE	P07STE	P07SNTE	Diff	Dmrt1 target	SN intermediate
CUST_574_PI428259748	ENSMUST00000021816	Susd3	1.067	1.382	2.129	1.616	-0.747	0	1
CUST_2217_PI428259748	ENSMUST00000045866	Ddx21	1.054	1.029	1.779	1.407	-0.751	0	1
CUST_6703_PI428259748	ENSMUST00000076989	Sohlh1	2.080	1.197	2.031	2.281	-0.835	1	0
CUST_6823_PI428259748	ENSMUST00000156895	Ifi203	2.430	0.397	1.259	1.420	-0.862	0	0
CUST_7076_PI428259748	ENSMUST00000099494	Btaf1	0.882	0.781	1.644	1.442	-0.863	0	1
CUST_5064_PI428259748	ENSMUST00000113457	Col4a3	1.144	1.250	2.151	2.137	-0.900	1	1
CUST_4872_PI428259748	ENSMUST00000114875	Mbnl3	1.718	0.987	1.949	1.773	-0.962	0	1
CUST_6373_PI428259748	ENSMUST00000101698	AU022751	1.398	0.998	2.106	1.796	-1.107	0	1
CUST_3178_PI428259748	ENSMUST00000063340	Rhox2a	2.003	0.764	2.157	1.843	-1.393	0	1
CUST_768_PI428259748	ENSMUST00000022708	Trim52	0.947	0.585	2.039	1.063	-1.454	1	1
CUST_1872_PI428259748	ENSMUST00000031876	Stra8	1.396	0.793	2.604	2.206	-1.811	1	1

Table S3.4.7: List of 1% most differently transcribed genes in brain at P09 between mice carrying the N or S allele of *Dmrt1*.

probe	ensembl_id	gene_id	P09MCNBR	P09NBR	P09SBR	P09SNBR	Diff	Dmrt1 target	SN intermediate
CUST_2619_PI428259748	ENSMUST00000060576	Lpar4	1.019	2.660	0.408	0.366	2.252	0	0
CUST_5486_PI428259748	ENSMUST00000112028	Erlin1	0.827	2.538	0.638	0.655	1.900	0	1
CUST_2682_PI428259748	ENSMUST00000038475	Fa2h	2.237	3.447	1.752	1.518	1.696	0	0
CUST_3002_PI428259748	ENSMUST00000051033	NA	0.983	2.509	0.999	1.075	1.510	0	1
CUST_6211_PI428259748	ENSMUST00000105590	Esr1	0.456	1.956	0.488	0.477	1.469	0	0
CUST_6768_PI428259748	ENSMUST00000072271	Agxt2l1	1.079	2.343	1.004	1.050	1.339	0	1
CUST_4013_PI428259748	ENSMUST00000055884	Htr5b	1.659	3.034	1.696	1.655	1.338	1	0
CUST_2684_PI428259748	ENSMUST00000053066	Qk	1.183	2.172	0.863	1.089	1.310	1	1
CUST_1559_PI428259748	ENSMUST00000024032	9530002B09Rik	0.164	1.227	0.166	0.177	1.062	1	1
CUST_2627_PI428259748	ENSMUST00000041105	Sfmbt2	1.278	1.826	0.798	0.698	1.028	0	0
CUST_1259_PI428259748	ENSMUST00000025224	Gfra3	1.295	1.978	0.962	0.682	1.016	1	0
CUST_1346_PI428259748	ENSMUST00000027988	Ccdc3	0.980	1.850	0.867	0.895	0.984	0	1
CUST_6784_PI428259748	ENSMUST00000132743	Tmcc3	0.676	1.759	0.794	0.599	0.965	0	0
CUST_145_PI428259748	ENSMUST00000002640	Scin	0.226	1.190	0.230	0.180	0.961	1	0
CUST_1404_PI428259748	ENSMUST00000031224	Tgfbr3	2.079	2.911	1.952	2.087	0.958	0	1
CUST_1594_PI428259748	ENSMUST00000026670	Nptx1	1.132	1.961	1.059	1.284	0.902	0	1
CUST_1496_PI428259748	ENSMUST00000029714	Pear1	1.100	1.788	0.912	1.131	0.876	1	1
CUST_7165_PI428259748	ENSMUST00000117172	Vwa3b	0.897	1.821	0.963	0.883	0.859	1	0
CUST_4761_PI428259748	ENSMUST00000108551	Gp1ba	1.150	1.738	0.889	0.995	0.849	1	1
CUST_7047_PI428259748	ENSMUST00000133298	Dlgap2	1.239	2.570	1.727	2.441	0.843	0	1
CUST_4927_PI428259748	ENSMUST00000093207	Hba-a2	0.793	1.745	0.926	1.042	0.819	0	1
CUST_1210_PI428259748	ENSMUST00000029995	4930547C10Rik	0.865	1.729	0.911	0.954	0.818	1	1
CUST_1490_PI428259748	ENSMUST00000023934	Hbb-b1	0.878	1.690	0.890	1.045	0.800	0	1
CUST_4606_PI428259748	ENSMUST00000070659	1700001K19Rik	1.362	1.548	0.778	1.315	0.769	0	1
CUST_3394_PI428259748	ENSMUST00000051153	3300002I08Rik	2.173	2.842	2.079	1.875	0.763	1	0
CUST_1459_PI428259748	ENSMUST00000023754	Aqp6	1.050	1.871	1.150	1.736	0.721	1	1
CUST_4115_PI428259748	ENSMUST00000041203	Cpne9	3.176	3.020	2.303	3.293	0.717	0	0
CUST_4336_PI428259748	ENSMUST00000121979	Cep135	0.302	1.141	0.434	0.293	0.707	0	0
CUST_4889_PI428259748	ENSMUST00000135107	Sox3	0.890	1.416	0.721	1.060	0.695	0	1
CUST_4818_PI428259748	ENSMUST00000066337	Alas2	0.802	1.537	0.867	0.842	0.670	0	0

probe	ensembl_id	gene_id	P09MCNBR	P09NBR	P09SBR	P09SNBR	Diff	Dmrt1 target	SN intermediate
CUST_1504_PI428259748	ENSMUST00000032194	Bhlhe40	1.160	1.966	1.307	1.485	0.659	0	1
CUST_3754_PI428259748	ENSMUST00000059794	Nhlh1	1.219	1.844	1.185	1.060	0.659	1	0
CUST_7019_PI428259748	ENSMUST00000109763	NA	0.971	1.698	1.055	1.046	0.644	0	0
CUST_3085_PI428259748	ENSMUST00000053175	Vstm4	0.185	0.785	0.157	0.155	0.628	0	0
CUST_5513_PI428259748	ENSMUST00000111064	Ntsr2	2.891	3.560	2.952	3.614	0.608	0	0
CUST_7238_PI428259748	ENSMUST00000105849	Luzp1	0.441	1.070	0.481	0.478	0.589	0	0
CUST_4966_PI428259748	ENSMUST00000161002	Ecel1	1.920	1.709	2.122	1.148	-0.412	0	0
CUST_7158_PI428259748	ENSMUST00000090558	Celsr2	1.939	1.272	1.689	1.381	-0.417	0	1
CUST_4055_PI428259748	ENSMUST00000042096	Emc1	1.911	1.433	1.852	1.578	-0.419	0	1
CUST_5700_PI428259748	ENSMUST00000090679	Tac1	2.003	2.091	2.510	1.922	-0.419	0	0
CUST_4385_PI428259748	ENSMUST00000172269	Sptbn4	2.263	1.481	1.906	1.926	-0.424	1	0
CUST_3350_PI428259748	ENSMUST00000060918	Amz1	2.077	1.972	2.400	1.719	-0.428	0	0
CUST_6352_PI428259748	ENSMUST00000089654	BC024139	2.350	1.105	1.539	1.096	-0.434	0	0
CUST_2042_PI428259748	ENSMUST00000029188	Wisp2	0.798	0.847	1.286	0.850	-0.439	1	1
CUST_5480_PI428259748	ENSMUST00000084677	Gm21093	1.382	0.815	1.274	0.640	-0.459	0	0
CUST_2717_PI428259748	ENSMUST00000034408	Gpr83	2.255	2.406	2.879	2.053	-0.473	1	0
CUST_824_PI428259748	ENSMUST00000022913	Dcstamp	0.792	0.834	1.314	0.766	-0.480	1	0
CUST_898_PI428259748	ENSMUST00000020027	Serinc1	1.106	1.875	2.365	2.201	-0.490	0	1
CUST_7367_PI428259748	ENSMUST00000105423	NA	1.496	1.235	1.726	1.250	-0.491	0	1
CUST_5771_PI428259748	ENSMUST00000070375	Penk	0.570	0.643	1.136	0.638	-0.493	0	0
CUST_5705_PI428259748	ENSMUST00000068681	Ngef	2.270	2.206	2.707	2.856	-0.501	0	0
CUST_1748_PI428259748	ENSMUST00000029852	NA	1.239	0.880	1.387	0.766	-0.507	0	0
CUST_4262_PI428259748	ENSMUST00000037779	NA	1.195	1.100	1.624	1.535	-0.524	0	1
CUST_2569_PI428259748	ENSMUST00000033545	Rab39b	1.407	1.901	2.445	1.708	-0.544	0	0
CUST_2791_PI428259748	ENSMUST00000042767	Slitrk5	1.514	1.228	1.777	1.518	-0.549	0	1
CUST_2834_PI428259748	ENSMUST00000045692	Fbxl16	2.393	1.957	2.517	2.219	-0.561	0	1
CUST_4616_PI428259748	ENSMUST00000162875	2010300C02Rik	2.033	2.094	2.662	2.529	-0.568	0	1
CUST_426_PI428259748	ENSMUST00000015987	Rxrg	1.434	1.248	1.836	1.555	-0.588	0	1
CUST_499_PI428259748	ENSMUST00000006559	Tpbg	1.269	1.340	1.931	2.150	-0.591	0	0
CUST_1624_PI428259748	ENSMUST00000027231	Slc9a2	1.223	1.253	1.866	1.120	-0.613	1	0
CUST_3050_PI428259748	ENSMUST00000048229	Myrfl	3.136	2.720	3.345	1.734	-0.625	0	0
CUST_1092_PI428259748	ENSMUST00000021932	Drd1a	1.711	1.338	1.976	1.905	-0.637	1	1
CUST_5409_PI428259748	ENSMUST00000102608	AU040320	1.036	0.844	1.482	0.883	-0.638	0	1
CUST_3233_PI428259748	ENSMUST00000048275	NA	0.601	0.572	1.238	1.025	-0.666	0	1

probe	ensembl_id	gene_id	P09MCNBR	P09NBR	P09SBR	P09SNBR	Diff	Dmrt1 target	SN intermediate
CUST_6551_PI428259748	ENSMUST00000090546	Tmem167b	0.664	0.595	1.266	0.349	-0.671	0	0
CUST_4592_PI428259748	ENSMUST00000089421	NA	1.600	0.847	1.551	0.885	-0.704	0	1
CUST_2297_PI428259748	ENSMUST00000057495	Tmem161b	0.853	0.516	1.282	0.863	-0.767	1	1
CUST_7001_PI428259748	ENSMUST00000155301	Fbxl12	1.198	0.473	1.293	0.775	-0.820	0	1
CUST_610_PI428259748	ENSMUST00000020920	Rgs9	2.158	2.169	3.020	2.771	-0.851	0	1
CUST_1602_PI428259748	ENSMUST00000026360	Itgb8	0.814	0.809	1.662	1.526	-0.853	0	1
CUST_4897_PI428259748	ENSMUST00000078694	Ppp1r1b	2.377	2.282	4.059	3.534	-1.777	0	1
CUST_3189_PI428259748	ENSMUST00000035061	Ngp	0.696	0.605	2.669	0.852	-2.064	0	1

Table S3.4.8: List of 1% most differently transcribed genes in testis at P09 between mice carrying the N or S allele of *Dmrt1*.

probe	ensembl_id	gene_id	P09MCNTE	P09NTE	P09STE	P09SNTE	Diff	Dmrt1 target	SN intermediate
CUST_1594_PI428259748	ENSMUST00000026670	Nptx1	0.721	2.475	0.526	0.673	1.950	0	1
CUST_6240_PI428259748	ENSMUST00000124516	Myh1	2.776	1.986	0.247	0.362	1.739	1	1
CUST_789_PI428259748	ENSMUST00000021907	Fbp2	1.092	2.545	1.283	1.850	1.262	0	1
CUST_2766_PI428259748	ENSMUST00000032910	Mylpf	1.966	1.470	0.221	0.367	1.249	0	1
CUST_1078_PI428259748	ENSMUST00000021778	Prl2c5	1.187	2.259	1.135	1.010	1.124	1	0
CUST_1192_PI428259748	ENSMUST00000029625	Sfrp2	1.806	2.125	1.005	1.757	1.120	1	1
CUST_1030_PI428259748	ENSMUST00000022185	F2rl1	0.298	2.471	1.408	1.792	1.064	1	1
CUST_7183_PI428259748	ENSMUST00000111751	Myl2	1.050	1.186	0.151	0.146	1.034	0	0
CUST_4351_PI428259748	ENSMUST00000168901	NA	1.731	1.785	0.759	1.005	1.026	0	1
CUST_5956_PI428259748	ENSMUST00000172019	NA	0.894	1.798	0.772	0.923	1.026	0	1
CUST_6536_PI428259748	ENSMUST00000068068	1700041G16Rik	1.379	1.846	0.909	0.798	0.937	1	0
CUST_6810_PI428259748	ENSMUST00000073391	Cyp26c1	1.140	1.675	0.745	0.851	0.930	0	1
CUST_4306_PI428259748	ENSMUST00000053441	Adam29	1.199	1.787	0.870	0.968	0.917	0	1
CUST_2218_PI428259748	ENSMUST00000035077	Ltf	1.370	2.786	1.878	1.802	0.908	0	0
CUST_2146_PI428259748	ENSMUST00000028794	Siglec1	1.668	1.748	0.949	1.065	0.799	1	1
CUST_2625_PI428259748	ENSMUST00000050544	Has2	0.540	1.225	0.454	0.344	0.771	0	0
CUST_2715_PI428259748	ENSMUST00000039480	Zswim4	0.888	2.036	1.272	1.124	0.765	0	0
CUST_4208_PI428259748	ENSMUST00000049768	Epx	0.856	2.025	1.294	1.086	0.732	1	0
CUST_6906_PI428259748	ENSMUST00000172064	NA	1.237	2.303	1.578	1.930	0.726	0	1
CUST_5305_PI428259748	ENSMUST00000094753	Ceacam20	0.967	1.691	1.019	1.117	0.672	1	1
CUST_6479_PI428259748	ENSMUST00000089289	Arl13b	0.797	1.421	0.753	0.851	0.668	0	1
CUST_3888_PI428259748	ENSMUST00000036072	5031414D18Rik	1.068	1.665	1.014	0.882	0.651	1	0
CUST_1438_PI428259748	ENSMUST00000028515	Chrna1	1.137	1.398	0.763	0.851	0.635	1	1
CUST_4536_PI428259748	ENSMUST00000075595	Olfr504	NA	1.328	0.699	0.631	0.630	0	0
CUST_1696_PI428259748	ENSMUST00000031221	Cdc7	0.977	1.485	0.858	0.725	0.627	0	0
CUST_6263_PI428259748	ENSMUST00000171644	Pparg	2.761	2.543	1.924	2.270	0.619	0	1
CUST_5481_PI428259748	ENSMUST00000165774	Gbp2	1.729	1.519	0.922	1.574	0.598	1	0
CUST_2212_PI428259748	ENSMUST00000033522	3830417A13Rik	1.587	2.383	1.787	1.420	0.596	0	0

probe	ensembl_id	gene_id	P09MCNTE	P09NTE	P09STE	P09SNTE	Diff	Dmrt1 target	SN intermediate
CUST_5497_PI428259748	ENSMUST00000074854	Ptpla	1.764	1.917	1.335	1.818	0.583	0	1
CUST_5886_PI428259748	ENSMUST00000068860	Epha6	0.095	0.677	0.097	NA	0.580	0	0
CUST_1983_PI428259748	ENSMUST00000031894	Clcn1	1.066	1.542	0.963	1.111	0.579	1	1
CUST_7294_PI428259748	ENSMUST00000099895	Olfr1054	1.181	1.352	0.776	1.050	0.576	0	1
CUST_2752_PI428259748	ENSMUST00000040583	Heatr5a	0.961	1.529	0.966	1.010	0.562	1	1
CUST_1460_PI428259748	ENSMUST00000030452	Ccdc163	0.860	1.655	1.096	1.111	0.559	0	1
CUST_6243_PI428259748	ENSMUST00000077248	NA	2.483	2.879	2.321	3.098	0.558	0	0
CUST_4606_PI428259748	ENSMUST00000070659	1700001K19Rik	1.270	1.300	0.744	1.061	0.555	0	1
CUST_6609_PI428259748	ENSMUST00000086556	Elk4	0.892	0.852	1.249	1.026	-0.397	0	1
CUST_7285_PI428259748	ENSMUST00000076831	Cdc42ep5	1.765	1.650	2.054	1.844	-0.404	0	1
CUST_7119_PI428259748	ENSMUST00000084828	Ncapg2	1.881	2.347	2.752	2.283	-0.405	0	0
CUST_4514_PI428259748	ENSMUST00000091851	Mapk4	0.604	0.444	0.852	0.625	-0.408	1	1
CUST_2414_PI428259748	ENSMUST00000062211	Gpat2	2.267	2.450	2.859	2.020	-0.409	0	0
CUST_6703_PI428259748	ENSMUST00000076989	Sohlh1	1.837	1.999	2.410	1.932	-0.411	1	0
CUST_3958_PI428259748	ENSMUST00000046485	Efcab11	1.640	1.256	1.670	1.334	-0.414	1	1
CUST_5191_PI428259748	ENSMUST00000106000	Cd151	2.577	0.511	0.925	2.535	-0.414	0	0
CUST_366_PI428259748	ENSMUST00000013458	Adh4	1.821	1.714	2.164	1.671	-0.450	1	0
CUST_4520_PI428259748	ENSMUST00000105502	Foxo3	1.104	0.838	1.295	1.809	-0.456	0	0
CUST_498_PI428259748	ENSMUST00000016902	Bik	2.154	1.862	2.321	1.716	-0.458	1	0
CUST_3913_PI428259748	ENSMUST00000033691	Tsx	1.757	2.642	3.101	2.318	-0.459	1	0
CUST_6452_PI428259748	ENSMUST00000071978	not_available	1.320	0.745	1.215	1.644	-0.470	0	0
CUST_2297_PI428259748	ENSMUST00000057495	Tmem161b	0.655	0.339	0.813	0.614	-0.475	1	1
CUST_921_PI428259748	ENSMUST00000021495	Serpina5	1.364	1.735	2.221	1.759	-0.486	0	1
CUST_5166_PI428259748	ENSMUST00000171587	Gbp11	0.825	0.795	1.293	0.839	-0.499	1	1
CUST_337_PI428259748	ENSMUST00000006667	Gvin1	0.844	1.072	1.579	1.181	-0.507	0	1
CUST_2501_PI428259748	ENSMUST00000059836	Kif7	1.990	0.787	1.295	1.080	-0.507	0	1
CUST_1602_PI428259748	ENSMUST00000026360	Itgb8	0.495	0.425	0.939	0.777	-0.514	0	1
CUST_4917_PI428259748	ENSMUST00000092298	Zfp750	0.782	1.185	1.741	1.016	-0.556	1	0
CUST_3749_PI428259748	ENSMUST00000035194	Mapkapk3	0.894	0.885	1.444	1.181	-0.559	1	1
CUST_3233_PI428259748	ENSMUST00000048275	not_available	0.799	0.644	1.211	1.035	-0.567	0	1
CUST_3186_PI428259748	ENSMUST00000048054	Chtf18	1.665	1.510	2.093	1.569	-0.583	0	1

probe	ensembl_id	gene_id	P09MCNTE	P09NTE	P09STE	P09SNTE	Diff	Dmrt1 target	SN intermediate
CUST_4474_PI428259748	ENSMUST00000082226	not_available	2.211	0.471	1.068	1.441	-0.597	0	0
CUST_4356_PI428259748	ENSMUST00000163360	D17H6S56E-5	0.782	0.885	1.533	1.645	-0.648	0	0
CUST_2763_PI428259748	ENSMUST00000032969	Pold3	0.995	1.153	1.803	1.277	-0.650	0	1
CUST_7471_PI428259748	ENSMUST00000081271	Spink12	2.154	1.960	2.618	2.283	-0.658	1	1
CUST_5975_PI428259748	ENSMUST00000163265	Arhgap19	1.331	1.242	1.923	1.345	-0.681	0	1
CUST_3189_PI428259748	ENSMUST00000035061	Ngp	0.577	0.511	1.212	0.564	-0.701	0	1
CUST_1427_PI428259748	ENSMUST00000031243	Spp1	0.252	0.410	1.143	1.415	-0.733	1	0
CUST_3937_PI428259748	ENSMUST00000055770	Hist1h1a	1.744	1.705	2.508	2.156	-0.803	1	1
CUST_429_PI428259748	ENSMUST00000017153	Sdc4	0.497	1.082	1.914	1.219	-0.831	0	1
CUST_2906_PI428259748	ENSMUST00000041052	Hist1h1t	1.348	1.557	2.596	2.280	-1.039	1	1
CUST_1925_PI428259748	ENSMUST00000028935	Cst9	2.999	0.418	1.592	1.378	-1.173	1	1
CUST_6252_PI428259748	ENSMUST00000084763	Nlrp14	2.057	1.008	2.293	2.896	-1.285	0	0
CUST_4300_PI428259748	ENSMUST00000048993	Polr3g	2.437	1.063	3.290	2.178	-2.227	0	1

Table S3.5.1: Overview about shared genes across the analyzed stages in brain.

genes	P01, 05	P01, 07	P01, 09	P05, 07	P05, 09	P07, 09	P01,05,07	P01,07,09	P05,07,09	P01,05,09	P01,05,07,09
Zic2	1	0	0	0	0	0	0	0	0	0	0
Dctd	0	1	0	0	0	0	0	0	0	0	0
Cyp26c1	0	0	0	0	0	0	1	0	0	0	0
AI597479	0	1	0	0	0	0	0	0	0	0	0
Fbxl12	0	0	1	0	0	0	0	0	0	0	0
Otx2	1	0	0	0	0	0	0	0	0	0	0
Lhx8	0	0	1	0	0	0	0	0	0	0	0
Bfsp2	0	1	0	0	0	0	0	0	0	0	0
Eya1	1	0	0	0	0	0	0	0	0	0	0
Ptx3	1	0	0	0	0	0	0	0	0	0	0
Olfr535	0	0	0	0	0	0	1	0	0	0	0
Gm21093	0	0	1	0	0	0	0	0	0	0	0
Cwc22	0	0	0	0	0	0	1	0	0	0	0
Gm11678	1	0	0	0	0	0	0	0	0	0	0
Tfap2b	0	0	0	1	0	0	0	0	0	0	0
Cep135	0	0	0	0	1	0	0	0	0	0	0
Itgb8	0	0	0	0	1	0	0	0	0	0	0
Agxt2l1	0	0	0	0	1	0	0	0	0	0	0
Amz1	0	0	0	0	1	0	0	0	0	0	0
Ngp	0	0	0	0	0	0	0	0	1	0	0
Slc9a2	0	0	0	0	1	0	0	0	0	0	0
Ankrd28	0	0	0	1	0	0	0	0	0	0	0
Timmdc1	0	0	0	1	0	0	0	0	0	0	0
Rsph1	0	0	0	1	0	0	0	0	0	0	0

genes	P01, 05	P01, 07	P01, 09	P05, 07	P05, 09	P07, 09	P01,05,07	P01,07,09	P05,07,09	P01,05,09	P01,05,07,09
Spt1	0	0	0	1	0	0	0	0	0	0	0
Tal1	0	0	0	1	0	0	0	0	0	0	0
Luzp1	0	0	0	0	0	1	0	0	0	0	0
Sfmbt2	0	0	0	0	0	1	0	0	0	0	0
Ecel1	0	0	0	0	0	1	0	0	0	0	0
Myrfl	0	0	0	0	0	1	0	0	0	0	0
Rxrg	0	0	0	0	0	1	0	0	0	0	0
Pcdh15	0	0	0	0	0	1	0	0	0	0	0
Tpbg	0	0	0	0	0	1	0	0	0	0	0
Fa2h	0	0	0	0	0	1	0	0	0	0	0
Ppp1r1b	0	0	0	0	0	1	0	0	0	0	0
Ntsr2	0	0	0	0	0	1	0	0	0	0	0

Table S3.5.2: Overview about shared genes across the analyzed stages in testis.

genes	P01, P05	P01, P07	P01, P09	P05, P07	P05, P09	P07, P09	P01,05,07	P01,07,09	P05,07,09	P01,05,09	P01,05,07,09
Bcat2	1	0	0	0	0	0	0	0	0	0	0
1700041G16Rik	0	0	0	0	0	0	0	0	0	1	0
Lipa	0	0	0	0	0	0	1	0	0	0	0
Myh3	1	0	0	0	0	0	0	0	0	0	0
Ifi203	0	0	0	0	0	0	1	0	0	0	0
H2-T22	1	0	0	0	0	0	0	0	0	0	0
Tnni1	1	0	0	0	0	0	0	0	0	0	0
Pou5f1	1	0	0	0	0	0	0	0	0	0	0
Tmem48	1	0	0	0	0	0	0	0	0	0	0
Gtpbp2	1	0	0	0	0	0	0	0	0	0	0
Btaf1	0	0	0	0	0	0	1	0	0	0	0
Coq7	0	1	0	0	0	0	0	0	0	0	0
Myl2	0	0	0	0	0	0	0	0	0	1	0
Spt1	1	0	0	0	0	0	0	0	0	0	0
Spp1	0	0	1	0	0	0	0	0	0	0	0
Tmco4	0	0	0	1	0	0	0	0	0	0	0
Has2	0	0	0	0	1	0	0	0	0	0	0
Scd4	0	0	0	1	0	0	0	0	0	0	0
Cst9	0	0	0	0	1	0	0	0	0	0	0
Mnda	0	0	0	1	0	0	0	0	0	0	0
Adamts1	0	0	0	1	0	0	0	0	0	0	0
Itgb8	0	0	0	0	1	0	0	0	0	0	0
Sdc4	0	0	0	0	1	0	0	0	0	0	0
Sfrp2	0	0	0	0	1	0	0	0	0	0	0
Gm12588	0	0	0	1	0	0	0	0	0	0	0
Chrna1	0	0	0	0	1	0	0	0	0	0	0
Sohlh1	0	0	0	0	0	0	0	0	1	0	0
Heatr5a	0	0	0	0	0	0	0	0	1	0	0
Stra8	0	0	0	1	0	0	0	0	0	0	0
Mylpf	0	0	0	0	1	0	0	0	0	0	0
Nlrp14	0	0	0	0	0	1	0	0	0	0	0

genes	P01, P05	P01, P07	P01, P09	P05, P07	P05, P09	P07, P09	P01,05,07	P01,07,09	P05,07,09	P01,05,09	P01,05,07,09
Ncapg2	0	0	0	0	0	1	0	0	0	0	0
Kif7	0	0	0	0	0	1	0	0	0	0	0
Hist1h1a	0	0	0	0	0	1	0	0	0	0	0