Evolution of Genes Involved in Mammalian Reproduction and Sex Determination

by

Ondrej Podlaha

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2007

Doctoral Committee:

Associate Professor Jianzhi Zhang, Chair
Professor Gerald R. Smith
Professor Priscilla K. Tucker
Associate Professor Phillip Myers
Assistant Professor Yin-Long Qiu

© Ondrej Podlaha 2007
All Rights Reserved

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

APPENDIX

INTRODUCTION

Sexual reproduction is a fundamental biological process common among eukaryotes. Many aspects of reproduction are facilitated by proteins that control crucial cellular events such as mating type determination, spermatogenesis, gamete recognition, sperm-egg binding, *etc*.. Contrary to their essential role, reproductive proteins are among the fastest evolving genes known to us. For example, lysin, a sperm protein of marine invertebrates, evolves up to 50 times faster than the most rapidly evolving mammalian gene (Metz, Robles-Sikisaka, and Vacquier 1998). Genome wide comparisons of >1800 human- mouse orthologous genes have shown that many genes directly related to gamete adhesion rank in the top 5% of the most divergent genes (Makalowski, Zhang, and Boguski 1996). Likewise, accessory gland proteins, which are part of *Drosophila* seminal fluids, are twice as diverse between species as non-reproductive *Drosophila* proteins (Civetta and Singh 1995). The list of similar cases is very long. The emerging picture here is that reproductive proteins are characterized by rapid evolution across all taxa.

This finding raises several intriguing questions. What are the selective forces governing reproductive protein evolution and what are the functional consequences of such rapid divergence? Answers to these questions have important implications in the field of evolutionary biology, as well as practical application in human fertility and health. Experimental studies show that changes in the functional domains of

reproductive proteins are sufficient to create reproductive barriers (Lyon and Vacquier 1999; Sainudiin et al. 2005). These processes can consequently have broad impact ranging from the variation in rates of speciation to the success rate of *in vitro* fertilization. In humans, for example, more than 10% of *in vitro* fertilization trials result in failure (Liu et al. 2001). Molecules that direct gamete recognition, sperm-egg binding, sperm-egg fusion –which we still know very little about – are likely to be the causal factors of reproductive deficiency. It is therefore of great interest to characterize the molecules that are involved in fertilization and the driving forces that govern their evolution.

Herein, I present case studies of genes whose function ties in closely with sperm motility, sperm-egg adhesion and sex-determination and propose scenarios and mechanisms by which natural selection shaped their evolution. In chapters 1 and 2, I document the first known case of positive selection acting on the length of a protein (CATSPER1) both in primates and rodents (Podlaha and Zhang 2003). CATSPER1 is a voltage-gated calcium channel that is expressed on the plasma membrane of the principal piece of sperm tail and its function is essential in sperm motility. My work suggests that positive selection affected the inactivation rate of this channel through the length variation of its N-terminus and consequently influenced sperm motility, which is an important determinant in sperm competition. In chapter 3, I trace the evolution of SED1. SED1 is secreted by the initial segment of caput epididimis and is an important factor in sperm-egg binding. I provide evidence that SED1 experienced structural modification in ancestral primates and describe the evolutionary forces accompanying this functional shift (Podlaha, Webb,

and Zhang 2006). And lastly, in chapter 4, I investigate an anomaly in sex determination system found among akodont rodents. Mammals have heterogametic sex-determination, which means that XY individuals develop into males whereas XX individuals develop into females. This mechanism is generally conserved across all therian mammals. In rare cases, however, the genetic sex does not correspond to the phenotypic sex. Such individuals have varying degrees of secondary sexual characteristics and are most often sterile (Camerino et al. 2006). Contrary to these general principles, naturally occurring population in eight species of akodont rodents have high frequency of XY sex reversed females that are fully fertile. In chapter 4, I analyze akodont *Sry* gene, the major testis-determining factor, to elucidate possible molecular mechanism that enables the development of fertile sex reversed female. Furthermore, it has been postulated that this phenomenon evolved multiple times in *Akodon* lineage. I provide strong phylogenetic evidence arguing otherwise and propose a new model for the propagation of sex reversal in *Akodon*.

CHAPTER I

POSITIVE SELECTION ON PROTEIN-LENGTH IN THE EVOLUTION OF A PRIMATE SPERM ION CHANNEL

Introduction

There have been dozens of reports on detection of positive Darwinian selection at the DNA sequence level (Hughes 1999; Yang and Bielawski 2000; Wolfe and Li 2003a) since the pioneering work by Hughes and Nei on mammalian Major-Histocompatibility-Complex genes (Hughes and NEI 1988a). The majority of the positively selected genes are involved in host-pathogen interactions (Hughes and NEI 1988a; Tanaka and NEI 1989; Fitch et al. 1991; Mindell 1996; Zhang, Dyer, and Rosenberg 2000) or reproduction (Lee, Ota, and Vacquier 1995b; Swanson and Vacquier 1995; Metz and Palumbi 1996; Tsaur and Wu 1997; Rooney and Zhang 1999b; Wyckoff, Wang, and Wu 2000b; Yang, Swanson, and Vacquier 2000; Swanson, Aquadro, and Vacquier 2001; Swanson and Vacquier 2002b; Galindo, Vacquier, and Swanson 2003), while a small number of the genes are of other functions (Zhang, Zhang, and Rosenberg 2002b; Goldberg et al. 2003). In all these cases, positive selection has been shown to promote nonsynonymous (amino acid-replacing) nucleotide substitutions that are presumably advantageous. In theory, certain insertion/deletion (indel) mutations in protein-coding regions may also be advantageous and subject to positive selection. Naturally occurring polymorphisms

of indels that alter protein function have been reported (Osterberg et al. 2002). However, there has been no evidence for the operation of positive selection promoting fixations of indel mutations. This is probably because a large proportion of indel substitutions would disrupt the reading frame of a gene and thus be subject to strong purifying selection, which makes it difficult to detect positive selection. We here, nevertheless, provide evidence for the operation of positive selection on indel substitutions in the primate CATSPER1 gene, and demonstrate that positive selection plays a role in the evolutionary change of protein-length.

CATSPER1 is a voltage-gated calcium ion channel that is exclusively found in the plasma membrane of the principal piece of the sperm tail (Ren et al. 2001). It is necessary for cAMP-induced Ca^{2+} influx, normal sperm motility, and penetration of the egg (Thompson et al. 1997a). Targeted disruption of the gene results in sperm immobility and male infertility in mice (Thompson et al. 1997a). The CATSPER1 protein contains an intracellular N-terminus region, 6 transmembrane domains, a pore-forming domain, and an intracellular C-terminus (Thompson et al. 1997a). In an alignment of the putative orthologous CATSPER1 sequences from the human and mouse, we noticed that the N-terminus region (mostly encoded by exon 1) contains multiple gaps and a large number of amino acid differences, while the rest of the sequences are conserved (Supplementary Fig. 1 – See APPENDIX A for chapter 1 supplementary figures and tables). Such a high frequency of gaps is unusual for orthologous mammalian proteins, which prompted us to examine this region in detail. Below, we describe the results from an analysis of 16 orthologous

CATSPER1 sequences from primates and demonstrate the action of positive selection on indels in this gene.

Materials and Methods

PCR and sequencing

Exon 1 (~1,200 nucleotides) of the CATSPER1 gene was amplified by polymerase chain reaction from genomic DNAs of the common chimpanzee (*Pan troglodytes*), pygmy chimpanzee (*Pan paniscus*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), talapoin monkey (*Miopithecus talapoin*), rhesus monkey (*Macaca mulatta*), baboon (*Papio hamadryas*), African green monkey (*Cercopithecus aethiops*), colobus monkey (*Colobus guereza*), woolly monkey (*Lagothrix lagotricha*), owl monkey (*Aotus trivirgatus*), squirrel monkey (*Saimiri sciureus*), spider monkey (*Ateles geoffroyi*), cotton-top tamarin (*Saguinus oedipus*), and ring-tailed lemur (*Lemur catta*). The PCR products were purified and cloned into pCR4TOPO vector (Invitrogen, Carlsbad, CA) before being sequenced in both directions using an automated DNA sequencer. In addition, we amplified and sequenced intron 1 (2050 nucleotides) of the rhesus monkey CATSPER1 gene. The primer sequences are given in Supplementary Fig. 2.

Sequence analysis

The protein sequences for CATSPER1 exon 1 from the above 15 primates and the human (GenBank accession number: NM_053054) were aligned using CLUSTAL X (Thompson et al. 1997a). The DNA sequences were then aligned

following the protein alignment. To examine the robustness of the alignment, we used a variety of penalty parameters for gap opening (go) and gap extension (ge): go=10, ge=0.2; go=20, ge=0.2; go=20, ge=0.4; go=40, ge=0.8; go=40, ge=0.2; go=10, ge=0.8; go=80, ge=0.2; go=60, ge=0.2; go=10, ge=10; and go=10, ge=20. The default parameters in the program were go=10 and ge=0.2. The phylogeny of the 16 primate species studied here is relatively well established, especially for the major divisions (Goodman et al. 1998; Page and Goodman 2001; Singer et al. 2003; Steiper and Ruvolo 2003), and use of alternative trees does not change our conclusion. We mapped the indels observed in the alignment of the CATSPER1 sequences to this phylogeny using the parsimony principle and counted the number of indel substitutions in each branch of the tree.

To compare the rates of synonymous and nonsynonymous nucleotide substitutions, we implemented the modified Nei-Gojobori method (Zhang, Rosenberg, and Nei 1998b) in a sliding window analysis using MEGA2 (Kumar et al. 2001). The window size was set to be 20 codons. In addition, we used the likelihood (Yang et al. 2000b) and parsimony (Suzuki and Gojobori 1999) methods to identify codons that are under positive selection. Ancestral gene sequences were inferred by the Bayesian method (Yang, Kumar, and Nei 1995b; Zhang and Nei 1997). Rates of conservative and radical nonsynonymous substitutions were estimated by the method of Zhang (Zhang 2000b). WinPep software (Hennig 1999) was used to identify intracellular N- and C- termini and transmembrane domains of CATSPER1 and other 6-transmembrane voltage-gated channels.

Results

The primate CATSPER1 exon 1 contains many indels

The exon 1 sequences of the CATSPER1 gene from 15 non-human primates were determined and compared with the sequence from the human. All sequences have an open reading frame throughout the exon as expected. But the sequence length of the exon varies among species, from 360 codons in the lemur to 443 codons in the orangutan. These sequences were conceptually translated and aligned by CLUSTAL X with the default parameters (Fig. 1.1), and the DNA sequences were subsequently aligned following the protein alignment. A gene tree of the 16 sequences was reconstructed using the neighbor-joining method (Saitou and NEI 1987b), which shows branching patterns (Supplementary Fig. 3) that are largely consistent with the known species tree (Fig. 1.2), indicating that the sequences obtained are orthologous to each other. Using the parsimony principle, we inferred events of indel substitutions in CATSPER1 exon 1 and mapped them onto the species tree. Note that parsimony makes our inference of the total number of indels conservative. Multiple parsimonious solutions are weighted equally. A total of 31 indel substitutions were found throughout the tree (Fig. 1.2). To investigate the robustness of this result, we used a wide variety of penalty parameters in alignment (see Materials and Methods). The resulting number of indels for the entire tree varied from 26 to 34. But, by our judgment, the alignment with 31 indels shown in Fig. 1.1, which was obtained using the default parameters, appears most reasonable, and further analysis is based on this alignment. However, our conclusion is valid even when alignments with fewer indels are used (see below).

Indels in CATSPER1 are under positive selection

To test whether the rate of indel substitutions in CATSPER1 exon 1 is significantly higher than the neutral expectation, it is necessary to first estimate the neutral rate of indel substitutions. For this, we used a recently published human-chimpanzee genomic comparison by Britten (Britten 2002a). In this comparison, 1019 indels were found in an alignment of 779,142 nucleotides. Because only 1.1-1.4% of the human genome contains protein-coding sequence (Human Genome Sequencing Consortium 2001), this alignment is largely comprised of noncoding sequences and may thus be regarded as neutrally evolving regions. The earliest hominid fossil known to date has an age of 6-7 million years (MY) (Brunet et al. 2002). We thus assume that the human and chimpanzee diverged about 6.5 MY ago. The neutral indel substitution rate is then estimated to be $1{,}019/779{,}142/(6.5\times10^{6}\times2)=(1.01\pm0.03)\times10^{-10}$ per site per year. Because we will compare the indel rates between noncoding and coding regions, it is more relevant to compute the neutral substitution rate of indels with sizes of multiples of 3 nucleotides (3*n* indels), as only 3*n* indels are potentially non-deleterious when they occur in protein-coding regions. In the above human-chimpanzee genomic data, there are 194 3*n* indels. Thus, the neutral substitution rate for 3*n* indels is $194/779{,}142/(6.5\times10^{6}\times2)=(1.92\pm0.14)\times10^{-11}$ per site per year. The reason that the number of 3*n* indels is smaller than one third the number of all indels is that the frequency of indels declines quickly with the increase of the indel size and there are many more indels with 1 (or 2) nucleotide(s) than with 3 nucleotides (Britten 2002a).

In addition to the use of the Britten data, we also used the result from Silva and Kondrashov, who conducted a genomic comparison between human and baboon for 1,448,332 nucleotides (Silva and Kondrashov 2002b). They identified 5883 indels, of which 1001 were 3*n* indels. Assuming that humans and baboons diverged 23 MY ago (Goodman et al. 1998; Glazko and Nei 2003), we estimated from their data that the neutral substitution rate for indels is $(8.83\pm0.12)\times10^{-11}$ per site per year and that for 3*n* indels is $(1.50\pm0.05)\times10^{-11}$ per site per year.

With these estimates of genomic neutral substitution rates for 3*n* indels, we computed the expected number of 3*n* indels in CATSPER1 exon 1 under the assumption that all 3*n* indels are neutral. Using the estimate from the human-chimpanzee genomic comparison, the expected number of 3*n* indels in exon 1 between hominoids and Old World (OW) monkeys is $1.92\times10^{-11}\times1329\times23\times10^{6}\times2=1.17$ (Table 1). Here 1329 is the number of nucleotides in the longest exon 1 sequence (orangutan) of the 16 primates. Use of this number makes our statistical test more conservative. The observed average number of 3*n* indels in exon 1 between the 5 hominoids and 5 OW monkeys is 6, which is significantly greater than the expected value of 1.17 under neutral evolution ($P<0.001$, Poisson test, Table 1). Similarly, the comparison between hominoids and New World (NW) monkeys and that between OW and NW monkeys yielded the same conclusion (Table 1). Use of the neutral rate estimated from the human-baboon genomic data shows even higher statistical significance (Table 1). Use of the alignment with the smallest number (*i.e.*, 26) of gaps gave similar results (Supplementary Table 1). These comparisons strongly suggest that 3*n* indels are positively selected for in the

evolution of primate CATSPER1 exon 1. In the above tests, we assumed that the rate of indel mutations at the CATSPER1 locus is similar to the genomic average. To verify this assumption, we sequenced the rhesus monkey CATSPER1 intron 1, which is adjacent to exon 1, and compared it with the orthologous human sequence. We found 9 indels in 2102 aligned sites of intron 1, which translates into a rate of $9/2102/(23\times10^{6}\times2)=(9.31\pm0.96)\times10^{-11}$ indels per site per year. This rate is very close to the rates estimated from the two genomic comparisons (1.01×10^{-10} and 8.83×10^{-11}), suggesting no elevation of the indel mutation rate at the CATSPER1 locus.

Longer indels are selectively favored

We further investigated whether indels of certain lengths are particularly favored in CATSPER1 by comparing the (3*n*) indel-size distributions for the CATSPER1 data and the two genomic data sets used above (Fig. 1.3). A significant distributional difference is detected between CATSPER1 and either of the two genomic data sets ($P<10^{-19}$, χ^2 test). Longer indels are preferentially selected for in CATSPER1. For instance, the proportion of 3*n* indels with 15 or more nucleotides is 8-9% in the two genomic data, but 58% in the CATSPER1 data. However, even for indels of 3 nucleotides, the number of observed indels in CATSPER1 is about 2.5 times the expected number from the genomic data, and their difference is statistically significant ($P<0.02$). This suggests that both short and long indels are selectively favored in CATSPER1, with longer ones being under stronger positive selection.

It should be noted that some sequence motifs such as simple nucleotide repeats are known to have relatively high rates of mutation (Hancook 1999b). Most of these mutations are due to slippage in DNA replication, which results in addition or deletion of one or occasionally a few repeats (Hancook 1999b). The CATSPER1 exon1 sequences do not contain such repetitive sequences except for a few regions with 3-4 histidines in a row. However, the indel-size distribution in Fig. 1.3 and the amino acid sequence alignment in Fig. 1.1 indicate that the majority of the indels in CATSPER1 were not due to mutations of simple repetitive sequences.

Amino acid substitutions may also be under positive selection

The alignment of the CATSPER1 exon 1 sequences from the 16 primates also reveals a relatively high number of amino acid substitutions (Fig. 1.1). To test whether the amino acid substitutions are under positive selection, we estimated the number of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions per site between each pair of the sequences. We found that the average d_N/d_S ratio is 1.05 from a total of 120 pairwise comparisons, with 46 of these comparisons showing $d_N > d_S$ (Supplementary Fig. 4). We further characterized the d_N/d_S ratio by a sliding-window analysis with a non-overlapping window size of 60 nucleotides. This analysis identified 9 out of 17 windows that have an average d_N greater than d_S, one of them being marginally significant (Fig. 1.4). We also conducted a likelihood analysis, which may be more powerful in detecting positive selection for nonsynonymous substitutions at individual codons (Yang et al. 2000b). Specifically, we compared a null model M7 with a more general model M8. M7 assumes that the

d_N/d_S ratio across codon sites follows a beta distribution between 0 and 1, while M8 adds to M7 an extra class of sites with any d_N/d_S. We found that M8 fits the data significantly better than M7 (χ^2=37.3, df=2, $P<10^{-8}$), with an additional class of sites of d_N/d_S=2.98. Five codons were identified to be under positive selection with posterior probabilities higher than 95% (Fig. 1.1). Similar results were obtained when models M1 and M2 were compared (see Yang et al. 2000 for details of the model description). Because the likelihood method is known to occasionally generate false positive results (Hancook 1999b), we also used a more conservative test based on parsimony (Suzuki and Gojobori 1999). With this test, none of the five aforementioned codons show significantly higher d_N than d_S. But when the five codons are tested together, the average d_N is significantly greater than d_S ($P<0.02$). These tests thus are consistent in suggesting some sort of positive selection for amino acid substitutions. To further characterize the substitutions that are favored by selection, we compared the rates of conservative and radical nonsynonymous substitutions with regard to amino acid polarity (Zhang 2000b). For this analysis, the ancestral sequences at all interior nodes of the tree (Fig. 1.2) were inferred by the Bayesian method, and the numbers of conservative and radical nonsynonymous substitutions were counted for each tree branch. For the entire tree, we observed 188 radical and 373 conservative nonsynonymous substitutions, respectively. The potential numbers of radical and conservative nonsynonymous sites are 195 and 478, respectively. Thus, the substitution rate at radical sites (188/195=0.96) is 1.23 times that at conservative sites (373/478=0.78), and their difference is statistically significant ($P<0.02$, binomial test). This pattern is in contrast to that observed from

a majority of mammalian genes (Zhang 2000b), suggesting positive selection favoring changes of the polarity profile of CATSPER1. A similar analysis did not yield significant results when amino acid charge is considered.

Discussion

By analyzing DNA sequences from 16 primates, we discovered exceptionally frequent incidences of indel substitutions in the evolution of the first exon of the CATSPER1 gene. In all likelihood the first exon is a functional part of CATSPER1, as no frame-shifting indels or nonsense substitutions are found in any of the species examined. We found that the indel substitution rate in exon 1 is 5-8 times that of the genomic average, which represents the neutral rate. Furthermore, larger indels (≥15 nucleotides) are significantly more prevalent in exon 1 than in the neutral genomic regions. These observations provide strong evidence that indel substitutions, particularly those with greater sizes, are positively selected for in the evolution of primate CATSPER1.

Why would indel substitutions be beneficial in CATSPER1 exon 1, which encodes the intracellular N-terminus region of the ion channel? To address this question, we turn to the structure and function of ion channels. Ion channels are transmembrane proteins that form pores through which ions can pass. A voltage-gated ion channel such as CATSPER1 is activated by depolarization (reduction in electric potential) of the cell membrane, which causes a conformational change of the channel and allows ions to pass through it (Fig. 1.5). Within 1 millisecond of activation, the channel is inactivated and is impermeable to the ions, even though the

membrane is still depolarized. The membrane must be repolarized or hyperpolarized to remove the channel from the inactive state and return it to the closed state where it is prepared for subsequent activation. Inactivation prevents the channel from remaining open, and is also responsible for the unidirectional propagation of action potential. The "ball-and-chain" model of ion channel inactivation (Fig. 1.5), proposed by Bezanilla and Armstrong (1977b) and demonstrated by Aldrich and colleagues (Hoshi, Zagotta, and Aldrich 1990b; Zagotta, Hoshi, and Aldrich 1990b), offers a possible scenario where the length of the N-terminus region plays an important functional role. Specifically, Aldrich and colleagues showed that the N-terminus of a *Drosophila* voltage-gated potassium (K_V) channel named *Shaker* acts to inactivate the channel (Hoshi, Zagotta, and Aldrich 1990b; Zagotta, Hoshi, and Aldrich 1990b). Here, a "ball on a chain" structure is located at the N-terminus of the channel and acts as a tethered plug, which is able to physically block the intracellular end of the ion channel pore region and cause inactivation of the channel (Hoshi, Zagotta, and Aldrich 1990b; Zagotta, Hoshi, and Aldrich 1990b) (Fig. 1.5). The first ~20 residues of the N-terminus of *Shaker* channel form the intracellular "plug" and the next ~60 residues represent the "tether" (Hoshi, Zagotta, and Aldrich 1990b; Zagotta, Hoshi, and Aldrich 1990b). It was found that the length of this tethered plug controls the rate of channel inactivation (Hoshi, Zagotta, and Aldrich 1990b). That is, lengthening or shortening of the tether resulted in slow or rapid channel inactivation, respectively. This is probably because a shorter tether restricts the space in which the "plug" wanders, making it easier for the "plug" to find the pore. Although this "ball-and-chain" model has only been demonstrated in K_V

channels, it is possible that CATSPER1 has a similar mechanism of regulating its inactivation. In fact, structurally, CATSPER1 resembles K_V channels more than voltage-gated Ca (Ca_V) or Na (Na_V) channels, as CATSPER1 and K_V channels are each formed by 4 identical peptides, each having a single, 6-transmembrane-spanning repeat, whilst Ca_V and Na_V channels are made of a single peptide with 4 repeats of 6-transmembrane-spaning regions. The amino acid sequence of the pore-forming region, however, is more similar between CATSPER1 and Ca_V, presumably reflecting the identical ion selectivity. The hydropathy profile shows a greater similarity of CATSPER1 to K_V than to Ca_V or Na_V channels (Supplementary Fig. 1.5). Evolutionarily, it is generally believed that K_V channels originated before Ca_V and Na_V channels (Lodish et al. 2000) and that metazoan Ca_V and Na_V channels each form a monophyletic group in exclusion of K_V channels (Piontkivska and Hughes 2003). To investigate the phylogenetic position of CATSPER1 in relation to other 6-transmembrane voltage-gated channels, we reconstructed a phylogeny using the human and mouse sequences of N-termini regions of several K_V, Na_V, and Ca_V channels (Fig. 1.6). The tree is consistent with the current understanding of the evolution of K_V, Na_V, and Ca_V channels (Lodish et al. 2000; Hille 2001; Piontkivska and Hughes 2003). Surprisingly, CATSPER1 does not cluster with other Ca_V channels, but resides outside the monophyletic group of Ca_V and Na_V channels (Fig. 1.6). This suggests that CATSPER1 is one of the earliest branches splitting from K_V channels, originating before the divergence of other Ca_V channels and Na_V channels. This would further suggest that the emergence of the structure of a single peptide with 4 repeats that is seen in non-CATSPER1 Ca_V channels and Na_V channels

postdated the origin of CATSPER1. It is unlikely that the branching pattern in Fig. 1.6 is due to long-branch attraction between CATSPER1 and K_V channels, because CATSPER1 does not cluster with the longest branch-the outgroups (legend to Fig. 1.6) and the same branching pattern is also observed when the non-N-terminal sequences, which are conserved in CATSPER1, are used in tree-making. Taken together, the evolutionary and structural analyses suggest similarity of CATSPER1 to K_V channels, which makes the "ball-and-chain" a more plausible model of channel inactivation for CATSPER1. If this model indeed works in CATSPER1, the indels in the N-terminus region can potentially affect the inactivation rate of the channel, as in the *Drosophila* K_V channel *Shaker*. Because CATSPER1 determines sperm motility by regulating the cellular Ca^{2+} concentration (Ren et al. 2001), it is likely that the rate of channel inactivation influences sperm motility. As sperm motility is one of the most important factors in sperm competition (Gomendio, Harbourt, and Roldan 1998), the exceptionally high rate of indel substitutions in CATSPER1 may be a signature of intense sperm competition. A population genetic study will be useful to further test this hypothesis. Our preliminary data from mice (*Mus musculus*) show intraspecific indel polymorphisms in the first exon of Catsper1 (J.Z. and colleagues, unpubl.).

In addition to indel substitutions, our results indicate that amino acid substitutions in the N-terminus region of CATSPER1, particularly those that alter amino acid polarity, are probably under positive selection as well. It is possible that such changes in hydrophobicity affect the folding of the N-terminus region and influence the rate of channel inactivation. In the future, it would be interesting to

use *in vitro* and *in vivo* assays to investigate the functional consequences of the indel and amino acid substitutions in the N-terminus of CATSPER1.

The phylogenetic tree in Fig. 1.6 also shows that the level of sequence divergence in the N-terminus region between the human and mouse orthologous CATSPER1 proteins is exceptionally high in comparison to that in other voltage-gated ion channels. All these channels are expressed in somatic tissues with the exception of CATSPER1, which is solely expressed in sperm. Recently, a second gene encoding a sperm calcium channel has been cloned and it is named CATSPER2 (Quill et al. 2001). Our preliminary analysis suggests a remote evolutionary relationship between CATSPER1 and CATSPER2 (data not shown). Furthermore, the N-terminus region of CATSPER2 appears conserved between the human and mouse, with only a few amino acid substitutions and virtually no indels. Thus, the physiological functions of CATSPER1 and CATSPER2 might be different.

Positive selection for nonsynonymous nucleotide substitutions has been documented in many genes (Hughes 1999; Yang and Bielawski 2000; Wolfe and Li 2003a). To our knowledge, CATSPER1 represents the first case in which positive selection for indel substitutions is detected. This success largely relies on the availability of genomic sequence data from closely related species, from which a neutral rate of indel substitution can be reliably estimated. From the present study, it seems that the statistical test for detecting selection on indels is relatively powerful. For instance, the number of expected 3*n* indels is about 1 for the CATSPER1 exon 1 sequences between hominoids and OW monkeys (Table 1). Under the assumption that the number of indels follows a Poisson distribution, an observation of 4 indels

would lead to the rejection of the null hypothesis of neutral evolution, with a statistical confidence of 98%. Because protein-length variation among orthologs and paralogs is quite common and indels are often seen in protein sequence alignments, we hypothesize that positive selection for indels is not rare. With the establishment of the basic methodology here and the estimation of neutral rates of indel substitutions from many more species, this hypothesis can be tested in the near future.

Acknowledgments

Results of this work have been published in the Proceedings of the National Academy of Science of the United States of America (Podlaha and Zhang 2003). We thank Priscilla Tucker, Xiaoxia Wang, David Webb, and the anonymous reviewers for valuable comments. This work was supported by a startup fund and a Rackham fellowship from the University of Michigan and by National Institutes of Health Grant GM67030 (to J.Z).

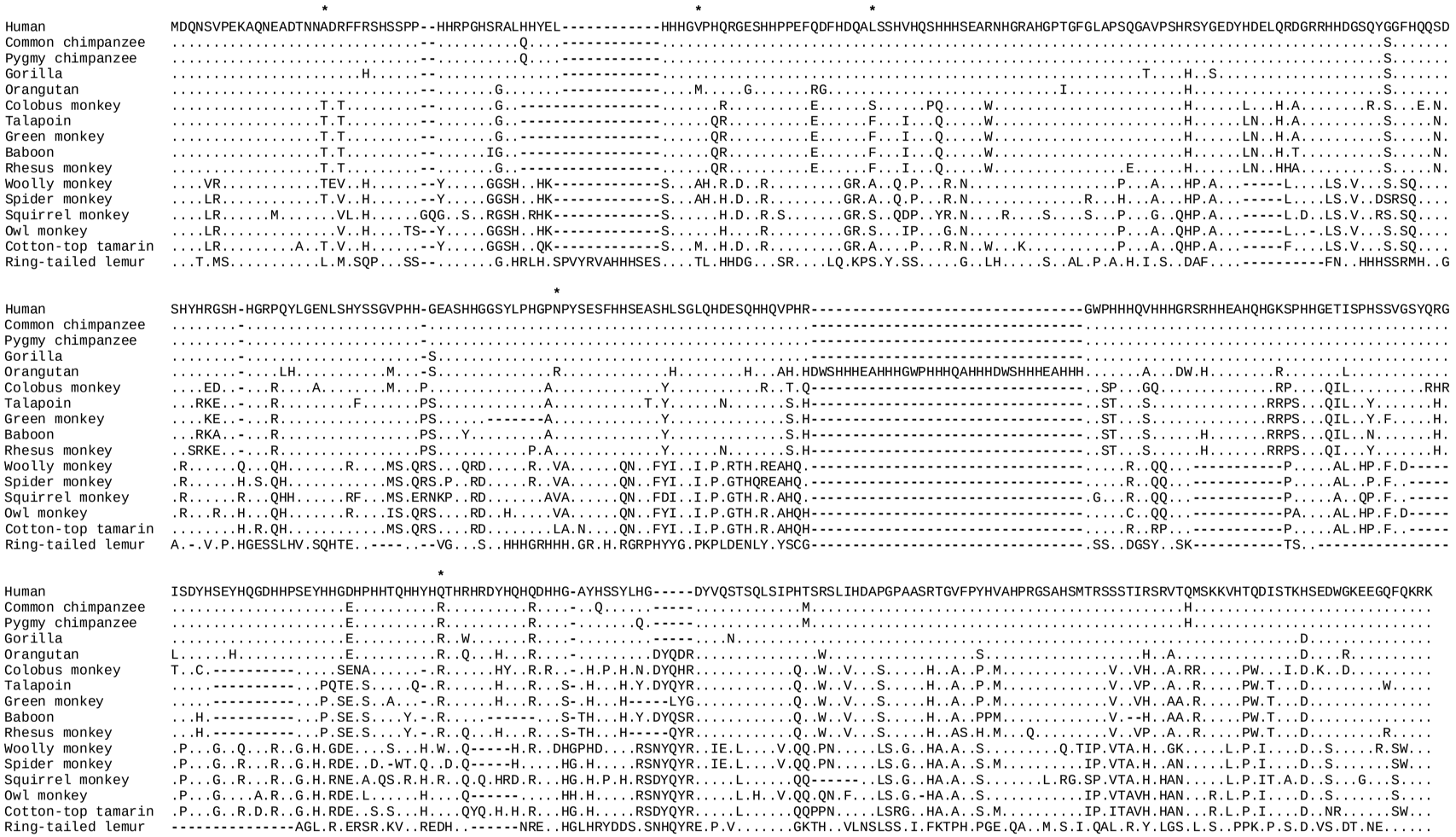
Table 1.1 Comparison of the substitution rates of 3*n* indels from genomic data and from primate CATSPER1 exon 1.

		Indel rate (per site per 10^{11} year)			Number of indels in CATSPER1	
Comparisons	Divergence [a] (MY)	From ref. (37)	From ref. (40)	From CATSPER1	Expectation from ref. (37)	Expectation from ref. (40)
Hominoids vs. OW monkeys	23↔2	1.92	1.50	9.81	1.17	0.92
Hominoids vs. NW monkeys	35↔2	1.92	1.50	10.3	1.79	1.40
OW vs. NW monkeys	35↔2	1.92	1.50	12.5	1.79	1.40

[a] The divergence times follow ref. (41).

[b] The probabilities of the observation given the expectation calculated from ref. (37) and (40) are computed under the assumptio of indels follows a Poisson distribution.

FIG. 1.1. Amino acid sequence alignment of the exon 1 of CATSPER1 from 16 primates. This alignment is derived from CLUSTAL X with default parameters. "." represents an identical amino acid to the first sequence and "–" represents an alignment gap. "*" show five sites at which positive selection for amino acid substitutions is inferred by the likelihood method with >95% posterior probabilities (see text).


Human
Common chimpanzee
Pygmy chimpanzee
Gorilla
Orangutan
Colobus monkey
Talapoin
Green monkey
Baboon
Rhesus monkey
Woolly monkey
Spider monkey
Squirrel monkey
Owl monkey
Cotton-top tamarin
Ring-tailed lemur

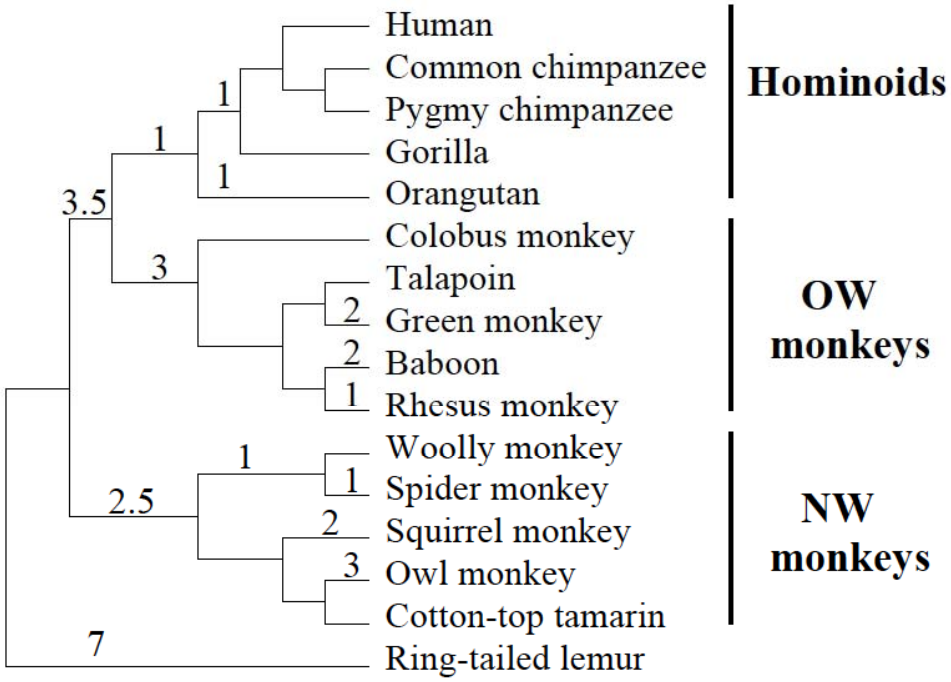


FIG. 1.2. Phylogeny of the 16 primates studied here. The numbers on tree branches are parsimony-inferred numbers of indel substitutions in CATSPER1 exon 1, based on the alignment shown in Fig. 1.1.

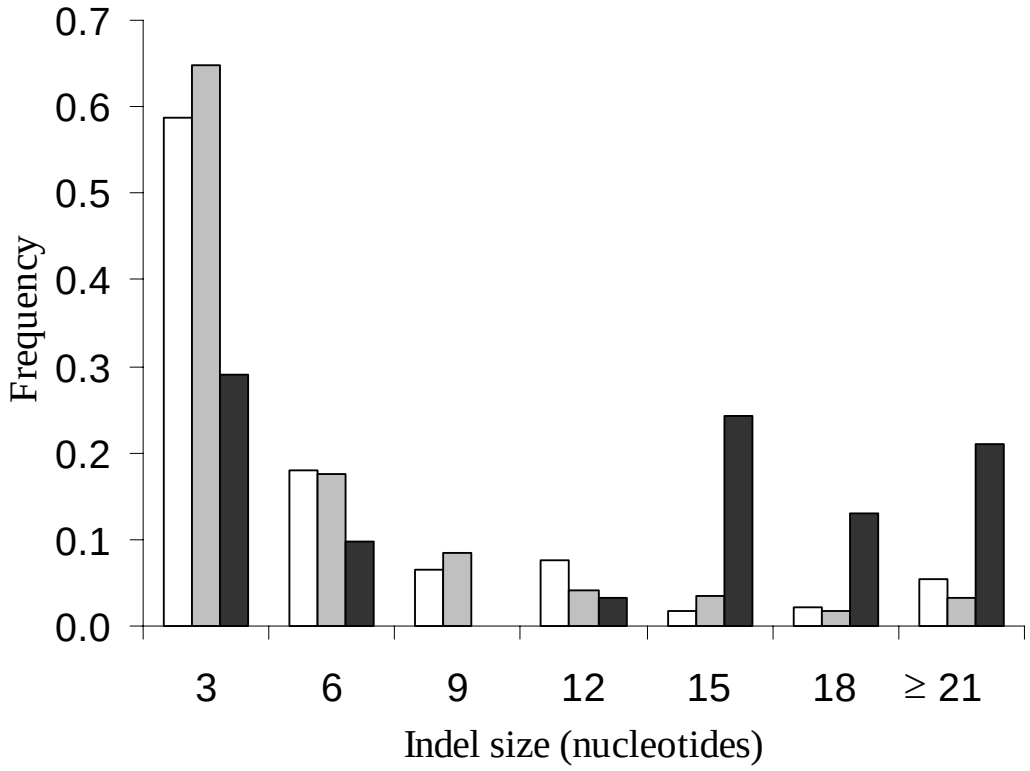


FIG. 1.3. Size distribution of 3*n* indels. White and gray bars represent 3*n* indels of the genomic data from ref. (37) and (40), respectively, and black bars represent 3*n* indels from CATSPER1 exon 1 sequences of 16 primates. The frequencies are calculated by the number of indels of a particular size divided by the total number of 3*n* indels.

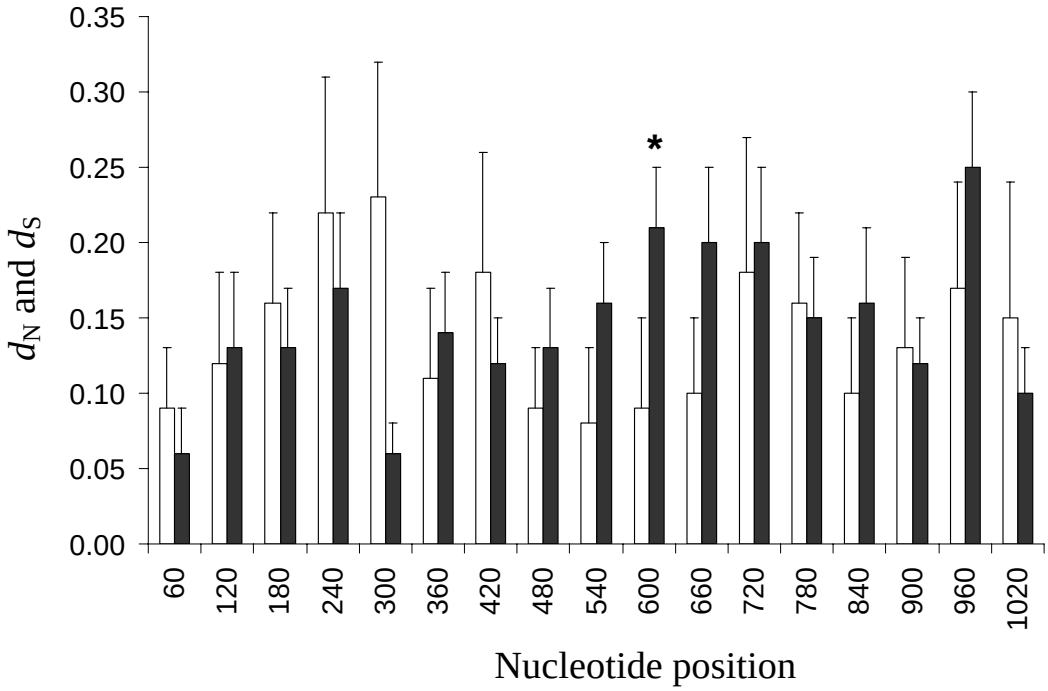


FIG. 1.4. Sliding-window analysis of CATSPER1 exon 1 sequences from 15 primates. The lemur sequence was not used here because of the presence of a large number of indels. The non-overlapping window size is 20 codons. Average numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions per site among the 15 sequences are shown by open and closed bars, respectively, with the error bars representing one standard error. One window has a marginally significant $d_N > d_S$ (Z test, P=0.053) and it is indicated by "*". The nucleotide positions are from an alignment of the 15 sequences with all the indels removed, and thus do not directly correspond to the positions in Fig. 1.1.

FIG. 1.5. Schematics of the “ball-and-chain” model of channel inactivation. **(A)** Closed state, **(B)** open state, **(C)** inactivated state. According to this model, the N-terminus represents a tethered plug that can physically block the intracellular side of the ion channel pore and cause inactivation. Different lengths of the N-terminus region result in different rates of channel inactivation (45, 46), where a shorter N-terminus **(D)** causes a more rapid inactivation in comparison to a longer N-terminus **(E)**.

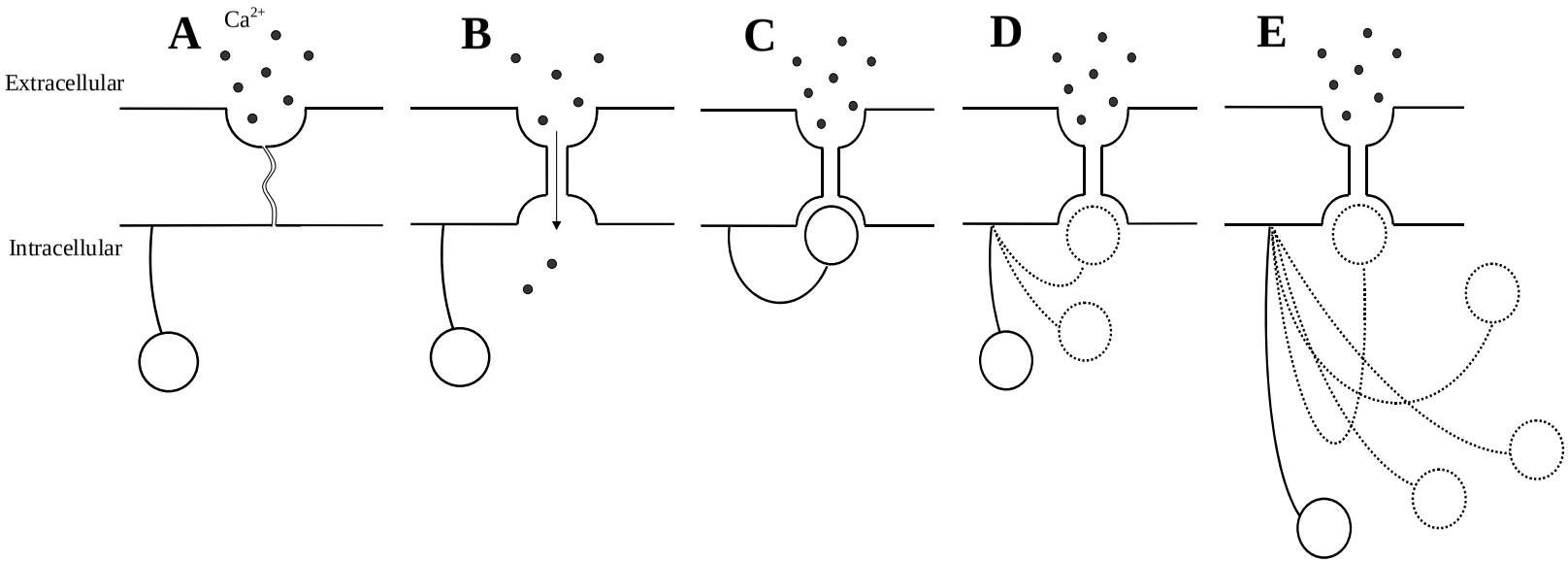
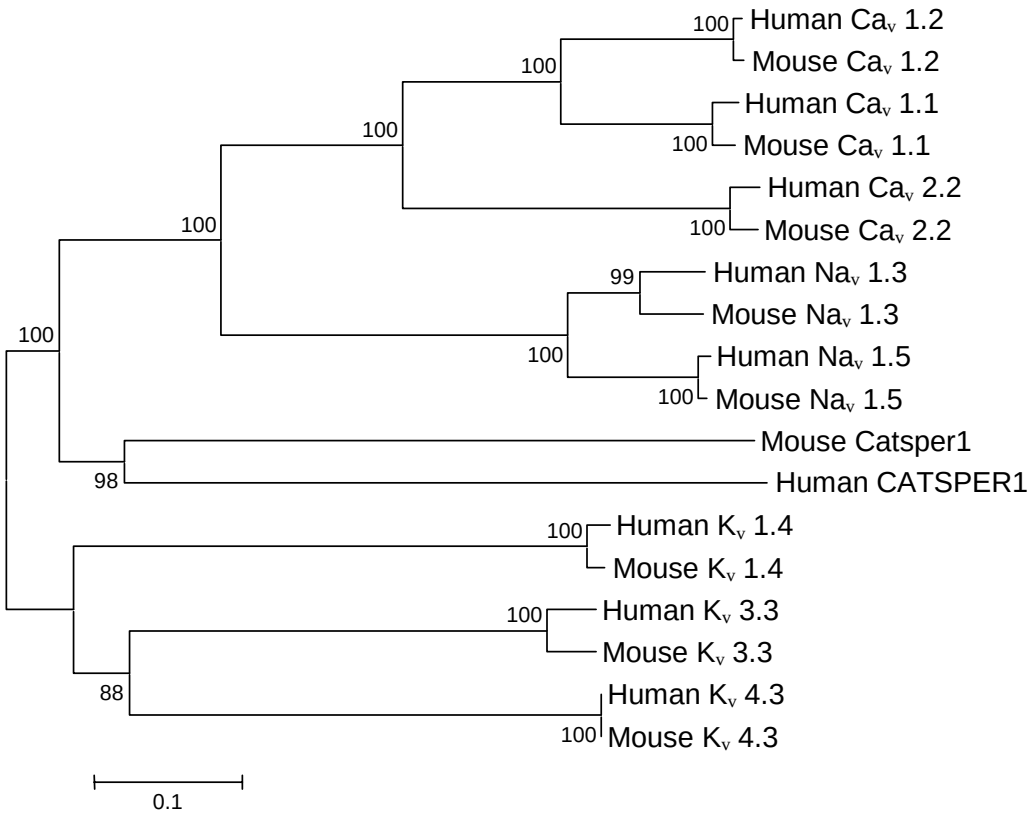

A
Ca2+
B
C
D
E
Extracellular
Intracellular

FIG. 1.6. The evolutionary relationship of CATSPER1 with mammalian K_v, Na_v, and Ca_v channels. The tree was reconstructed with the N-terminus region of each ion channel, which was determined by hydropathy analysis. The neighbor-joining method with protein *p*-distance (29) was used. Numbers at interior nodes are bootstrap percentages from 1000 replications. Branch lengths are drawn to scale (number of amino acid substitutions per site). The root of the tree was determined to be on the deepest branch shown here, by using *Drosophila* and vertebrate inward-rectifier K channels as outgroups (49). When the entire sequences of the ion channels are used, the tree topology remains the same with the exception of the interrelationships among K_V1.4, K_V3.3, and K_V4.3. The GenBank accession numbers are: human K_V1.4, A39922; mouse K_V1.4, NP_067250.1; human K_V3.3, NP_004968.1; mouse K_V3.3, Q63959; human K_V4.3, NP_004971.1; mouse K_V4.3, NP_064315.1; human Ca_V2.2, Q00975; mouse Ca_V2.2, O55017; human Ca_V1.2, NP_000710.1; mouse Ca_V1.2, A44467; human Ca_V1.1, A55645; mouse Ca_V1.1, NP_055008.1; human Na_V1.3, Q9NY46; mouse Na_V1.3, NP_035453.1; human Na_V1.5, Q14524; mouse Na_V1.5, NP_067519.1; human CATSPER1, AF407333_1; mouse Catsper1 AF407332_1.


Human Cav 1.2
Mouse Cav 1.2
Human Cav 1.1
Mouse Cav 1.1
Human Cav 2.2
Mouse Cav 2.2
Human Nav 1.3
Mouse Nav 1.3
Human Nav 1.5
Mouse Nav 1.5
Mouse Catsper1
Human CATSPER1
Human Kv 1.4
Mouse Kv 1.4
Human Kv 3.3
Mouse Kv 3.3
Human Kv 4.3
Mouse Kv 4.3
100
100
100
100
100
100
99
100
100
100
98
100
100
88
100
0.1

CHAPTER II

POSITIVE SELECTION FOR INDEL SUBSTITUTION IN THE RODENT SPERM PROTEIN CATSPER1

Introduction

The occurrence of positive Darwinian selection at the molecular level has been reported in many genes, particularly among those involved in immunity or reproduction (Hughes and Nei 1988b; Lee, Ota, and Vacquier 1995a; Rooney and Zhang 1999a; Wyckoff, Wang, and Wu 2000a; Swanson and Vacquier 2002a; Zhang and Rosenberg 2002; Zhang, Zhang, and Rosenberg 2002a; Wolfe and Li 2003b; Wang and Zhang 2004; Zhang and Webb 2004). In all of these genes, it was shown that positive selection promotes amino acid replacements. In theory, positive selection may also act upon other types of sequence changes. In a previous report, we showed that positive selection was responsible for elevated fixation rates of insertions and deletions (indels) in the primate *Catsper1* gene (Podlaha and Zhang 2003), which codes for a voltage-gated calcium ion channel essential for sperm mobility (Ren et al. 2001). Mice with the *Catsper1* gene deleted are sterile (Ren et al. 2001). Interestingly, the excess of indel substitutions was observed only in the first exon of *Catsper1*, which codes for the intracellular N-terminus of the channel. Based on the predicted structure of Catsper1 that includes 6 transmembrane domains and intracellular N-

and C- termini (Ren et al. 2001), we hypothesized that the length of the N-terminus may affect the rate of channel inactivation (Podlaha and Zhang 2003), as in the so-called “ball-and-chain” model (Bezanilla and Armstrong 1977a; Hoshi, Zagotta, and Aldrich 1990a; Zagotta, Hoshi, and Aldrich 1990a). Thus, the indels in Catsper1 may influence sperm mobility and consequently be an important factor in sperm competition (Podlaha and Zhang 2003). The mouse Catsper1 protein is found in the plasma membrane of the principle piece of the sperm tail (Ren et al. 2001). But this location does not tell whether the products of one allele or both alleles of the gene are present in an individual sperm, as it is possible that the progenitor diploid cells of sperm express both alleles whose products get incorporated into sperm. A recent study, however, showed that *Catsper1* expression is confined to the postmeiotic stage of spermatogenesis (Schultz, Hamra, and Garbers 2003), strongly suggesting that only the product of one allele is present in each sperm. This provides the possibility of haploid selection on (Joseph and Kirkpatrick 2004) and meiotic drive of (Hartl and Clark 1997) *Catsper1* if the two alleles confer different fitnesses. These hypotheses about the function of and selection on *Catsper1* are difficult to test in primates because of the difficulties in housing animals and in obtaining biological materials. Rodents would be more convenient and suitable for addressing these questions. We here sequence the *Catsper1* gene in multiple species belonging to the rodent subfamily Murinae of family Muridae and multiple individuals of the species *Mus musculus* and *M. macedonicus*. We show that the phenomenon of accelerated indel substitutions observed in primates also occurs in rodents. This

opens the possibility for experimental examination of the functional and evolutionary significance of the indel substitutions in rodent Catsper1.

Materials and Methods

PCR and sequencing

Exon 1 (933 nucleotides in *Mus musculus*) of the *Catsper1* gene was amplified by polymerase chain reaction (PCR) from the genomic DNAs of *Mus pahari*, *M. spretus*, *M. spicilegus*, *M. macedonicus* (2 individuals), *M. cervicolor*, *M. cookii*, *M. caroli*, *M. musculus musculus* (1 individual from Loppi, Finland; 1 individual from Bratislava, Slovakia; 1 individual from Studenec, Czech Republic), *M. musculus domesticus* (1 individual from Centreville, MD, USA; 1 individual from Canton Vaud, Switzerland; 1 individual from Bern, Switzerland; 1 individual from Nurnberg, Germany; 1 individual from Erfoud, Morocco; 1 individual from Tubingen, Germany), *M. musculus molossinus* (Fukuoka, Japan), and *M. musculus castaneus* (Thailand). PCR products were purified and sequenced in both directions using an automated DNA sequencer. The nucleotide sequences from the above 19 individuals, mouse strain C57BL/6J (GenBank accession no. NM_139301) and *Rattus norvegicus* (GenBank accession no. XM_219698) were aligned based on the protein sequence alignment made by CLUSTAL_X (Thompson et al. 1997b). The complete exon 1 sequences except for the first 18 and last 17 codons were obtained and analyzed.

Estimating rates of indel substitutions

We assumed that the phylogenic relationships of the rodent species used in this study are the same as presented in Fig. 2B of (Tucker, Sandstedt, and Lundrigan 2005), which was the maximum-likelihood tree from a combined analysis of eight nuclear and mitochondrial genes. The number of indel substitutions that occurred in the rodent *Catsper1* sequences was counted using the parsimony principle based on the species tree. In cases where multiple equally parsimonious solutions for a single alignment gap were encountered, all solutions were weighted equally. Because there are no unequivocal estimates of absolute divergence times among the *Mus* taxa included in our study, we calculated the rate of indel substitutions per unit of d_S distance (number of synonymous substitutions per synonymous site), rather than per unit of time. This approach allows us to make reliable comparisons of indel rates across all lineages and avoid the imprecision inherent in divergence time estimates. Branch lengths for a given tree topology were calculated using the BNBS program (Zhang, Rosenberg, and Nei 1998a). The genomic average rate of indel substitutions in mouse and rat (Rat Genome Sequencing Consortium 2004) was used as an approximation of the neutral rate of indel substitutions.

Testing for positive selection

Positive selection promoting single-nucleotide and indel substitutions in exon 1 of *Catsper1* was tested. The maximum likelihood method implemented in PAML (Yang 1997) was used to test positive selection for single-nucleotide substitutions.

We tested whether indel mutations have been fixed at a higher-than-neutral rate in *Catsper1* exon 1 by comparing it with the genomic indel rate following Podlaha and Zhang (2003). We also modified the McDonald and Kreitman's (1991) test to examine whether the ratio of the number of indels to the number of synonymous differences is the same for intraspecific and interspecific data. A significant difference between these ratios suggests variable selective pressures on indel mutations at the intraspecific and interspecific levels.

Results

We determined from the mouse (*Mus musculus*) genome sequence that *Catsper1* is located in chromosome 19 and is composed of 12 exons totaling 2061 nucleotides. A comparison of the mouse *Catsper1* and its human ortholog revealed a large number of alignment gaps in the first exon, whereas the rest of the sequences remained relatively conserved. Because our previous study in primates showed that exon 1, which codes for the majority of the N-terminal intracellular region of the ion channel, has been under positive selection for indel substitutions, we here focus on the same exon in rodents. In fact, the alignment of the mouse and rat Catsper1 sequences showed a high frequency of both amino acid replacements and indels in the region encoded by exon 1 (Fig. 2.1). We obtained DNA sequences of *Catsper1* exon 1 from 19 individuals of 9 *Mus* species. No heterozygous individuals were found. In addition, we downloaded from GenBank one *Catsper1* sequence from *Mus musculus* and one from *Rattus norvegicus*. Thirteen different nucleotide sequences and 12 different protein sequences were found among the 21 *Catsper1*

exon 1 sequences obtained (Fig. 2.2). These raw sequences varied in length from 808 to 856 nucleotides, with the open reading frame being intact in each of them. Sequence alignment by CLUSTAL_X was straightforward and robust, without the need for manual adjustments. The alignment of the sequences, spanning 910 nucleotides, contained numerous gaps ranging in number from 5 in *M. pahari* to 10 in one of the *M. musculus* individuals (Fig. 2.2). Interestingly, within the *M. musculus* species, there are two indel polymorphisms. One of them involves an indel of one codon (found in *M. m. molossinus* from Japan), whereas the other involves two codons (found in *M. m. musculus* from Finland).

A gene tree of the 21 sequences was reconstructed using the neighbor-joining method (Saitou and Nei 1987a), which shows a branching pattern that is largely consistent with the assumed species phylogeny (Fig. 2.3), indicating that the sequences are indeed orthologous to each other. Furthermore, all allelic sequences within species cluster in species-specific clades. Using the parsimony principle, we inferred the number of indels in *Catsper1* exon 1 and mapped them onto the species tree (Fig. 2.3). For example, the branch that leads to *M. musculus* has 2.17 indels, because this branch is representative of 12 individuals from the *M. musculus* species and while all of them share 2 indels, two individuals show one unique indel each in the alignment. Therefore, averaged across all 12 individuals, there are 0.17 individual-specific indels in addition to the 2 indels shared by all. Note that parsimony makes our inference of the number of indels conservative. A total of 22.47 indel substitutions were found throughout the tree (Fig. 2.3). We then calculated the average number of indel events for different parts of the tree, dividing

the taxa into groups M1 (*M. musculus, M. macedonicus, M. spicilegus*, and *M. spretus)*, M2 (*M. cervicolor, M. cookii*, and *M. caroli)*, M3 *(M. pahari)*, and R (*Rattus norvegicus*) (Fig. 2.3 and Table 1). The average number of indel substitutions from the common ancestor of a group to all terminal nodes in the group is computed by averaging the number of indels between sister lineages in the group in a hierarchical manner. For example, the number for M1 is ((((0+0)/2+0)+1)/2+0+2.17)/2=1.34. The average number of indels between species of two groups was subsequently computed by adding the number of indels between the common ancestors of the two groups and the numbers from the common ancestors to terminal nodes (Table 1). For example, the number of indels between M1 and R is 4.83+5.67+1.34 =11.84 (Fig. 2.3 and Table 1).

To test whether the rate of indel substitutions in *Catsper1* exon1 is significantly greater than the neutral expectation, it is necessary to first estimate the neutral rate of indel substitutions. For this we used the recently published mouse-rat genomic comparison (Cooper et al. 2004). It was estimated that indel substitutions in bulk genomic DNA has a frequency of 12.46 indels/kb between mouse and rat. This rate of indel substitution should be very close to the neutral rate as only 1-1.5% of the mammalian genome are coding sequences (Human Genome Sequencing Consortium 2001). Because indels of sizes other than multiples of three nucleotides (3*n*) will almost certainly be highly deleterious in coding regions, it is necessary to compute the neutral substitution rate for 3*n* indels. The proportion of 3*n* indels is less than one third of all indels because the frequency of indels declines rapidly with size in a non-linear fashion. Studies of the indel size distribution in mammalian

genomes show that approximately 17–19% of all indels are 3*n* indels (Britten 2002b; Silva and Kondrashov 2002a; Makova, Yang, and Chiaromonte 2004). In particular, the proportion of 3*n* indels for mouse and rat was 17.3% (Makova, Yang, and Chiaromonte 2004). Applying this value to the neutral rate of indel substitutions, we estimated that the neutral substitution rate of 3*n* indels is 2.156 indels/kb between mouse and rat.

The divergence times for most taxa included in our study are either unknown or ambiguous. Therefore, rather than considering substitution rates of indels per site per year, we calculated the number of indel substitutions per site per unit of synonymous distance (d_S). Using d_S instead of absolute time is a more reliable way for comparing indel substitution rates among different lineages because it allows us to avoid errors in divergence time estimation. Applying the genomic average d_S=0.19 between mouse and rat (Rat Genome Sequencing Consortium 2004), the neutral substitution rate of 3*n* indels becomes $2.156\times10^{-3}/0.19=0.01135$ per site per unit of d_S.

With the above neutral substitution rate of 3*n* indels, we computed the expected number of 3*n* indels in exon 1 of *Catsper1* if all 3*n* indels are neutral. For example, we expect to observe $0.01135\times759\times0.2648 = 2.29$ indels between groups M1 and R (Table 1). Here, 0.01135 is the neutral substitution rate of 3*n* indels per site per unit of d_S, 759 is the length of exon 1 (alignment gaps excluded for conservative statistical tests), and 0.2648 is the average d_S in *Catsper1* exon 1 between groups M1 and R (Fig. 2.3). The observed number of indels between M1 and R is 11.84, which is significantly greater than the neutral expectation (P<0.001,

Poisson test; Table 1). Similarly, we computed the expected numbers of indels for other between-group comparisons (Table 1). The observed numbers of indels in exon 1 of rodent *Catsper1* are 5-18 times the neutral expectations (Table 1). Except for one comparison that yields marginal significance, all other comparisons show highly significant differences between the observed and expected numbers of indels (Table 1).

The presence of increased numbers of indels in exon 1 of *Catsper1*, when compared to the genomic average, could be explained by positive selection promoting the fixation of indel mutations and/or by a higher indel mutation rate at the *Catsper1* locus. To discriminate between these two possibilities, we assessed the indel mutation rate in the *Catsper1* locus by examining the intron sequences. Since the majority of sites in introns are more or less neutral, the indel substitution rate in *Catsper1* introns reflects the local indel mutation rate at the locus. We found a total of 106 indels in the 6739 nucleotide long alignment of the 11 introns of *Catsper1* between mouse and rat. Applying the genomic neutral indel substitution rate to the *Catsper1* introns, we would expect to see 86.2 indels between mouse and rat, which is significantly different from the observed number (P=0.02, Poisson test). The indel mutation rate at *Catsper1* is 23% greater than the genomic average. However, even when this elevated mutation rate is considered, we still find the observed number of indels in *Catsper1* exon 1 to be 4–15 times the neutral expectation at the locus. Except for the M1-M2 comparison, all other group comparisons are statistically significant (Table 1). These results strongly suggest that 3n indels have been favored by positive selection in rodent *Catsper1* exon 1. We further tested whether

indels of a particular size are preferentially fixed in exon 1 of *Catsper1* by comparing the *Catsper1* data to a genomic indel size distribution (Makova, Yang, and Chiaromonte 2004). No significant difference was detected (χ^2=0.97, d.f.=2, *P*>0.1), although this could be due to a relatively small sample size.

Comparison of intraspecific polymorphism data and interspecific divergence data is often used to test for natural selection. In particular, one can compare the ratio of the number of nonsynonymous differences to that of synonymous differences within species and between species (Mcdonald and Kreitman 1991). Here we modify this test by replacing nonsynonymous differences with indel differences, where each indel is counted as a single event regardless of the intel size. From the intraspecific data of *Mus musculus* and *M. macedonicus*, we found 2 indel and 1 synonymous polymorphisms. The number of fixed differences between the two species is 2 and 4 for indels and synonymous changes, respectively. The interspecific indel/synonymous ratio of 2:4 is not significantly different from the intraspecific ratio of 2:1 (Fisher's exact test, *P*>0.2). Therefore, this modified McDonald and Kreitman test does not reject the hypothesis that the selective intensity on indels is identical between the interspecific and intraspecific data. The above test has a low power here because both intraspecific and interspecific differences are small in our data. When the entire tree of Fig. 2.3 is considered, the indel/synonymous ratio of 22:61 is found. However, this ratio is again not significantly different from the intraspecific indel/synonymous ratio of 2:1 found in *M. musculus* and *M. macedonicus* (Fisher's exact test, *P*>0.1). Nevertheless, the

ratio appears higher for the intraspecific data than for the interspecifc data at face value.

The sequences of rodent *Catsper1* exon 1 show many amino acid replacements in addition to indels. Over 50% of pair-wise comparisons have more nonsynonymous substitutions per nonsynonymous site (d_N) than synonymous substitutions per synonymous site (d_S) (Fig. 2.4), with an average d_N/d_S ratio of all comparisons being 1.18. The standard McDonald and Kreitman test, comparing nonsynonymous/synonymous ratios within species (3/1 for *Mus musculus* and *M. macedonicus*) and between species (229/61 for the entire tree), showed no significant differences ($P>0.5$), suggesting that the selective pressure on nonsynonymous changes is indistinguishable between the intra- and inter-specific levels. The average d_N/d_S ratio of approximately 1 in the pairwise comparisons may be a result of several factors. One possibility is that several sites in exon 1 of *Catsper1* are under positive selection while others are under purifying selection. Alternatively, the region may be experiencing a relaxation of functional constraint. It is unlikely that exon 1 of *Catsper1* would evolve completely neutrally because deleting the gene in mice leads to infertility (Ren et al. 2001). Also, despite many indels present in the first exon of *Catsper1*, the open reading frame remains undisrupted. It is therefore more plausible that several sites are evolving under positive selection for amino acid replacements, while other sites are under functional constraints, rendering the overall d_N/d_S ratio approximately 1. To test this hypothesis, we performed a likelihood analysis using PAML (Yang 1997). Specifically, we compared models M7 and M8 (Yang et al. 2000a). M7 is a null model, which assumes that the d_N/d_S ratio across

codons follows a beta distribution in the range of 0 to 1. M8 is identical to M7 except for an additional class of codons with a free d_N/d_S ratio. Our likelihood ratio test showed that M8 fits our data significantly better than M7 ($\chi^2 = 46.77$, d.f. =2, $P<$ 0.001), with an additional class of sites having $d_N/d_S = 5.78$. It has been suggested that a better way to test for positive selection is to compare the alternative hypothesis of M8 with the null hypothesis of M8a (Swanson, Nielsen, and Yang 2003; Wong et al. 2004). M8a is a special case of M8, where instead of estimating one additional class of sites with any d_N/d_S ratio, M8a fixes this d_N/d_S ratio to 1. The use of M8a seems to reduce false positives (Suzuki and Nei 2001; Zhang 2004) caused by codon sites under weak or no selection because these get absorbed in the neutral class of sites with d_N/d_S =1 (Wong et al. 2004). In our analysis, M8 fit the data significantly better than M8a ($\chi^2 = 45.03$, d.f. = 1, $P<0.001$), suggesting that positive selection for nonsynonymous substitutions may indeed be acting on the first exon of rodent *Catsper1*, a result consistent with that found in primates.

Discussion

Our analysis of exon 1 of *Catsper1* from 22 individuals of 9 rodent species revealed strong evidence for positive selection that promotes fixation of indel mutations. These substitutions resulted in different lengths in the N-terminus of Catsper1 among different species. Within species, the level of indel polymorphism also appears higher than expected from interspecific comparisons, although our data are too limited to provide a statistically significant result. It is interesting to note that some genes that have been shown to evolve rapidly between species also show

exceptionally high levels of intraspecific polymorphism, possibly due to balancing selection (e.g., Drosophila accessory gland protein Acp26Aa; (Tsaur, Ting, and Wu 2001).

It is known that insertion and deletion mutations in DNA sequences are often due to slippage during DNA replication (Hancook 1999a). This type of mutation is particularly common in regions containing simple nucleotide repeats. With the exception of a short C_5A repeat in the 3' end, *Catsper1* exon 1 does not have apparent sequence repeats. However, it should be noted that *Catsper1* exon 1 has a strong amino acid compositional bias, with histidine and serine having a total frequency of 32%. With the exception of the first two alignment gaps that are within C_5A repeats, the majority of the indels observed are not located near repeats, suggesting that the unusually high abundance of indels in *Catsper1* exon 1 is not caused by repeat expansion/contraction.

What selective agents are behind the extraordinarily rapid pace of indel substitutions in the sperm ion channel encoded by *Catsper1*? Genes belonging to the *Catsper* family encode putative voltage-gated Ca^{2+} (Ca_V) ion channels. A recent study showed that Catsper1 is necessary for depolarization-evoked Ca^{2+} entry and for hyperactive sperm movement (Carlson et al. 2003). The onset of hyperactive movement occurs near the site of fertilization and is crucial for penetrating the egg's zona pellucida. Although direct evidence showing that Catsper1 forms a voltage-gated ion channel is still lacking, the requirement of Catsper1 for depolarization-evoked entry of Ca^{2+} strongly supports such a hypothesis (Carlson et al. 2003). Several other members of the Catsper family have been identified and all share a

common expression in sperm (Quill et al. 2001; Lobley et al. 2003). It was thus conjectured that Catsper1, 2, 3 and 4 form a functional hetero-tetramer in sperm; however, empirical evidence for this hypothesis is lacking (Lobley et al. 2003). The putative ion channel formed by Catsper1 resembles K_V channels more than Ca_V channels in structure, as Catsper1 and K_V channels are formed by four peptides rather than a single long peptide with four repeats. This structural similarity to Kv ion channels prompted us to hypothesize a regulatory mechanism for Catsper1 similar to the "ball-and-chain" model (Bezanilla and Armstrong 1977a) and in which strong positive selection could occur for indel substitutions in the N-terminus of Catsper1 (Podlaha and Zhang 2003). Specifically, it was found that the action of the Kv channel Shaker in *Drosophila* follows the “ball-and-chain” model, in which the length of the N-terminus directly affects the rate of the channel’s inactivation (Hoshi, Zagotta, and Aldrich 1990a). If a similar regulatory mechanism indeed operates in Catsper1, different N-terminus lengths could have a direct functional effect on the channel’s performance and consequently on sperm motility (Podlaha and Zhang 2003).

The physiological function and molecular evolutionary pattern of *Catsper1* suggests that it is a candidate for direct involvement in sperm competition. Further understanding of the evolutionary forces acting on this gene requires that questions about Catsper1 expression and fitness effects be addressed. Since *Catsper1* is expressed in sperm, could it possibly be under haploid selection? For a gene to be haploid selected, it is necessary to show that each sperm contains the product of only one allele of the gene and that the two alleles confer different fitnesses (Joseph and

Kirkpatrick 2004). Rare examples of haploid selection have been documented in sperm adhesion molecule 1 *Spam1* (Zheng, Deng, and Martin-DeLeon 2001), segregation distortion responder *Rsp* (Kusano et al. 2003), and the *t*-locus responder *Tcr* (Herrmann et al. 1999). Although no direct demonstration of haploid selection in *Catsper1* exists, two lines of evidence lead us to believe that this is the case. First, *Catsper1* (as well as *Catsper2*, *Catsper3*, *Catsper4*) is solely expressed in testis and no other reproductive tissue (Genomics Institute of the Novaris Research Foundation SymAtlas; symatlas.gnf.org/SymAtlas/). Second, the timing of *Catsper1* expression is confined to the postmeiotic stage of spermatogenesis (Schultz, Hamra, and Garbers 2003), strongly suggesting that only one allele is expressed in each sperm. Haploid selection can result in meiotic drive, which refers to the phenomenon of nonrandom segregation in heterozygotes (Hartl and Clark 1997). Such haploid selection and meiotic drive could generate an "arms-race" between alleles and result in high levels of intraspecific polymorphism and interspecific divergence. Haploid selection and meiotic drive can be tested by genotyping the offspring of heterozygous males and examining the competitive advantage of a particular length variant of Catsper1 over another variant. Such experiments are feasible in mice, where obtaining individuals of a particular genotype and acquiring data on reproductive success is faster and easier than in primates. *Catsper1* is a promising candidate gene for sperm competition as sperm motility has been shown to be one of its decisive factors (Birkhead et al. 1999). The present study, as well as our previous analysis in primates, reveals strong positive selection on the length of Catsper1. Furthermore, it provides a necessary foundation for future experimental

investigations of Catsper1's function in sperm physiology and role in sperm competition using rodent models.

Acknowledgments

Resluts of this work have been published in the journal of Molecular Biology and Evolution (Podlaha et al. 2005). We thank A. Brege for technical assistance, M. Potter (National Cancer Institute Contract N01-CB-71085) and E. Eicher (the Jackson Laboratory) for providing frozen tissue/DNA from various stocks and strains of wild mice and R. D. Sage for providing frozen tissue/DNA from wild mice collected in the field. This work was supported in part by the NIH grant GM67030 to J.Z.

Table 2.1 Test of positive selection for indel substitutions in rodent *Catsper1* exon 1.

Comparison between groups	d_S [2]	Expected number of indels in *Catsper1* exon 1[3]		Observed number of indels in *Catsper1* exon 1	Probability[4]	
		Genomic	*Catsper1*		Genomic	*Catsper1*
M1-M2	0.0383	0.33	0.41	2.34	4.4×10^{-2}	6.3×10^{-2}
M1-M3	0.0865	0.75	0.92	12.00	5.1×10^{-10}	4.3×10^{-9}
M1-R	0.2648	2.29	2.81	11.84	2.8×10^{-5}	1.7×10^{-4}
M2-M3	0.0728	0.63	0.77	11.66	8.6×10^{-11}	7.3×10^{-10}
M2-R	0.2511	2.17	2.67	11.50	1.7×10^{-5}	1.1×10^{-4}
M3-R	0.2609	2.25	2.77	11.50	2.5×10^{-5}	1.5×10^{-4}

[1] See Fig. 3 for definitions of the groups.

[2] Average d_S distance between groups.

[3] Expected numbers of indels were calculated by multiplying the neutral substitution rate of $3n$ indels per d_S, average d_S between groups, and the length of the sequence concerned. To make the statistical test more conservative, we used the sequence length of 759 nucleotides, after removing all alignment gaps. Two neutral rates, one estimated from the mouse-rat genomic comparison and the other from introns of Catsper1, were used.

[4] The probability of the observation given the expectation was computed under the assumption that the number of indels follows a Poisson distribution.

FIG. 2.1. Alignment of the full-length Catsper1 sequences of the mouse (*Mus musculus*) and rat (*Rattus norvegicus*). Dots represent amino acids identical to the first sequence and dashes represent alignment gaps. The region encoded by exon 1 is boxed. The six transmembrane domains (S1-S6) and the pore-forming region (P) are shaded.

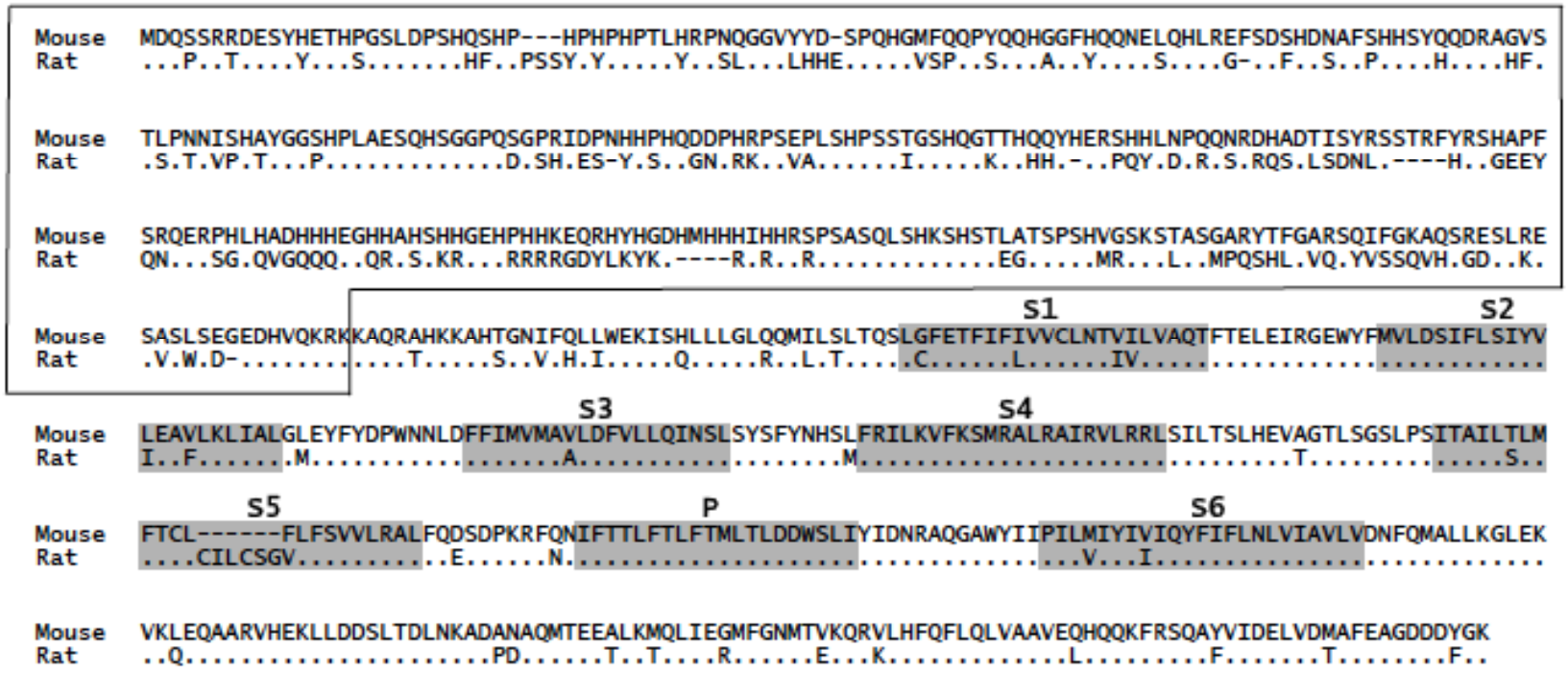
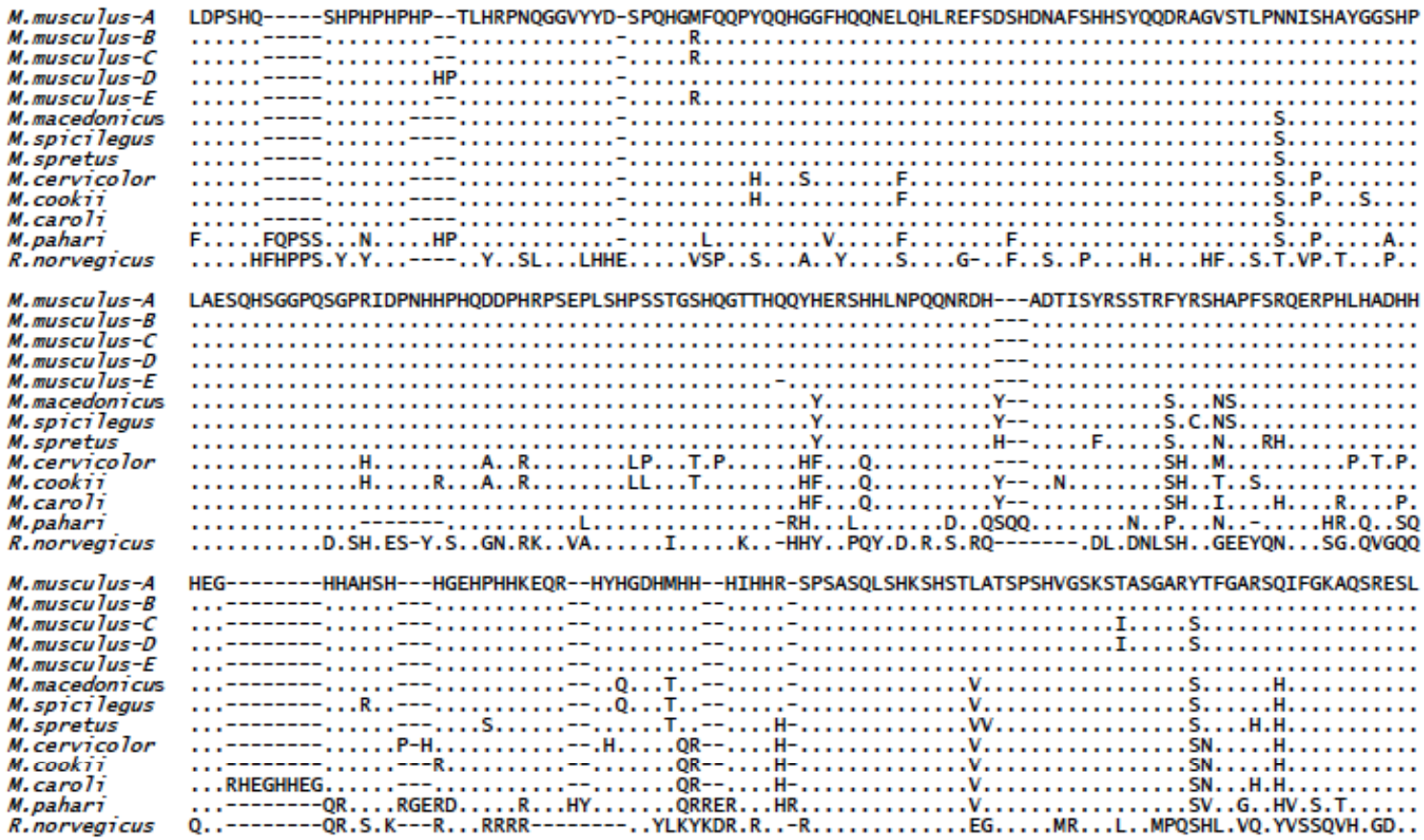

Mouse MDQSSRRDESYHETHPGSLDPSHQSHP---HPHPHPTLHRPNQGGVYYD-SPQHGMFQQPYQQHGGFHQQNELQHLREFSDSHDNAFSHHSYQQDRAGVS
Rat ...P..T....Y...S.......HF..PSSY.Y.....Y..SL...LHHE.....VSP..S...A..Y....S....G-..F..S..P....H....HF.
Mouse TLPNNISHAYGGSHPLAESQHSGGPQSGPRIDPNHHPHQDDPHRPSEPLSHPSSTGSHQGTTHQQYHERSHHLNPQQNRDHADTISYRSSTRFYRSHAPF
Rat .S.T.VP.T...P.............D.SH.ES-Y.S..GN.RK..VA......I.....K..HH.-..PQY.D.R.S.RQS.LSDNL.----H..GEEY
Mouse SRQERPHLHADHHHEGHHAHSHHGEHPHHKEQRHYHGDHMHHHIHHRSPSASQLSHKSHSTLATSPSHVGSKSTASGARYTFGARSQIFGKAQSRESLRE
Rat QN...SG.QVGQQQ..QR.S.KR...RRRRGDYLKYK.----R.R..R.............EG.....MR...L..MPQSHL.VQ.YVSSQVH.GD..K.
S1
S2
Mouse SASLSEGEDHVQKRKKAQRAHKKAHTGNIFQLLWEKISHLLLGLQQMILSLTQSLGFETFIFIVVCLNTVILVAQTFTELEIRGEWYFMVLDSIFLSIYV
Rat .V.W.D-.....T.....S..V.H.I.....Q.....R..L.T....C......L......IV.....
S3
S4
Mouse LEAVLKLIALGLEYFYDPWNNLDFFIMVMAVLDFVLLQINSLSYSFYNHSLFRILKVFKSMRALRAIRVLRRLSILTSLHEVAGTLSGSLPSITAILTLM
Rat I..F.....M.....A.....M.....T.....S..
S5
P
S6
Mouse FTCL------FLFSVVLRALFQDSDPKRFQNIFTTLFTLFTMLTLDDWSLIYIDNRAQGAWYIIPILMIYIVIQYFIFLNLVIAVLVDNFQMALLKGLEK
Rat ....CILCSGV.....E.....N.....V...I.....
Mouse VKLEQAARVHEKLLDDSLTDLNKADANAQMTEEALKMQLIEGMFGNMTVKQRVLHFQFLQLVAAVEQHQQKFRSQAYVIDELVDMAFEAGDDDYGK
Rat ..Q.....PD.....T..T....R.....E...K.....L.....F.....T.....F..

FIG. 2.2. Alignment of translated rodent *Catsper 1* exon 1 sequences obtained in this study. Only non-redundant protein sequences are shown. Dots represent amino acids identical to the first sequence and dashes represent alignment gaps. *M. musculus-A* sequence was downloaded from GenBank (mouse strain C57BL/6J, GenBank accession no. NM_139301) and represents a haplotype shared by all six *M. musculus domesticus* mice and one *M. musculus castaneus* individual. *M. musculus-B* represents a haplotype carried by two *M. musculus musculus* individuals, one from Czech Republic and one from Slovakia. *M. musculus-C* represents a haplotype of one *M. musculus musculus* from Slovakia. *M. musculus-D* represents a haplotype of one *M. musculus musculus* from Finland. *M. musculus-E* represents a unique haplotype found in *M. musculus molossinus* from Japan. Sequences from two *M. macedonicus* individuals were identical and are represented by a single *M. macedonicus* haplotype.


M.musculus-A LDPSHQ-----SHPHPHPHP--TLHRPNQGGVYYD-SPQHGMFQQPYQQHGGFHQQNELQHLREFSDSHDNAFSHHSYQQDRAGVSTLPNNISHAYGGSHP
M.musculus-B ......-----.........--.............-.....R.....
M.musculus-C ......-----.........--.............-.....R.....
M.musculus-D ......-----.........HP.............-.....
M.musculus-E ......-----.........--.............-.....R.....
M.macedonicus ......-----.......----.............-.....S.....
M.spicilegus ......-----.......----.............-.....S.....
M.spretus ......-----.........--.............-.....S.....
M.cervicolor ......-----.......----.............-.........H...S.......F.....S..P.....
M.cookii ......-----.......----.............-.........H.........F.....S..P...S....
M.caroli ......-----.......----.............-.....S.....
M.pahari F.....FQPSS...N.....HP.............-......L.........V.....F........F.....S..P.....A..
R.norvegicus .....HFHPPS.Y.Y...----..Y..SL...LHHE.....VSP..S...A..Y....S....G-..F..S..P....H....HF..S.T.VP.T...P..
M.musculus-A LAESQHSGGPQSGPRIDPNHHPHQDDPHRPSEPLSHPSSTGSHQGTTHQQYHERSHHLNPQQNRDH---ADTISYRSSTRFYRSHAPFSRQERPHLHADHH
M.musculus-B .....---.....
M.musculus-C .....---.....
M.musculus-D .....---.....
M.musculus-E .....-.....---.....
M.macedonicus .....Y.....Y--.....S...NS.....
M.spicilegus .....Y.....Y--.....S.C.NS.....
M.spretus .....Y.....H--.....F.....S...N...RH.....
M.cervicolor .....H.....A..R.....LP...T.P.....HF...Q.....---.....SH..M.....P.T.P.
M.cookii .....H.....R...A..R.....LL...T.....HF...Q.....Y--..N.....SH..T..S.....
M.caroli .....HF...Q.....Y--.....SH..I....H....R....P.
M.pahari .....-------.....L.....-RH...L.....D..QSQQ.....N..P...N..-.....HR.Q..SQ
R.norvegicus .....D.SH.ES-Y.S..GN.RK..VA.....I.....K..-HHY..PQY.D.R.S.RQ-------.DL.DNLSH..GEEYQN...SG.QVGQQ
M.musculus-A HEG--------HHAHSH---HGEHPHHKEQR--HYHGDHMHH--HIHHR-SPSASQLSHKSHSTLATSPSHVGSKSTASGARYTFGARSQIFGKAQSRESL
M.musculus-B ...--------.....---.....--.....--.....-.....
M.musculus-C ...--------.....---.....--.....--.....-.....I.....S.....
M.musculus-D ...--------.....---.....--.....--.....-.....I.....S.....
M.musculus-E ...--------.....---.....--.....--.....-.....
M.macedonicus ...--------.....---.....--..Q...T..--.....-.....V.....S.....H.....
M.spicilegus ...--------...R..---.....--..Q...T..--.....-.....V.....S.....H.....
M.spretus ...--------.....---....S.....--.....T..--....H-.....VV.....S....H.H.....
M.cervicolor ...--------.....P-H.....--.H.....QR--....H-.....V.....SN.....H.....
M.cookii ...--------.....---R.....--.....QR--....H-.....V.....SN.....H.....
M.caroli ...RHEGHHEG.....---.....--.....QR--....H-.....V.....SN...H.H.....
M.pahari ...--------QR....RGERD.....R...HY.....QRRER...HR.....V.....SV..G..HV.S.T.....
R.norvegicus Q..--------QR.S.K---R...RRRR--------..YLKYKDR.R..-R.....EG.....MR...L..MPQSHL.VQ.YVSSQVH.GD..

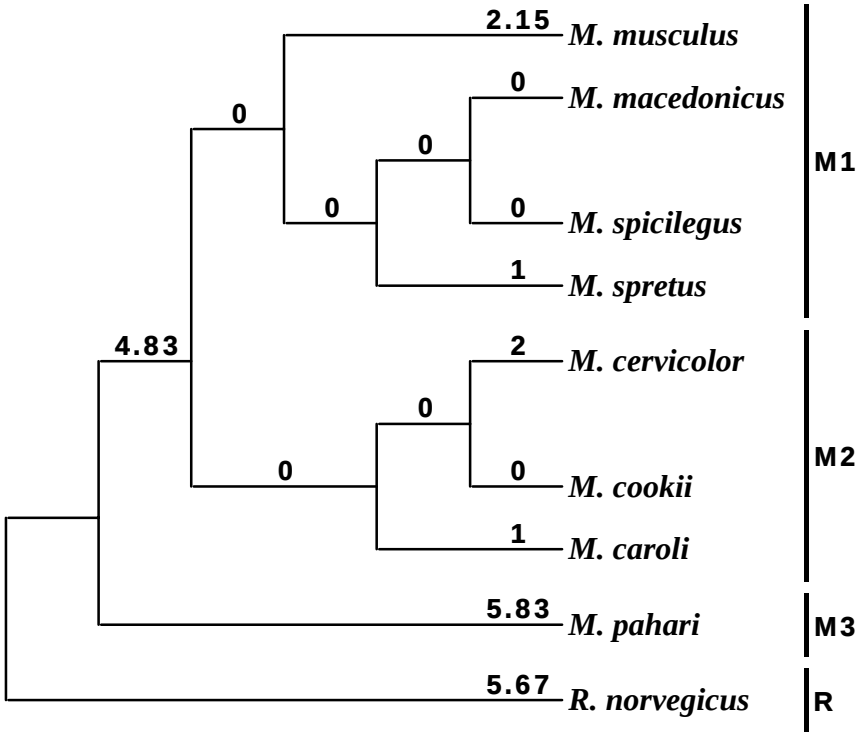


FIG. 2.3. Indel substitutions in the evolution of rodent *Catsper1* exon 1. The phylogeny of the nine rodent species included in our study was assumed to follow (Tucker, Sandstedt, and Lundrigan 2005). Numbers above branches represent parsimony-inferred numbers of indel substitution events. Capital letters on the right depict groupings of rodent species.

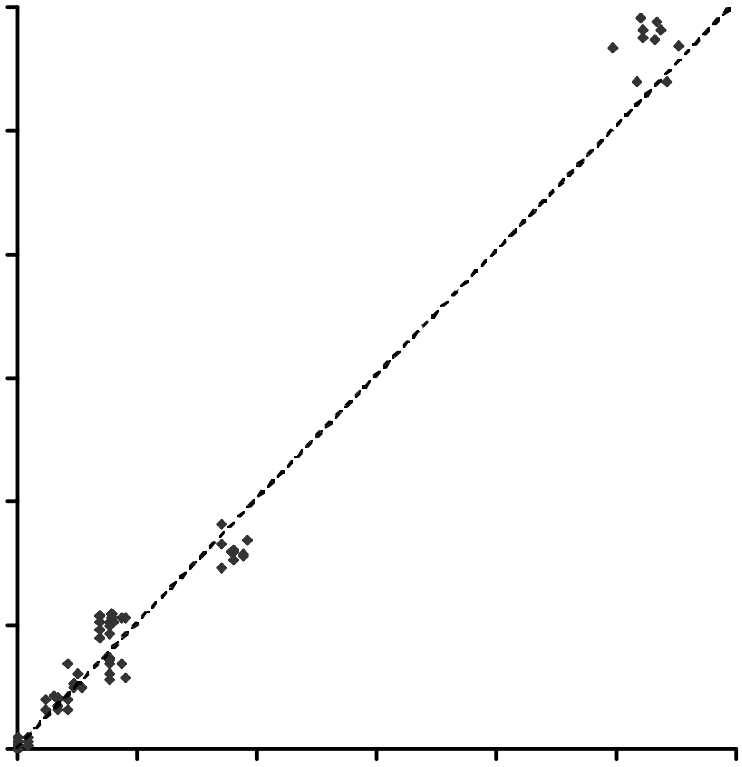


FIG. 2.4. Number of synonymous (d_S) and nonsynonymous (d_N) substitutions per site between rodent *Catsper1* exon 1 sequences. The 14 non-redundant sequences were used in this pairwise comparison. Dotted line represents $d_N/d_S = 1$.

CHAPTER III

ACCELERATED EVOLUTION AND LOSS OF A DOMAIN OF THE SPERM-EGG BINDING PROTEIN SED1 IN ANCESTRAL PRIMATES

Introduction

Genes involved in sperm-egg binding often show a rapid pace of evolution driven by positive diversifying selection (reviewed in Swanson and Vacquier 2002a). SED1 is a recently identified sperm protein from the mouse *Mus musculus* (Ensslin and Shur 2003). It is expressed in spermatogenic cells and secreted by the initial segment of the caput epididymis, resulting in its localization on the sperm plasma membrane overlying the acrosome. SED1 binds to the zona pellucida, specifically the glycoproteins ZP2 and ZP3, of unfertilized oocytes, but not to the zona of fertilized eggs. The fertility of SED1-null male mice is about one third that of wildtype mice, and SED1-null sperm cannot bind to the egg coat in vitro (Ensslin and Shur 2003). Here we report the finding of a structural change and positive directional selection of SED1 in an ancestral primate lineage and suggest that the SED1 function may have changed in primates.

Mouse SED1 is a short splice form of the well known milk fat globule-EGF factor 8 (MFGE8) gene. SED1 contains two Notch-like EGF domains responsible for protein-protein interaction and two discoidin/F5/8 type C domains involved in protein-cell membrane interaction (Fig. 3.1). To examine the evolutionary pattern of

SED1, we downloaded eight mammalian SED1 sequences from GenBank, including those from the human, mouse, rat, pig, cow, horse, dog, and opossum. Sequence alignment revealed conserved domain organization of SED1 across all examined taxa with the exception of the human, which lacks the N-terminal EGF domain (Fig. 3.1). One structural-functional model of SED1 is that it forms a homodimer through the interaction of EGF domains (Ensslin and Shur 2003) and it is believed that at least two EGF domains are required for successful protein-protein binding (Lawrence et al. 2000; Balzar et al. 2001). If correct, loss of one of these domains would reduce or prohibit the dimerization of SED1, suggesting a potential functional change of SED1 in primates. To narrow down the time when the EGF-like domain was lost, we sequenced five additional species that represent major lineages of higher primates (chimpanzee, gorilla, orangutan, rhesus monkey, and spider monkey). The gene tree of the SED1 sequences from six primates and five representatives of non-primate mammals showed a topology consistent with the well established species tree (Murphy, Pevzner, and O'Brien 2004; Goodman, Grossman, and Wildman 2005), indicating that the SED1 gene sequences under investigation are orthologous. We did not use the dog and opossum sequences because they were from draft genome sequences that may contain errors. The SED1 of all six primates have the same domain structure (Fig. 3.1), indicating that the first EGF domain was lost after the separation of primates from rodents, but before the divergence of platyrrhines (New World monkeys) and catarrhines (Old World monkeys, apes, and humans). For convenience, we will refer to this period of time as an ancestral lineage of primates.

Materials and Methods

The SED1 gene was amplified via polymerase chain reaction (PCR) from the genomic DNAs of the chimpanzee *Pan troglodytes*, gorilla *Gorilla gorilla*, orangutan *Pongo pygmaeus*, rhesus monkey *Macaca mulatta*, and spider monkey *Ateles geoffroyi*, purified, and sequenced by automatic DNA sequencing. Attempts to amplify prosimian SED1 genes were unsuccessful. The SED1 sequences from the human *Homo sapiens*, cow *Bos taurus*, pig *Sus scrofa*, horse *Equus caballus*, mouse *Mus musculus*, and rat *Rattus norvegicus* were obtained from GenBank. The DNA sequences were aligned following a protein alignment by ClustalX (Thompson et al. 1997b). Several methods were used to compare rates of synonymous and nonsynonymous substitutions in SED1. First, we used the modified Nei-Gojobori method (Zhang, Rosenberg, and Nei 1998a) to estimate pairwise synonymous and nonsynonymous distances between extant sequences and the least-squares method to estimate synonymous and nonsynonymous branch lengths of a given tree (Zhang, Rosenberg, and Nei 1998a). Second, we used a Bayesian method (Yang, Kumar, and Nei 1995a) to infer ancestral SED1 gene sequences at all interior nodes of the SED1 gene tree and compare the numbers of synonymous and nonsynonymous substitutions for individual tree branches using Fisher's exact test (Zhang, Kumar, and Nei 1997). Finally, we used an improved branch-site likelihood method (Zhang, Nielsen, and Yang 2005) implemented in PAML (Yang 1997) to test for positive selection. Rates of conservative and radical nonsynonymous substitutions were estimated using the method of (Zhang 2000a). Mega3.1 (Kumar, Tamura, and Nei

2004) was used for phylogenetic analysis. The domain structure of SED1 was also examined for the preliminary sequences identified from the dog and opossum draft genome sequences.

Results

An indicator of a protein functional shift is the occurrence of positive directional selection, which can be tested by comparing the rates of synonymous and nonsynonymous nucleotide substitutions for the tree branch in which the functional shift is suspected (Nei and Kumar 2000). We used three different methods to conduct such a test for the ancestral primate branch in which the N-terminal EGF domain of SED1 was lost. First, we estimated branch lengths in the SED1 gene tree, in terms of the number of synonymous substitutions per synonymous site (b_S) and the number of nonsynonymous substitutions per nonsynonymous site (b_N), respectively, using the least-squares method (Zhang, Rosenberg, and Nei 1998) (Fig. 3.2). The ancestral primate branch (bolded in Fig. 3.2) exhibits a distinct pattern of $b_N > b_S$. Using a two-tail *z*-test, we found b_N (0.183±0.020) to be significantly greater than b_S (0.059±0.020) ($P<10^{-4}$). This large-sample test (Zhang, Kumar, and Nei 1997) is appropriate here because the inferred numbers of synonymous and nonsynonymous substitutions are both greater than 10 for the concerned branch (see below). Second, we inferred the ancestral gene sequences at all interior nodes of the SED1 tree by a Bayesian method (Yang, Kumar, and Nei 1995a). There were n = 120.5 nonsynonymous and s = 34.5 synonymous differences between the two nodes that are at the ends of the ancestral primate branch. The potential numbers of

nonsynonymous and synonymous sites in SED1 are N = 688.1 and S = 304.9, respectively. Thus, the n/s ratio is 3.49, significantly greater than its neutral expectation (N/S=2.26) (P<0.01, Fisher's exact test). Finally, we used a likelihood method known as the branch-site test of positive selection or test 2 (Zhang, Nielsen, and Yang 2005), which compares the likelihood of a null model that does not invoke positive selection with that of an alternative model that invokes positive selection in a predetermined tree branch (i.e., the bolded branch in Fig. 3.2). For SED1, the null neutral model is rejected in favor of the alternative model (χ^2 = 7.3, df = 1, P<0.007), with p_2=20% of sites being estimated to be under positive selection in the ancestral primate branch (nonsynonymous/synonymous rate ratio ω_2=5.42 for these sites). Thus, the distance-, parsimony-, and likelihood-based methods all show a significantly higher substitution rate at nonsynonymous sites than synonymous sites in the ancestral primate branch of the SED1 tree, strongly suggesting the operation of positive selection.

To examine which domain of the ancestral primate SED1 was under positive selection, we examined each domain separately by the branch-site likelihood method. The null neutral hypothesis is strongly rejected for the retained EGF domain (χ^2=16.8, df=1, P<0.001), but is only marginally rejected or not rejected for the two discoidin/F5/8 type C domains (abbreviated C1 and C2; C1, χ^2=4.24, df=1, P=0.039; C2, χ^2=3.56, df=1, P=0.059). The ω_2 values for the EGF, C1, and C2 domains are estimated to be 999 (i.e., exceeding the largest ω_2 examined by the program) (p_2=0.38), 5.35 (p_2=0.15), and 3.4 (p_2=0.03), respectively. Interestingly, the identified positively selected sites (with posterior probabilities >0.95) in the C

domains are enriched in spike regions, which are involved in membrane binding (Shur, Ensslin, and Rodeheffer 2004). When substitution rates are calculated from ancestral sequences, only the EGF domain had significantly higher n/s over N/S (P=0.0023, Fisher's exact test). We also investigated whether any particular type of nonsynonymous substitutions was preferentially fixed in the ancestral primate branch by comparing the rates of conservative and radical nonsynonymous substitutions (Zhang 2000a). When amino acid polarity is considered, the number of radical nonsynonymous substitutions per radical nonsynonymous site (d_R) is significantly greater than the number of conservative nonsynonymous substitutions per conservative nonsynonymous site (d_C) in the EGF domain (d_R/d_C=4.5, P<0.001, Fisher's exact test), but not in the other domains. More specifically, 8 of the 10 polarity-altering amino acid replacements in EGF were from polar to nonpolar residues, suggesting that nonpolar amino acids may have been selectively favored in the ancestral primate branch.

Discussion

Our evolutionary analyses of the mammalian SED1 sequences provide strong evidence that SED1 was subject to positive selection in ancestral primates. The selection appears to target the EGF domain and favors changes of amino acid polarity. A difficulty in deciphering the selective agent on SED1 is that it is a short splicing variant of the *MFGE8* gene. Ensslin and Shur (2003) named the short variant SED1 (secreted protein containing N-terminal Notch-like type II EGF repeats and c-terminal discoidin/F5/8 C domains) to distinguish it from the previously

known long splicing variant (MFGE8) that is found in milk. MFGE8 participates in many different physiological functions including vascular endothelial growth factor-dependent neovascularization (Silvestre et al. 2005), inhibition of blood coagulating enzymes (Shi and Gilbert 2003), inhibition of rotaviral infections (Kvistgaard et al. 2004), and somatic cell-to-cell interaction (Ishii et al. 2005). In mice, SED1 has a broader expression than MFGE8. SED1 expression was detected through RT-PCR in liver, intestine, kidney, skin, stomach, heart, testis, brain, spleen, mammary glands, and lung tissues, whereas MFGE8 was detected primarily in the mammary glands and at much lower levels in skin, stomach, testis, spleen, and lung (Watanabe et al. 2005). Primate SED1, however, is less well studied. Information of expressed sequence tags show that human SED1/MFGE8 is expressed in many tissues, including the testis (http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Hs.3745).

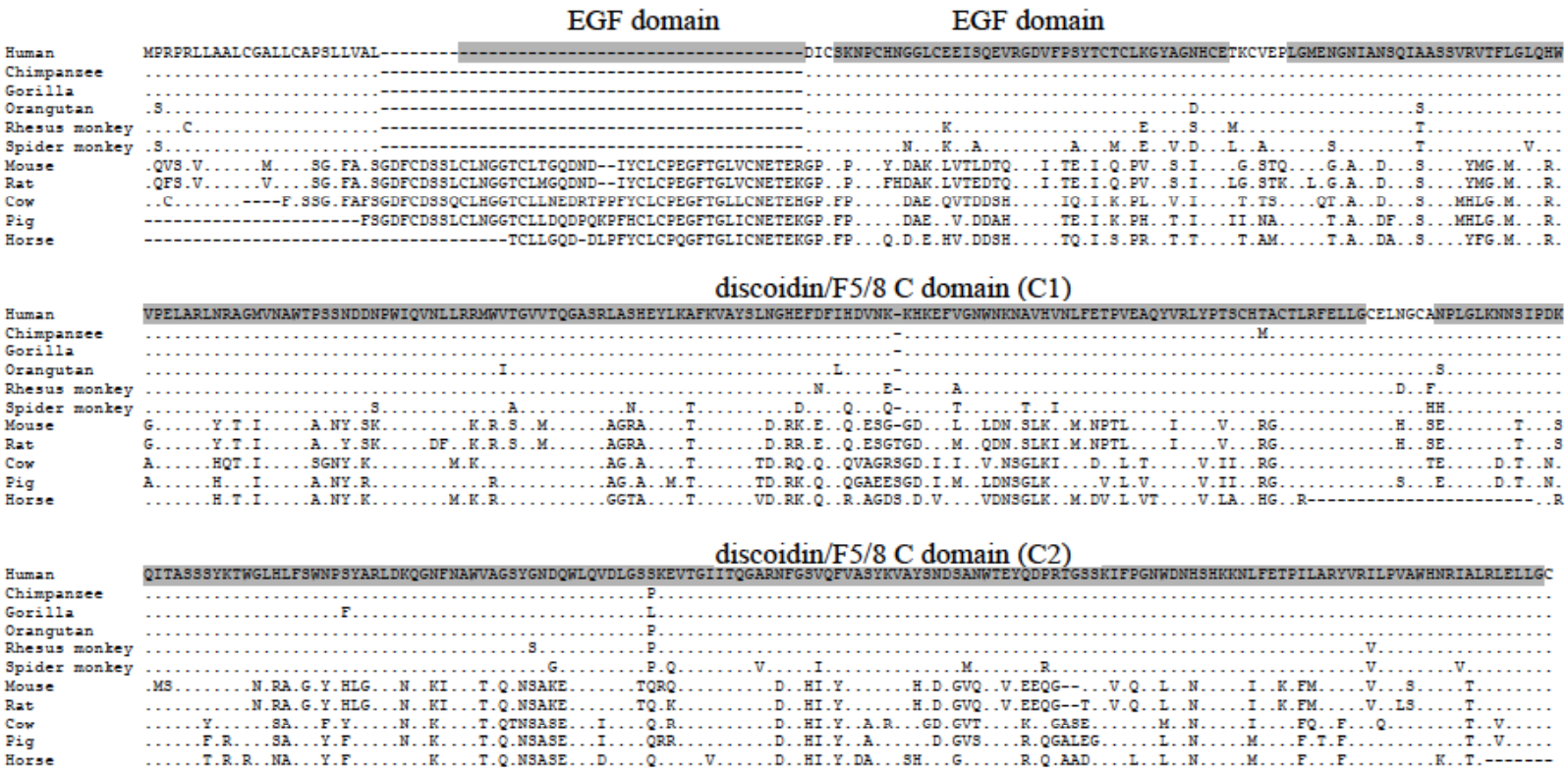
Many sperm-egg binding proteins show rapid evolution driven by diversifying selection occurring in multiple evolutionary lineages (reviewed in Swanson and Vacquier 2002). Such diversifying selection may alter the binding efficiency or specificity, but is unlikely to change the basic function of the protein. In the case of mammalian SED1, however, positive selection was identified in a single lineage (ancestral primates), with additional characteristics of directional selection and functional shifts such as the loss of an EGF domain and increase of nonpolar residues in the other EGF domain. Although the exact selective agent on primate SED1 is unknown, in part due to the complication of multifunctionality and alternative splicing, the discrepancy in evolutionary pattern between SED1 and other

sperm-egg binding proteins is intriguing. A functional change of a sperm-egg binding protein during primate evolution, if proven, would have significant evolutionary, physiological, and medical implications. Our results call for a thorough functional assay of primate SED1.

Acknowledgments

Results of this work have been published in the journal of Molecular Biology and Evolution (Podlaha, Webb, and Zhang 2006). We thank Daniel Green for experimental assistance and three anonymous reviewers for valuable comments. This work was supported by research grants from the University of Michigan and the National Institutes of Health to J.Z.

FIG. 3.1. Protein sequence alignment of SED1 from 11 mammals. Dots represent identical amino acids to the first sequence and dashes represent alignment gaps. The pig and horse sequences are partial, having incomplete N- and C-termini. Protein domains are superimposed over the first sequence with the grey color.


EGF domain
EGF domain
discoidin/F5/8 C domain (C1)
discoidin/F5/8 C domain (C2)
Human
Chimpanzee
Gorilla
Orangutan
Rhesus monkey
Spider monkey
Mouse
Rat
Cow
Pig
Horse

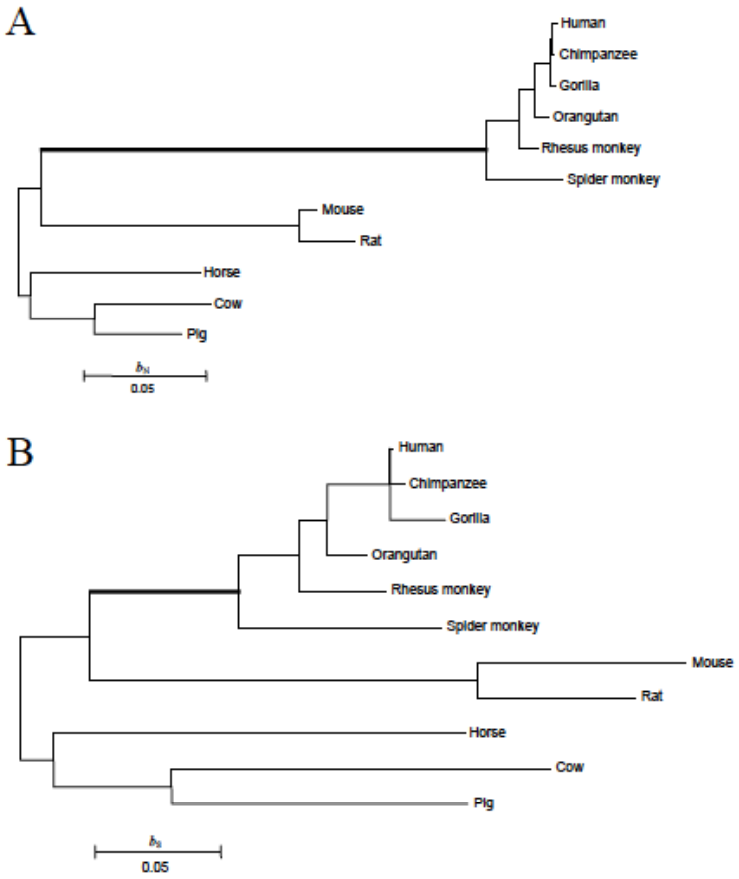


FIG. 3.2. The SED1 gene tree, with branch lengths measured by (A) the number of nonsynonymous substitutions per nonsynonymous site and (B) the number of synonymous substitutions per synonymous site. The modified Nei-Gojobori method was used to estimate the synonymous and nonsynonymous distances between pairs of extant species and the least-squares method was used to estimate the branch lengths.

CHAPTER IV

EVOLUTION OF THE MALE SEX-DETERMINATION GENE *SRY* AND SEX-REVERSAL IN AKODONT RODENTS

Introduction

Therian mammals have a heterogametic system of sex determination. Very simply put, the presence or the absence of the Y chromosome determines whether gonadal premordia develop into testis or ovaries, which in turn triggers the acquisition of secondary male or female characteristics (Harley, Clarkson, and Argentaro 2003). Although this system of sex determination is relatively well conserved across therian mammals, several exceptions exist (Fredga 1983). For example the wood lemming (*Myopus*) and a vole (*Microtus cabrerae*) both have XY females, which are thought to be a result of mutation on the X chromosome (Fredga et al. 1976; Burgos et al. 1988). Females in European species of the genus *Talpa* (moles) are hermaphroditic but the genetic mechanism causing this abnormality is still unknown (Sanchez et al. 1996).

Most of the functional information about mammalian sex determination has been obtained using the mouse model system. The progress leading to the discovery of the sex-determining region on the Y chromosome has been hampered by the seemingly complicated nature of the sex determination pathway itself and the genes involved. In the early 1990's, series of experiments have established that the *Sry*

gene (Sex determining region on the Y chromosome) is the primary molecular switch that triggers a cascade of events leading to testis development. Simple transgenetic experiments where *Sry* has been added to XX mouse led to the development of male phenotype and conversely the ablation of *Sry* function in XY mice failed to trigger male development (Gubbay et al. 1990; Lovell-Badge and Robertson 1990; Koopman et al. 1991). Since the discovery of *Sry* 17 years ago, however, little else about this gene has proven to be simple or straightforward.

Sry is a transcription factor encoded by a single exon on the short arm of the Y chromosome. The expression of *Sry* is tightly spatiotemporally controlled. In mice, *Sry* expression in the genital ridges of gonadal premordia reaches maximum 11.5 days past conception (dpc) and fades away before 12.5dpc. What induces *Sry* expression is still being speculated, as much of the evidence is only correlatory. Several autosomal genes expressed in the undifferentiated gonadal premordia, however, are particularly interesting candidates. Mutations in the Wilm's tumor-1 (*Wt-1*) are known to cause sex reversal in humans and two of the many *Wt-1* protein products (WT1-KTS and WT1+KTS) were reported to upregulate *Sry* via two different mechanisms (Hammes et al. 2001). WT1+KTS enhances *Sry* translation by stabilizing *Sry* mRNA (Hammes et al. 2001; Bor et al. 2006), whereas WT1-KTS, along with several other factors, bind *Sry* promoter region (Hossain and Saunders 2001; Pilon et al. 2003). Similarly to *Wt-1*, mutations in genes belonging to the insulin receptor tyrosine kinase family (*Igfl1r*, *Irr*, and *Ir*) as well as *Gata4*, *Fog2*, *Sp1*, and *Sf1* dramatically reduce expression of *Sry* and *Sox9* (Barbara et al. 2001; Tevosian et al. 2002; Nef et al. 2003; Assumpcao et al. 2005), a gene further

downstream in the sex determination pathway. In all of these cases, it is not known whether the reduction is due to lower levels of *Sry* expression or to fewer cells expressing *Sry*. It also remains controversial whether *Sry* targets multiple or just a single gene that carries out all functions necessary for male testis development. Arguably the best candidate target gene of *Sry* is *Sox9*. *Sox9* transcription is upregulated 4 hours immediately after the onset of *Sry* (Wilhelm et al. 2005), both displaying overlapping pattern of expression in the developing pre-Sertolli cells. Whether *Sry* induces *Sox9* directly or indirectly still needs to be determined. Interestingly, transgenic mice overexpressing SOX9 in the genital ridges developed testes even in the absence of *Sry* (Bishop et al. 2000; Vidal et al. 2001; Qin and Bishop 2005). Furthermore, mutations in the SOX9 result in XY sex reversal, implying that *Sox9* is capable of activating the entire male pathway and *Sry* function may be only limited to upregulation of *Sox9*. Although all of the interacting proteins within the testis determination pathways have not been elucidated yet, the primacy of *Sry* as the major molecular switch in testis determination seems to be well established and broadly accepted.

Much of what we know today about sex determination comes from detailed studies where experimentally induced sex reversal events were linked to particular gene mutations. Sex reversal, however, is not only restricted to laboratory manipulations. The incidence of XY sex reversal in humans, for example, is about 1 in 3000 newborns and typically involves diverse genetic mechanisms. XX sex reversal on the other hand is rarer, found in only 1 in 20,000 newborns, and is caused primarily by translocation of *Sry* onto the X chromosome or autosome (Camerino et

al. 2006). Spontaneous sex reversals can also be found in laboratory populations of mice. In most cases, these individuals suffer from gonadal disgenesis, have reduced fertility or are sterile (Camerino et al. 2006).

Contrary to these observations, one group of South American rodents has persistent high frequencies of sex reversed females in natural populations. Furthermore, these XY sex reversed females seem to be quite fertile, with higher implantation rates than XX females (Lizarralde, Bianchi, and Merani 1982; Espinosa and Vitullo 1996). Bianchi and Contreras (1967) reported that about half of *Akodon azarae* females had sex chromosomes indistinguishable from those of males. Over the 35 years since this discovery, 7 more species in the genus *Akodon* were found to have naturally occurring XY females: *A. boliviensis*, *A. subfuscus*, *A. torques*, *A. kofordi*, *A. puer* (= *lutescens*), *A. varius*, and *A. mollis*. The prevalence of heterogametic females in these species in the wild ranges up to 66% (Hoekstra and Edwards 2000; Bianchi 2002). The evidence that the cause of the sex reversal lies on the Y includes cytogenetic, molecular, breeding and phylogenetic studies (Lizarralde, Bianchi, and Merani 1982; Vitullo et al. 1986; Bianchi et al. 1993; Espinosa and Vitullo 1996; Hoekstra and Edwards 2000). No YY or XYY individuals have ever been found among *Akodon* rats (Lizarralde, Bianchi, and Merani 1982; Hoekstra and Edwards 2000; Bianchi 2002).

What is the molecular mechanism causing XY sex reversal in *Akodon* females and how many times has it evolved? The obvious candidate for the causal factor of sex reversed females is the *Sry* gene. Numerous clinical cases of XY sex reversal documented mutations in the *Sry* coding region as well as in noncoding

flanking sequence (Berta et al. 1990; Koopman et al. 1991; Mcelreavy et al. 1992; Kwok et al. 1996; McElreavey et al. 1996). *Sry* is also one of the first genes in the sex determination cascade and one of only a few genes located on the Y chromosome (Brennan and Capel 2004). Several attempts have been made to obtain the complete coding sequence of *Akodon Sry* (Bianchi et al. 1993) and to reconstruct the evolutionary history of sex reversal events (Hoekstra and Edwards 2000). Because the Y chromosome is notoriously difficult to work with due to GC-poor content and highly repetitive sequence, only153bp of the *Akodon Sry* High Mobility Group domain have been sequenced to date (Bianchi et al. 1993) without any apparent degenerative mutations in males or females. Character state reconstructions also did not give clear resolution to questions about single vs. multiple origins of sex reversal within *Akodon* lineage (Hoekstra and Edwards 2000). Answers to these questions will not only help us understand the evolutionary history of akodont rodents but also further our understanding of the labiality and fate of mammalian sex determination.

To elucidate the molecular mechanism of akodont sex reversal, we analyzed DNA sequences of the entire *Sry* gene from 22 individuals belonging to males and females of 7 *Akodon* species. Furthermore, we reconstructed the history of male and female Y chromosomes, using *Sry* as a marker. Herein, we report on our results and propose a new testable model for the origin of sex-reversed females.

Materials and Methods

Tissue Samples

Twenty-four tissue samples of *Akodon azarae* were obtained from the University of Michigan Museum of Zoology for DNA extraction and analysis. These include: *A. azarae* frozen tissue samples GD063, GD067, GD068, GD069, GD073, GD-76, GD079, GD140, GD142, GD146, GD262, GD264, GD271, GD280, GD282, GD283, GD284, GD298, GD299, GD300, GD308, GD538, GD553, and GD562 (see APPENDIX D for more information on *Akodon azarae* tissue samples). DNA samples from *A. bolviensis*, *A. kofordi*, *A. molinae*, *A. subfuscus*, *A. torques*, and *A. varius* were generously provided by Hopi Hoekstra.

Sex reversal screening

To detect the presence of sex reversal in phenotypic females, polymerase chain reaction (PCR) genotyping was performed on all available female samples. PCR primers (SmcyAkodonFwd1 5' GAGTGTAAGAGTCCCCCTGAAGCTTTTGG and SmcyAkodonRev1 3' CCTCAATGCTGCTCACCAGCCTCCAGAA) were designed using mouse genomic DNA sequence as template to target the *Smcy* gene. SmcyAkodonFwd1 and SmcyAkodonRev1 primers are anchored in exons 9 and 10 respectively and amplify 255bp product containing most of both exons and the entire intron 9. *Smcy* is a single copy gene located on the Y chromosome. It is also relatively well conserved across mammals, and one set of these primers can be readily used to genotype a variety of

mammalian species. *Smcx* is a homolog of *Smcy* and is located on the X chromosome. Under lower annealing temperatures of 48C, these primers hybridize to both *Smcy* and *Smcx*, yielding two bands. The PCR product from these two genes differs in length (255bp in *Smcy*, 302 bp in *Smcx*). Therefore standard procedure PCR done on individuals carrying XX chromosomes under annealing temperature of T =48C will yield one 302bp band and on individuals carrying XY chromosomes will yield two bands 255bp and 302bp in length (Fig. 4.1). This unique set of primers therefore requires no positive control. Both bands were initially sequenced from males every time a new species was genotyped to confirm product identity. For each species, PCR was conducted following standard procedures with annealing temperatures of T=48C.

PCR amplification

Prior to our study, only 153bp of incomplete coding sequence from the *Akodon azarae Sry* gene was known. By implementing a PanHandle PCR technique (Jones and Winistorfer 1997), we were able to obtain ~830 bp upstream and ~500bp downstream of this region. The resulting 1,500bp genomic fragment contains the entire 540bp of *Sry* coding sequence, ~700bp of upstream and ~270bp of downstream non-coding sequences. Using SryAkF1 5’ AAACCTTTTATACACAAAGAGTGATGAAA and SryAkR2 3’ GTAGGTAAAAGCACTACAACCATTCTTC, annealing temperature of T = 53C (hot start) and standard PCR procedures, this 1500bp segment was amplified from males and females of *A. azarae*, *A. boliviensis*, *A. kofordi*, *A. molinae* (only males),

A. subfuscus, *A. torques*, and *A. varius*. All PCR products were purified and cloned into pCR2.1 vector (Invitrogen) before being sequenced in both directions using an automated DNA sequencer.

Alignment and Phylogeny

Alignments of DNA sequences were generated using ClustalX software (Thompson et al. 1997a). Protein coding parts of *Sry* from multiple mammal species were initially aligned per amino acid sequence and subsequently reverted back to nucleotide sequence. Maximum likelihood method implemented in GARLI 0.95 (Zwickl 2006) was used to reconstruct *Sry* gene tree. Best fitting nucleotide substitution model (HKY+G), gamma distribution shape parameter ($\alpha = 0.432$), transition-transvertion ratio ($T_i/T_v = 2.390$) and proportion of invariable sites ($p_{invar} = 0$) were estimated in Modeltest3.7 (Posada and Crandall 1998). Bootstrap support values were generated from 200 replicates.

Transcription factor binding site prediction

To investigate whether sequence upstream of *Akodon Sry* contains transcription binding sites necessary for expression, we compared the transcription binding site pattern of upstream *Sry* regions of *Akodon*, mouse, rat, and guinea pig using Alibaba2.1 software (http://www.gene-regulation.com/pub/programs/alibaba2/index.html?).

Results and Discussion

Molecular mechanisms of sex reversal in Akodon

Results of the PCR screening for sex-reversed females identified 14 XY females: *A. azarae* GD063, GD068, GD073, GD146, GD262, GD280, GD284, GD538; *A. boliviensis* 13321 (number corresponds to Hoekstra's unique sample ID); *A. kofordi* 202, *A. subfuscus* 13983, *A. varius* 21718. Implementing the PCR panhandle amplification technique, we were able to obtain sequence for a ~1500 base pair (bp) fragment containing a single exon coding region of *Sry* and flanking sequence (Fig. 4.2) from 22 individuals belonging to 7 species of *Akodon*. These include: *A. azarae* - 3 males and 8 females, *A. boliviensis* - 1 male and 1 female, *A. torques* – 1 male and 1 female, *A. kofordi* – 1 male and 1 female. *A. mollis* – 1 male, *A. subfuscus* – 1 male and 1 female, and *A. varius* - 1 male and 1 female (APPENDICES A and B). It has been previously reported that *Akodon* Y chromosome contains multiple copies of *Sry* (Bianchi et al. 1993; Hoekstra and Edwards 2000; Bianchi 2002). This finding is not unique only to *Akodon*. Other mammalian taxa including *Rattus norvegicus* (Turner et al. 2007), *Microtus cabrerae* (Fernandez et al. 2002), and numerous African rodents (Lundrigan and Tucker 1997) show up to six *Sry* copies on the Y chromosome, many of which are putatively functional.

Consistent with previous reports, we were able to recover more than one haplotype from a single individual from each akodont species. To ascertain that multiple sequences are not generated through PCR or sequencing errors, we re-

amplified the *Sry* genomic fragment from 1 male (GD067) and 2 females (GD063 and GD146) from *A. azarae* using High Fidelity Taq polymerase (Promega), cloned all PCR products and sequenced 10 colonies from male GD067 and female GD063 and 7 colonies from female GD146. Six haplotypes were recovered from male GD067 and 7 haplotypes in each of the females (GD063 and GD146). These results strongly suggest that multiple *Sry* copies are present on the *Akodon* Y chromosome.

Although sex determination in mammals involves many genes, *Sry* is considered the primary molecular switch. Laboratory mouse studies have documented cases where a nonfunctional *Sry* allele failed to trigger testes determination leading to a complete XY sex reversal (Mcelreavy et al. 1992; Kwok et al. 1996; McElreavey et al. 1996). This observation prompted us to inspect all *Sry* alleles for possible mutations that would render *Sry* nonfunctional in *Akodon* XY females.

Considering just nonsense mutations first, only *A. subfuscus* females showed a 2-nucleotide deletion causing a premature stop codon. The truncated *Sry* protein was 108 amino acids (aa) in length compared to the typical 181aa in other *Akodon*. We sequenced only 2 colonies from this female, yielding two copies, both of which were truncated. It is therefore possible that another copy without a similar frame-shifting mutation is still present in *A. subfuscus* XY females. Surprisingly, single males in *A. kofordi* and *A. varius* also showed alleles with premature stop codons, shortening the protein products by 24 aa and 22 aa respectively. In the case of the *A. kofordi* male, however, a second full-length copy was also recovered. Our data

therefore suggest that both *Akodon* males and sex reversed females might have at least one *Sry* copy with an open reading frame (ORF).

Other types of mutations in the coding region can also render genes nonfunctional. Nonsynonymous mutations alter amino acid composition causing potential problems with protein folding and function (Harley et al. 1992; Jager et al. 1992; Giese, Pagel, and Grosschedl 1994; Poulat et al. 1994). *Sry* is a member of the SOX family of developmental transcription factors. Like other SOX proteins, it is characterized by an ~80aa motif called the High Mobility Group (HMG) domain, which is crucial in binding to specific regions in the minor groove of the DNA. Additionally, in mouse and human, *Sry* causes DNA to bend at angles of 60-85 degrees (Giese, Pagel, and Grosschedl 1994; Werner et al. 1995). In most cases, *Sry* coding mutations causing sex reversal affect the ability of *Sry* to bind specifically or bend DNA at proper angle (Harley et al. 1992; Pontiggia et al. 1994). The functionality of the *Sry* C-terminus has not been well demonstrated but several studies indicate a potential role in *Sry* conformation (Li et al. 2006). We therefore attempted to identify nonsynonymous mutations localized within the HMG domain or mutations outside of this domain that would segregate with either male or female phenotypes, but we found none.

A number of amino acid replacements at more than 20 different sites in *Sry* have been previously reported in clinical studies to cause sex reversal in humans and mice (Harley, Clarkson, and Argentaro 2003; Polanco and Koopman 2007). We therefore mapped these positions onto *Akodon Sry* sequences to see whether any substitutions are preferentially located at these positions. No clear pattern, however,

arose from this comparison. The lack of substitutions in *Akodon* that would segregate with either male of female phenotypes or no apparent similarity between polymorphism at sites correlated with human or mouse clinical mutations suggests that the molecular cause of XY sex reversal in *Akodon* females is not localized to the *Sry* coding region.

Predicting the functional impact of individual amino acid replacements is complicated by the high degree of variability in size and amino acid composition of *Sry* among mammals. This variability underscores the fast pace of *Sry* evolution (Tucker and Lundrigan 1993; Whitfield, Lovellbadge, and Goodfellow 1993; Pamilo and ONeill 1997; Katoh and Miyata 1999). The HMG domain of *Akodon* has ~87% identity to mouse or rat, whereas the C-terminus is only 69% identical. This is also reflected by differences in rates of synonymous and nonsynonymous substitutions between the HMD domain region and C-terminus (HMG $d_N/d_S = 0.26$, C-terminus $d_N/d_S = 0.90$ in *A. azarae*). Such a remarkable difference could be explained either by functional relaxation at the C-terminus or adaptive evolution of a few sites within the terminal region. Inferences of function for specific amino acid positions from mice or humans may therefore be irrelevant.

Aside from the coding region, mutations in the flanking sequence can also disrupt a gene's function. In genes under relaxed selection or in the early stages of gene pseudogenization, expression may deteriorate prior to any changes in the ORF (Zhang, Carriero, and Gerstein 2004). In several cases of sex-reversed individuals with an intact *Sry* ORF, mutations as far as 2kb 5' and 3' were proposed to influence *Sry* enhancer elements and change expression (Mcelreavy et al. 1992; Kwok et al.

1996; McElreavey et al. 1996). Comparisons of *Sry* 5' regulatory regions in 10 different mammals have revealed several conserved motifs, whose function fits well with the known expression of *Sry* (Veitia, Fellous, and McElreavey 1997; Margarit et al. 1998). GATA transcription factor is known to induce proliferation and differentiation of Sertoli-cell lineages (Hannon et al. 1991; Golay et al. 1996), which is an important stage in testis development. GATA expression overlaps the onset of *Sry*, possibly influencing *Sry* regulation (Gong, Stern, and Dean 1991; Fiddler et al. 1995). Sp1 transcription factor binds to G-C rich motifs and is very common in most mammalian genes. It has a more ubiquitous regulatory function but its high levels of expression in mouse fetal cells and spermatids suggest a potentially important role in the early determination of genes (Saffer, Jackson, and Annarella 1991). NF1 element has also been connected with *Sry* expression due to similar function in inducing growth regulator expression during early phases of embryogenesis (Lemaigre et al. 1989; Courtois et al. 1990; Lemaigre et al. 1990; Saffer, Jackson, and Annarella 1991). We therefore subjected our data to computational prediction of transcription binding sites to investigate whether changes in the 5' region of *Sry* are likely to disrupt expression of *Sry* in XY sex-reversed females. We implemented this approach using Alibaba2.1 (http://www.gene-regulation.com/pub/programs/alibaba2/index.html?) and focused on previously described binding motifs, namely, GATA, Sp1 and NF1 (Margarit et al. 1998). 500 bp DNA fragments upstream of mouse (from genome sequence assembly NCBI m36), rat (AJ222688), and guinea pig (AJ003126) *Sry* were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/) for comparison. Our first observation was that the

Akodon 5’ region is barely alignable to any of the rodent taxa above. While *Akodon* shares a more recent common ancestor with mouse and rat than with guinea pig (Adkins et al. 2001), guinea pig and mouse 5’ regions are more conserved. Guinea pig shows 85-97% similarity with mouse and rat upstream region compared to ~45% similarity with *Akodon*. Even though this implies that upstream regions of *Akodon* are not homologous with those of other rodent species, analysis of individual sequences revealed the presence of GATA, Sp1, and NF1 binding motifs in all species, including *Akodon*. No significant differences in binding motif frequency could be found among *Akodon* males and sex-reversed females. This implies that although the sequence upstream of *Sry* in *Akodon* has significantly diverged, the ability to bind transcription factors to initiate expression is still retained.

Origins of sex reversed females

Having acquired large segment of *Sry* locus gives us the opportunity to investigate yet another aspect of *Akodon* evolution. Our data confirm that at least 7 *Akodon* species (and 1 additional taxon *A. puer* [= *lutescens*] reported elsewhere [(Hoekstra and Edwards 2000) have sex reversed females. This observation brings forth a question whether sex reversal widespread among all these lineages has arisen once in their common ancestor or multiple times, independently in each species?

In an attempt to address this question, a previous study (Hoekstra and Edwards 2000) used mtDNA to reconstruct the *Akodon* phylogeny and mapped sex reversal events onto the tree using ACCTRAN (accelerated transitions) or DELTRAN (delayed transitions) methods (Maddison and Maddison 1992). This

approach has unfortunately yielded little resolution because the dataset used equally supported single and multiple origins of the sex-reversed female system. Additionally, to resolve the relationships between male and sex reversed female Y chromosomes, both within and between species, Hoekstra and Edwards (2000) examined the restriction pattern variation in XY males and XY females using *Sry* specific probe. Southern blot experiment revealed that restriction patterns were always the same within each of the six tested species but differed significantly between species, suggesting that sex reversal originated independently multiple times within *Akodon* genus. Congruent with these results, phylogenetic reconstruction of our *Sry* sequences from 7 *Akodon* species shows males and sex reversed females to predominantly cluster according to the species rather than gender (Fig. 4.3). The single exception is *A. varius,* whose male and female *Sry* sequences form two separate lineages basal to *A. mollis* and *A. torques* group. This discrepancy could be caused by unequal rates of gene conversion among *Sry* copies, or the possible non-orthology of these sequences. Overall, the gene tree is largely consistent with our current understanding of *Akodon* evolution (D'Elia 2003; Goncalves et al. 2006). In light of previously published results of Hoekstra and Edwards (2000), it could be said that *Sry* data further support multiple origins of sex reversal in *Akodon* females. However, we believe that ours results, as well as Hoekstra and Edwards', can be interpreted differently, arguing for a single origin of the sex reversal mechanism in akodont rodents.

Hoekstra and Edwards's conclusion on multiple XY female origins is based on the assumption that male and sex reversed female Y chromosomes form distinct

lineages, where male Y is inherited only paternally and sex reversed female Y only maternally (Hoekstra and Edwards 2000). The Southern blot data shows that male and sex reversed female Y chromosomes create identical restriction patterns. Under the assumption that both Y chromosomes are distinct, this would suggest that the origin would have to be very recent in all six species tested, which we find very unlikely. We would anticipate the within species comparison to yield some pattern differences because male and female Y chromosomes would be expected to diverge over time simply due to continuing mutation process and drift. Identical restriction patterns between males and sex reversed females would suggest that indeed the Y chromosomes do not evolve independently in each sex. In support of this notion, cytogenetic studies on *Akodon* karyotype claim male and sex reversed female Y chromosomes to be indistinguishable (Bianchi and Contrera.Jr 1967; Bianchi et al. 1971; Bianchi and Merani 1984; Vitullo et al. 1986). Restriction polymorphism is a very crude way to compare DNA sequences because only mutations within restriction sites or insertions and deletions between restriction sites will generate visible differences. All other sequence variation will not be detectable. Restriction pattern polymorphism was however the only possible comparison technique prior to our study because long enough Y chromosome sequences have not been obtained from *Akodon* species. Our dataset therefore enables us to investigate the question of female sex reversal origin in potentially far greater detail.

There are two possible mechanisms that would allow for male and sex reversed female Y chromosomes to be essentially “the same”. First, the cause of the female sex reversal could lie outside of the Y chromosome, in which case sex

reversed individuals could still be generated even if they inherited male Y. This mechanism might involve a “feminizing” factor (either *Sry* suppressor or ovary determining factor) on the X chromosome or an autosome. Breeding experiments, however, show that sex reversed individuals are only generated from crosses involving sex reversed XY females and never XX females (Bianchi and Contrera.Jr 1967; Lizarralde, Bianchi, and Merani 1982; Vitullo et al. 1986; Espinosa and Vitullo 1996). This observation therefore strongly argues against this possibility because XX female carrier of the “feminizing factor” should also give birth to sex reversed XY females. In the second scenario, the cause of the sex reversal lies on the Y chromosome. This second model requires a dynamic process, where the pool of male Y chromosomes degenerates at a particular rate, causing the male Y to loose its ability to trigger testis determination. A male with such degenerative mutation on the Y chromosome (Y*) would sire only females, because all XY* offspring would be sex reversed. If these sex reversed XY* females were fertile, then Y* could further propagate and persist in the population. Given a particular Y => Y* mutation rate, this process provides a constant flow of male Y into sex reversed females within each species and would consequently result in similar restriction patterns within species and conspecific phylogenetic clustering of male and sex reversed female Y chromosomes. Furthermore, phylogenetic analysis should also reveal that within species Y and Y* do not form monophyletic groups, but rather intermingle.

Is there any support for these predictions in our data? The *Sry* gene tree topology does reflect conspecific clustering of male and female sequences but because of the gene’s multicopy nature, the pattern of within species clustering (as is

seen in *A. azarae* for example) cannot be reliably used in support of the second prediction. Analysis of a single copy gene on the Y chromosome should resolve this issue. This is important because if Y* originated only once in each *Akodon* species, than the independent evolution of Y and Y* should result in Y and Y* monophyletic groups within species. If, however, male Y degenerates at a particular rate and enters sex reversed female Y chromosome pool, then Y and Y* should intermingle within each species, giving support to the second model.

This brings forth another issue. To avoid any confusion due to semantics, it is necessary to clarify what is meant by the origins of sex reversal. By addressing the question of sex reversed female origins in *Akodon*, we are pursuing to elucidate the origins of the mechanism that allows for sex reversal in *Akodon* females, and not the origins of sex reversed females themselves. This is an important distinction, because as it implies from the second model, such process could allow for single origin of the sex reversal mechanism (the ability of the *Akodon* sex determination pathway to allow existence of fertile sex reversed XY females) and at the same time multiple origins of sex reversed females between or even within species. Hoekstra and Edwards (2000), in their article about the origins of sex reversed XY females in *Akodon*, do not explicitly phrase the "origin of sex reversed XY females" to mean the origin of sex determination mechanism, which allows for fertile sex reversed females in *Akodon*. But we assume that is what Hoekstra and Edwards implied.

The results of our *Sry* phylogenetic analysis show that sequences, regardless of whether from male or female, group according to the species. This suggests that a trait, which allows for the development of fertile sex reversed female, is shared by

all taxa examined. In conclusion, the most parsimonious explanation for the evolution of such developmental mechanism is that it had evolved only once in the common ancestor of these akodont rodents.

In mice and humans, XY sex reversed individuals often suffer from severe gonadal dysgenesis and are either sterile or subfertile (refs. in (Camerino et al. 2006). How do we then explain the change in the *Akodon* sex determination pathway that allows for fully fertile sex reversed females? The severe side effects accompanying XY sex reversal in mice and humans could be a result of deleterious mutations in multifunctional sex determination genes. For example previously mentioned gene *Igf1r*, a member of the insulin receptor tyrosine kinase family, upregulates expression of both *Sry* and *Sox9*. It is also involved in smooth muscle signaling, skin development and differentiation, insulin-receptor signaling and liver regeneration (http://www.ncbi.nlm.nih.gov/projects/GeneRIF/). Binding motif of *Sp1* transcription factor are found not only upstream of *Sry*, but upstream of most mammalian genes (Harley, Clarkson, and Argentaro 2003). It is therefore easy to imagine the pleiotropic effects of mutations in multifunctional sex determination genes. However, the magnitude of pleiotropic effect for a gene in one species may be different from that in another species. Subfunctionalization process among duplicate genes may reduce pleiotropy and result in less severe side effects, should a mutation render one homolog nonfunctional. Is there any evidence for similar shifts in the sex determination pathway?

Studies in two species of mole voles (*Ellobius*) suggest that, despite their seemingly crucial role, the function of certain sex-determining genes may be either

taken up by other genes or might be dispensable altogether (Just et al. 1995). In *E. lutescens,* both sexes are XO and in *E. tancrei* both sexes are XX. It had been postulated that *Sry* gene in these two species had simply translocated to another chromosome. However, it is now clear, that *Sry*, along with other male specific genes, such as *Zfy*, completely disappeared along with the entire Y chromosome. Evidently, a new sex determination systems had to emerge in *Ellobius*. Similarly, sex-determining genes in *Akodon* might have either taken up or given up various roles enabling these rodents to suffer no fertility reduction when a molecular switch fails to trigger testis development. The precise molecular mechanism that allows this still remains unknown.

Our analysis of the sex determination factor *Sry* in *Akodon* has left us with more questions than answers as to how *Sry*, if at all, contributes to the bizarre phenomenon of persisting sex-reversed females. Although we present the largest *Sry* dataset for *Akodon* to date, we did not find any mutations that would be indicative of loss of function in sex reversed females. Our phylogenetic evidence supports the notion that the ability to produce fertile sex reversed females may not have evolved multiple times, but only once in the akodont ancestor of the seven species examined here.

Acknowledgments

We thank Hopi Hoekstra for providing valuable tissue samples from *Akodon* rodents and dissertation committee for valuable comments. This work was supported by research grants from the University of Michigan and the National Institutes of Health to Jianzhi Zhang.

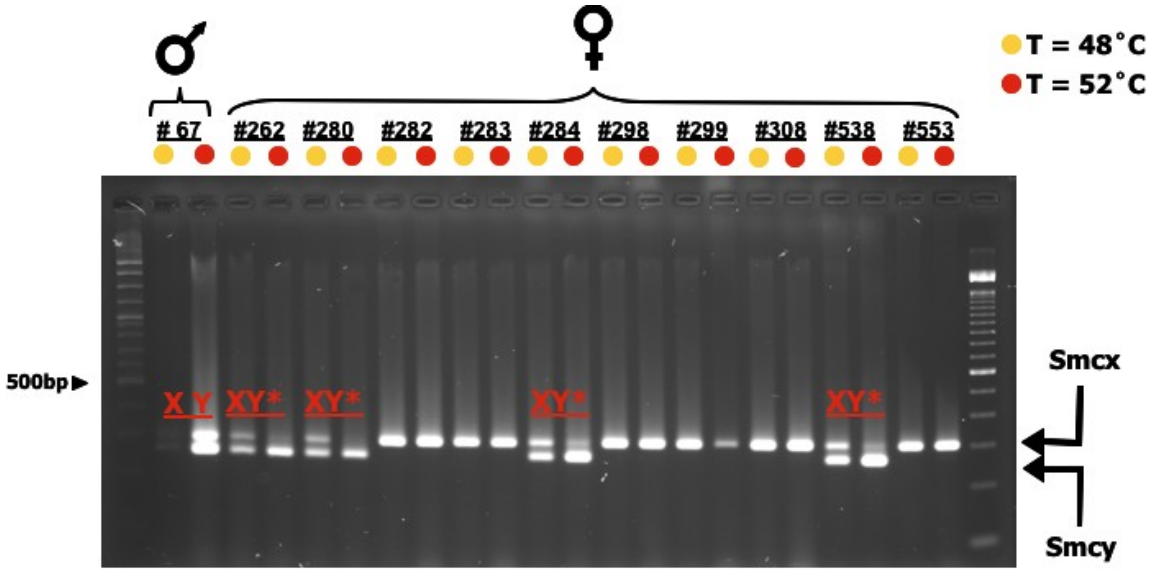


FIG. 4.1. PCR genotyping was performed to screen for XY* females. Using SmcyAkoronF1 and SmcyAkodonR1 set of primers, XY individuals will yield 2 bands whereas XX individuals only one. In this sample of 10 females, 4 females are found to be XY sex-reversed (see methods for details)

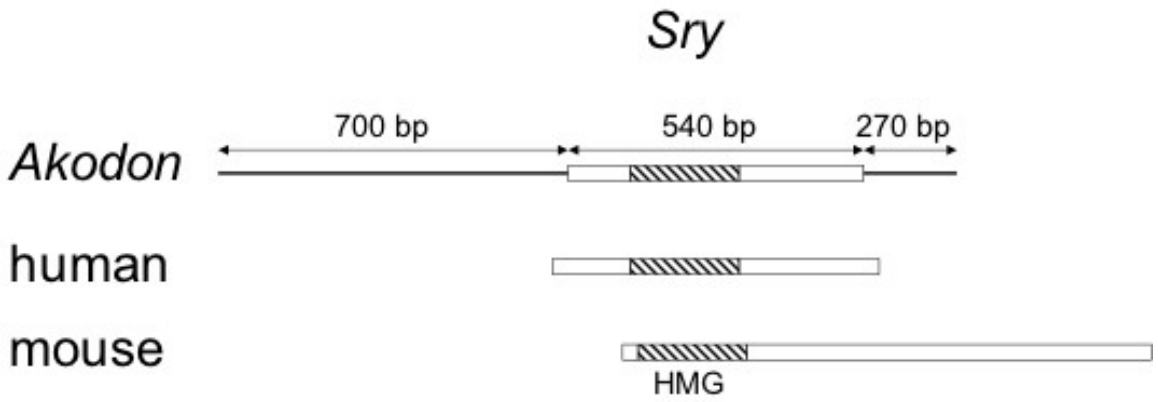


FIG.4.2. The diversity of *Sry* gene structure between *Akodon* , humans, and mice. High Mobillity Group domain (HMG), which is responsible for site specific binding of *Sry* in the minor groove of DNA, is depicted as hashed region. Corresponding lengths of sequenced segments of *Sry* and flanking regions are shown for *Akodon*.

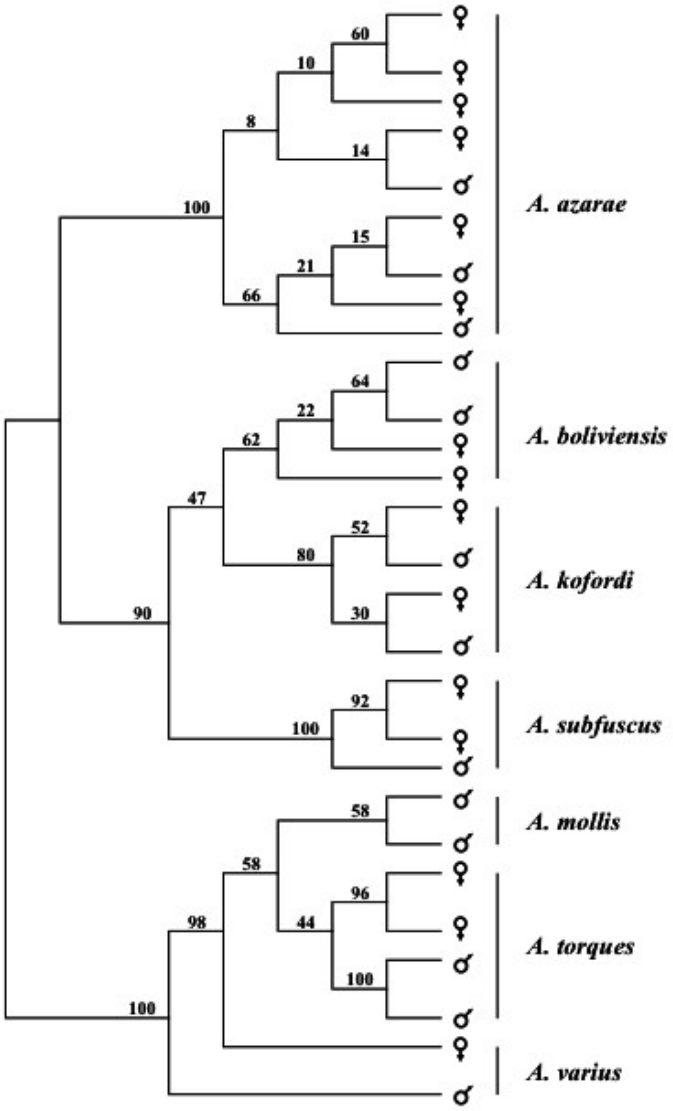


FIG. 4.3. Phylogenetic relationships of *Sry* homologs from males and females of various akodont species. This phylogeny was reconstructed using maximum likelihood implemented in GARLI 0.95 (Zwickl 2006). Bootstrap values derived from 200 replications are shown on interior branches.

APPENDICES

APPENDIX A

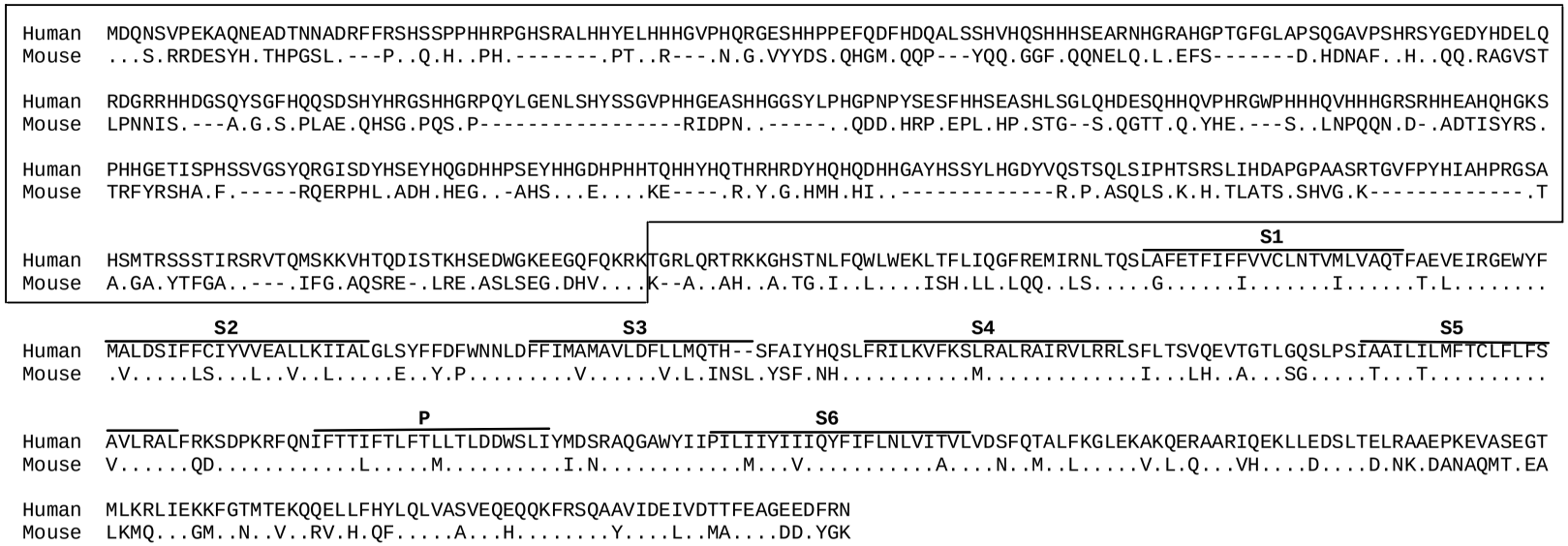
Supplementary table and figures to Chapter 1.

Supplementary Table 1. Comparison of the substitution rates of 3*n* indels from genomic data and from primate CATSPER1 exon 1 based on the alignment with fewest gaps.

		Indel rate (per site per 10^{11} year)			Number of indels in CATSPER1 exon 1			Probability	
Comparisons	Divergence [a] (MY)	From ref. (37)	From ref. (40)	From CATSPER1	Expectation from ref. (37)	Expectation from ref. (40)	Observation	under ref. (37)	un ref.
Hominoids vs. OW monkeys	23×2	1.92	1.50	9.81	1.17	0.92	6	1.3×10^{-3}	3.9
Hominoids vs. NW monkeys	35×2	1.92	1.50	11.4	1.79	1.40	10.6	1.9×10^{-5}	2.2
OW vs. NW monkeys	35×2	1.92	1.50	11.4	1.79	1.40	10.6	1.9×10^{-5}	2.2

[a] The divergence times follow ref. (41).

[b] The probabilities of the observation given the expectation calculated from ref. (37) and (40) are computed under the assumption that the number of indels follows a Poisson distribution.



Supplementary Fig. 1. Alignment of the complete CATSPER1 sequences of humans and mice. The alignment is generated by CLUSTAL X with the default parameters. "." represents an identical amino acid to the first sequence and "–" represents an alignment gap. Six putative transmembrane domains are marked S1-S6 and the pore region P. The boxed region is encoded by exon 1, which constitutes the majority of the intracellular N-terminus of this ion channel. The C-terminus region that is downstream of S6 is also intracellular.

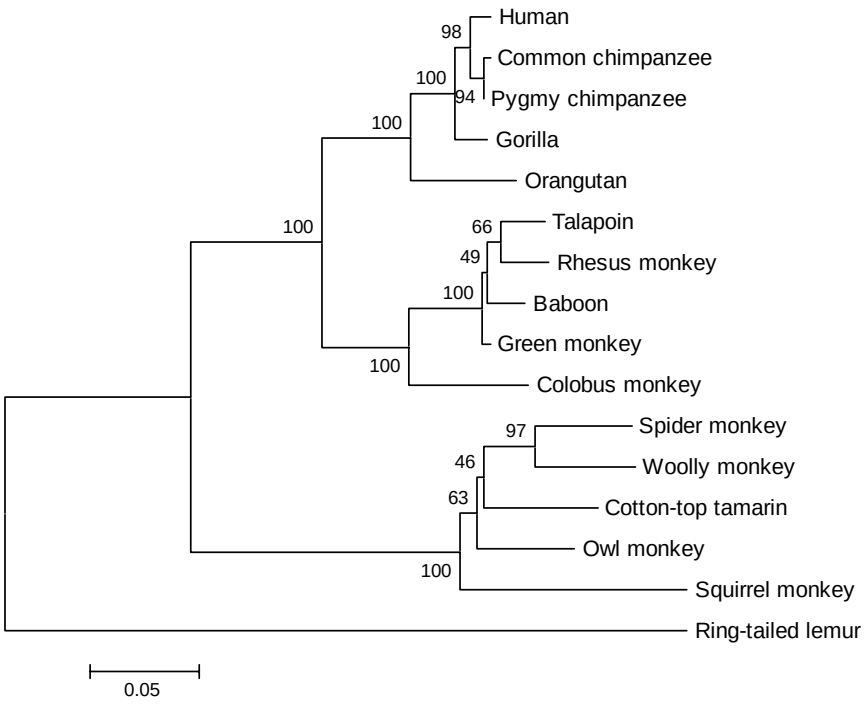
Primers for exon 1 of CATSPER1 for all 16 primates

5’	AGTTCCCAACACAGTCATGGATCAA	3’
5’	AGCAAAGACTCACTTTTGCGTTTCTG	3’
5’	ACAGCGAGGCTTCCCACCTTAG	3’
5’	ATGCTGGGACTCATCGTGTTGG	3’
5’	AGTTCCCAACACAGTCATGGATCA	3’

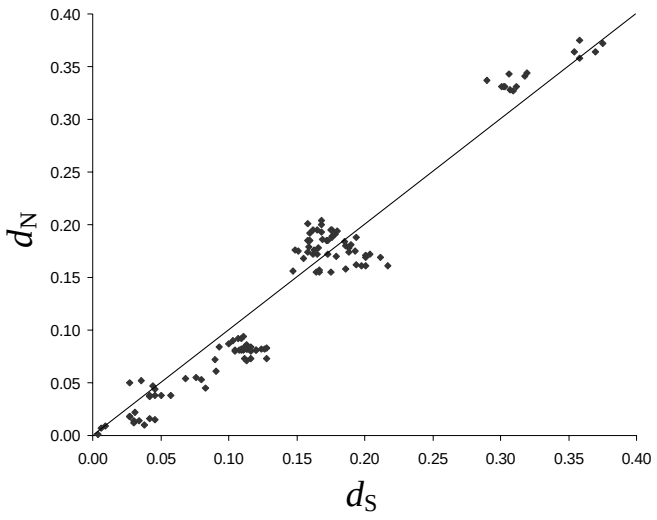
Primers for intron 1 of CATSPER1 for the rhesus monkey

5’	CCCACAGCCACTGGAAGAGATT	3’
5’	AAGGCCAAGGATTGGGTCAG	3’

Supplementary Fig. 2. Primer sets used to amplify exon 1 and intron 1 of CATSPER1.



Supplementary Fig. 3. Gene tree of CATSPER1 exon 1 from 16 primates. The neighbor-joining method (36) with protein *p*-distance (29) was used to reconstruct the tree. Bootstrap percentages from 1000 replications are shown at interior nodes. The branching patterns that are supported with >70% bootstrap values are identical to those of the species tree (Fig. 2).

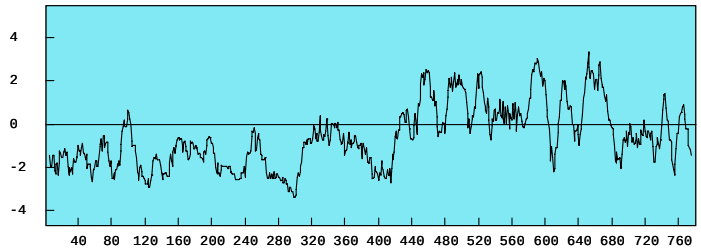


Supplementary Fig. 4. Pairwise comparisons of d_S and d_N for CATSPER1 exon 1 sequences of 16 primates. The diagonal line represents $d_N = d_S$.

Supplementary Fig. 5. Hydropathy plots for human **(A)** CATSPER1, **(B)** $K_V1.4$, **(C)** $Ca_V1.1$, and **(D)** $Na_V1.5$ ion channels, by WinPep. Hydrophobic and hydrophilic residues are represented above and below the line, respectively.

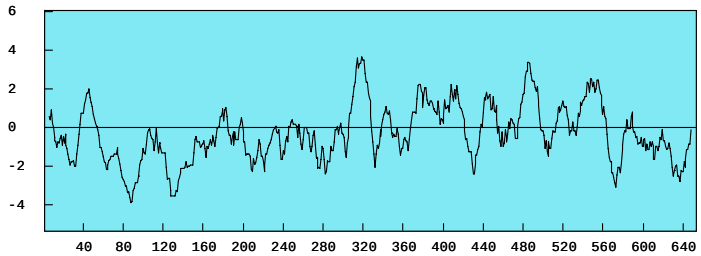
Mean Hydropathy = -0.614
Scale: Kyte and Doolittle (1982)

A



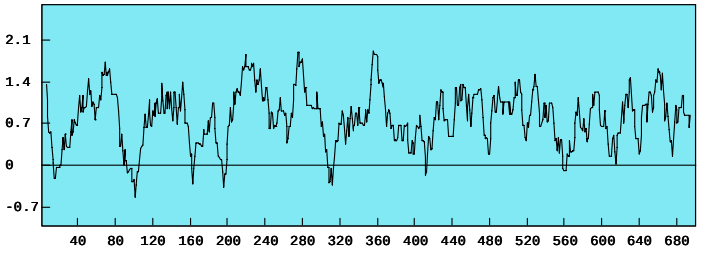
Mean Hydropathy = -0.328
Scale: Kyte and Doolittle (1982)

B



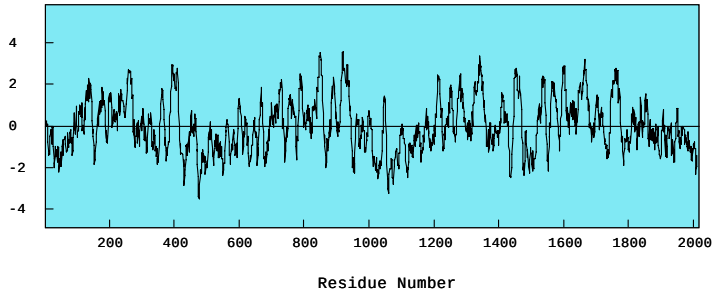
Mean Hydropathy = 0.808
Scale: Kyte and Doolittle (1982)

C



Mean Hydropathy = 0.00595
Scale: Kyte and Doolittle (1982)

D



APPENDIX B

Nucleotide sequence alignment of *Sry* from 7 *Akodon* species. This alignment is derived from CLUSTALX with default parameters. Dashes represent alignment gaps and dots represent same nucleotide to the first sequence. Highlighted area depicts *Sry* coding regio

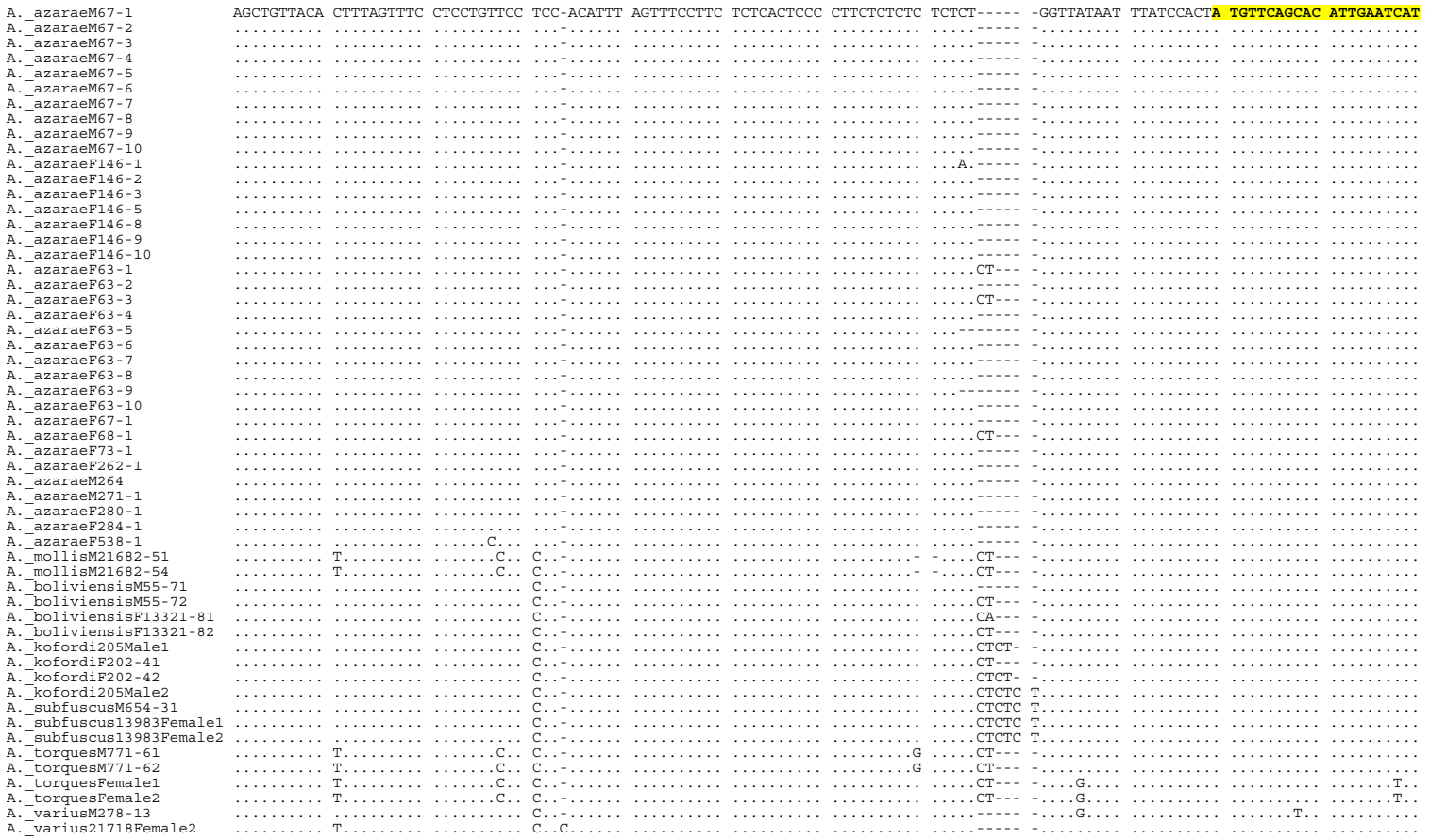
```
A._azaraeM67-1           -AAACCTTTT ATACACAAAG AGTGATGAAA TTTAAAATAG TATTAACATA CTACAAAACC ATTAGGGCTT ATAAGTTCCC AAGTAAACCT GATAGAAAAC ATAGCAGTTT TTCTTGGCTT 120
A._azaraeM67-2           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeM67-3           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeM67-4           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeM67-5           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeM67-6           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeM67-7           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeM67-8           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeM67-9           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeM67-10          --........ ......... -......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF146-1          -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF146-2          -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF146-3          -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF146-5          A..C.T.... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF146-8          -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF146-9          -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF146-10         -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF63-1           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF63-2           -......... ........-. ......... ......... ......... ..G...... ......... ......... ......... ......... ......... .........
A._azaraeF63-3           ---------- ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF63-4           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF63-5           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF63-6           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF63-7           -......... ......... .......G.. ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF63-8           -......... ........-. ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF63-9           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF63-10          -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF67-1           --........ ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF68-1           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF73-1           -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF262-1          -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeM264            -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeM271-1          -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF280-1          -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF284-1          -......... ......... ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._azaraeF538-1          --........ ........-. ......... ......... ......... ......... ......... ......... ......... ......... ......... .........
A._mollisM21682-51       -......... ......... ......... .........A ......... ......... ......... ......... .....T.... ......... ......... .........
A._mollisM21682-54       -......... ......... ......... .........A ......... ......... ......... ......... .....T.... ......... ......... .........
A._boliviensisM55-71     -......C.. ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._boliviensisM55-72     -......... ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._boliviensisF13321-81  -......... ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._boliviensisF13321-82  -......... ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._kofordi205Male1       -......... ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._kofordiF202-41        --........ ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._kofordiF202-42        --........ ....-..... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._kofordi205Male2       -......... ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._subfuscusM654-31      -......... ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._subfuscus13983Female1 -......... ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._subfuscus13983Female2 -......... ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._torquesM771-61        -......... ......... ......... .........A ......... ......... ......... ......... .....T.... ......... ......... .........
A._torquesM771-62        -......... ......... ......... .........A ......... ......... ......... ......... .....T.... ......... ......... .........
A._torquesFemale1        -......... ......... ......... .........A ......... ......... ......... ......... .....T.... ......... ......... .........
A._torquesFemale2        -.....-... ......... ......... .........A ......... ......... ......... ......... .....T.... ......... ......... .........
A._variusM278-13         -......... ......... ......... .........A ......... ......... ......... ......... ......... ......... ......... .........
A._varius21718Female2    -......... ......... ......... .........A ......... ......... ......... ...T..... .....T.... ......... ......... .........
```

```
A._azaraeM67-1           CAAAATTGAG TTCATGCATT ACTACTATGG ACTATTGAAT GAGGTTATAC TATTTACTTG TTTGTTCAGT CCTAAACTTC CTTAGAGGAC TAGGATACAA ATGACAGGTT TCATGGTCAT 240
A._azaraeM67-2           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-3           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-4           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-5           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-6           ....G..... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-7           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-8           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-9           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-10          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-2          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-3          .......... .......... .......... .......... .......... .......... .......... ..........T.......... .......... .......... ..........
A._azaraeF146-5          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-8          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-9          .......... .......... .......... .......... ..........C..... .......... .......... .......... .......... .......... ..........
A._azaraeF146-10         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-1           .......... .......... .......... .......... ..........C..... .......... .......... .......... .......... .......... ..........
A._azaraeF63-2           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-3           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-4           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-5           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-6           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-7           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-8           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-9           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-10          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF67-1           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF68-1           .......... .......... .......... .......... ..........C..... .......... .......... .......... .......... .......... ..........
A._azaraeF73-1           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF262-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM264            .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM271-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF280-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF284-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF538-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._mollisM21682-51       .......... ........C. .......... ........C. .......... .......... .......... .......... .......... .......... .......... ..........
A._mollisM21682-54       .......... ........C. .......... ........C. .......... .......... .......... .......... .......... .......... .......... ..........
A._boliviensisM55-71     .......... ........C. .......... .......... .......... .......... .......... .......... ........T. .......... .G........ ..........
A._boliviensisM55-72     .......... ........C. .......... .......... .......... .......... .......... .......... ........T. .......... .G........ ..........
A._boliviensisF13321-81  .......... ........C. .......... ........G. .......... .......... .......... .......... .......... .......... .G........ ..........
A._boliviensisF13321-82  .......... ........C. .......... .......... .......... .......... .......... .......... .......... .......... .G........ ..........
A._kofordi205Male1       .......... ........C. .......... .......... .......... .......... .......... .......... .......... .......... .......... ......C...
A._kofordiF202-41        .......... ........C. .......... .......... .......... .......... .......... .......... .......... .......... .G........ ......C...
A._kofordiF202-42        .......... ........C. .......... .......... .......... .......... .......... .......... .......... .......... .......... ......C...
A._kofordi205Male2       .......... ........C. .......... .......... .......... .......... .......... .......... .......... .......... .......... ......C...
A._subfuscusM654-31      .......... ........C. .......... .......... .......... .......... .......... .......... .......... .......... .......... ........G.
A._subfuscus13983Female1 .......... ........C. .......... .......... .......... .......... .......... .......... .......... .......... .......... ........G.
A._subfuscus13983Female2 .......... ........C. .......... .......... .......... .......... .......... .......... .......... .......... .......... ........G.
A._torquesM771-61        .......... ........C. .......... .........C .......... .......... .......... .......... .......... .......... .......... ..........
A._torquesM771-62        .......... ........C. .......... .........C .......... .......... .......... .......... .......... .......... .......... ..........
A._torquesFemale1        .......... .....A..C. .......... .......... .......... .......... .......... .......... .......... C......... .......... .........C
A._torquesFemale2        .......... ........C. .......... .......... .......... .......... .......... .......... .......... .......... .......... .........C
A._variusM278-13         .......... C.......C. .......... .........C .......... .......... .......... .......... .......... .......... .......... ..........
A._varius21718Female2    .......... ........C. .......... .........C .......... .......... .......... .......... .......... .......... .......... ..........
```

```
A._azaraeM67-1           CCTTTTCACT GTCCCTATTG GTTTCTTTAG TTAATTAGTT TTCATAGTAG TTCATCATTA AAAAAAA-TC ACATTACAGG ACAAAAAAGT TCACTGGTAT ACTCTTTGTT GTACCCTTCA 360
A._azaraeM67-2           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeM67-3           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeM67-4           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeM67-5           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeM67-6           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeM67-7           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeM67-8           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeM67-9           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeM67-10          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF146-1          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF146-2          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF146-3          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF146-5          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF146-8          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF146-9          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF146-10         .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF63-1           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF63-2           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF63-3           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF63-4           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF63-5           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF63-6           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF63-7           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF63-8           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF63-9           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF63-10          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF67-1           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF68-1           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF73-1           .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF262-1          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeM264            .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeM271-1          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF280-1          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF284-1          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._azaraeF538-1          .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ..........
A._mollisM21682-51       .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ...T......
A._mollisM21682-54       .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ...T......
A._boliviensisM55-71     .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... .........G
A._boliviensisM55-72     .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... .........G
A._boliviensisF13321-81  .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... .........G
A._boliviensisF13321-82  ....C..... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... .........G
A._kofordi205Male1       .......... .......... .......... .......... .......... .......... .......A.. .......... .......... .......... .......... .........G
A._kofordiF202-41        .......... .......... .......... .......... .......... .......... .......A.. .......... .......... .......... .......... .........G
A._kofordiF202-42        .......... .......... .......... .......... .......... .......... .......A.. .......... .......... .......... .......... .........G
A._kofordi205Male2       .......... .......... .......... .......... .......... .......... .......A.. .......... .......... .......... .......... .........G
A._subfuscusM654-31      .......... .......... .......... .......... .......... .......... .......A.. .......... G......... .......... .......... .........G
A._subfuscus13983Female1 .......... .......... .......... .......... .......... .......... .......A.. .......... .......... .......... .......... .........G
A._subfuscus13983Female2 .......... .......... .......... .......... .......... .......... .......A.. .......... .......... .......... .......... .........G
A._torquesM771-61        .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ...T......
A._torquesM771-62        .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ...T......
A._torquesFemale1        .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ...T......
A._torquesFemale2        .......... .......... .......... .......... .......... .......... .......-.. .......... .......... .......... .......... ...T......
A._variusM278-13         .......... .......... .......... .......... .....G.... .......... .......-.. .......... .......... .......... .......... ..........
A._varius21718Female2    .......... .......... .......... .......... .......... .......... .......A.. .......... .......... .......... .......... ...T......
```

```
A._azaraeM67-1          CCTTGATGTG AAAAACTGAT AGAGGAAATG ATGTGTTTTT AAGGAGTGTA GAAGTGCAGC ATCTAAGTTA TAGTGATTTT GGTTCACAAT AGGTAGCTCT CTACAGGGTG CCGAGCTGAC 480
A._azaraeM67-2          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-3          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-4          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-5          .......... .......... .......... .......... .......... ......T... .......... .......... .......... .......... .......... ..........
A._azaraeM67-6          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-7          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-8          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-9          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-10         .T........ .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-1         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-2         .......... .......... .......... .......... .......... .......... .......... .......... .....G.... .......... .......... ..........
A._azaraeF146-3         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-5         .......... .......... .......... .......... .......... .......T.. .......... .......... .......... .......... .......... ..........
A._azaraeF146-8         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-9         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-10        .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-2          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-3          .......... .......... .......... .......... .......... .......... .....G.... .......... .......... .......... .......... ..........
A._azaraeF63-4          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-5          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-6          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-7          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-8          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-9          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-10         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF67-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF68-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF73-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF262-1         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM264           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM271-1         .......... .......... .......... .......... ......C... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF280-1         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..T....... ..........
A._azaraeF284-1         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF538-1         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._mollisM21682-51      .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._mollisM21682-54      .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._boliviensisM55-71    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._boliviensisM55-72    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._boliviensisF13321-81 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._boliviensisF13321-82 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._kofordi205Male1      .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._kofordiF202-41       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._kofordiF202-42       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._kofordi205Male2      .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._subfuscusM654-31     .......... ....G..... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._subfuscus13983Female1 .......... ....G..... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._subfuscus13983Female2 .......... ....G..... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._torquesM771-61       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._torquesM771-62       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._torquesFemale1       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..T.......
A._torquesFemale2       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..T.......
A._variusM278-13        .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... - .......... ..........
A._varius21718Female2   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
```

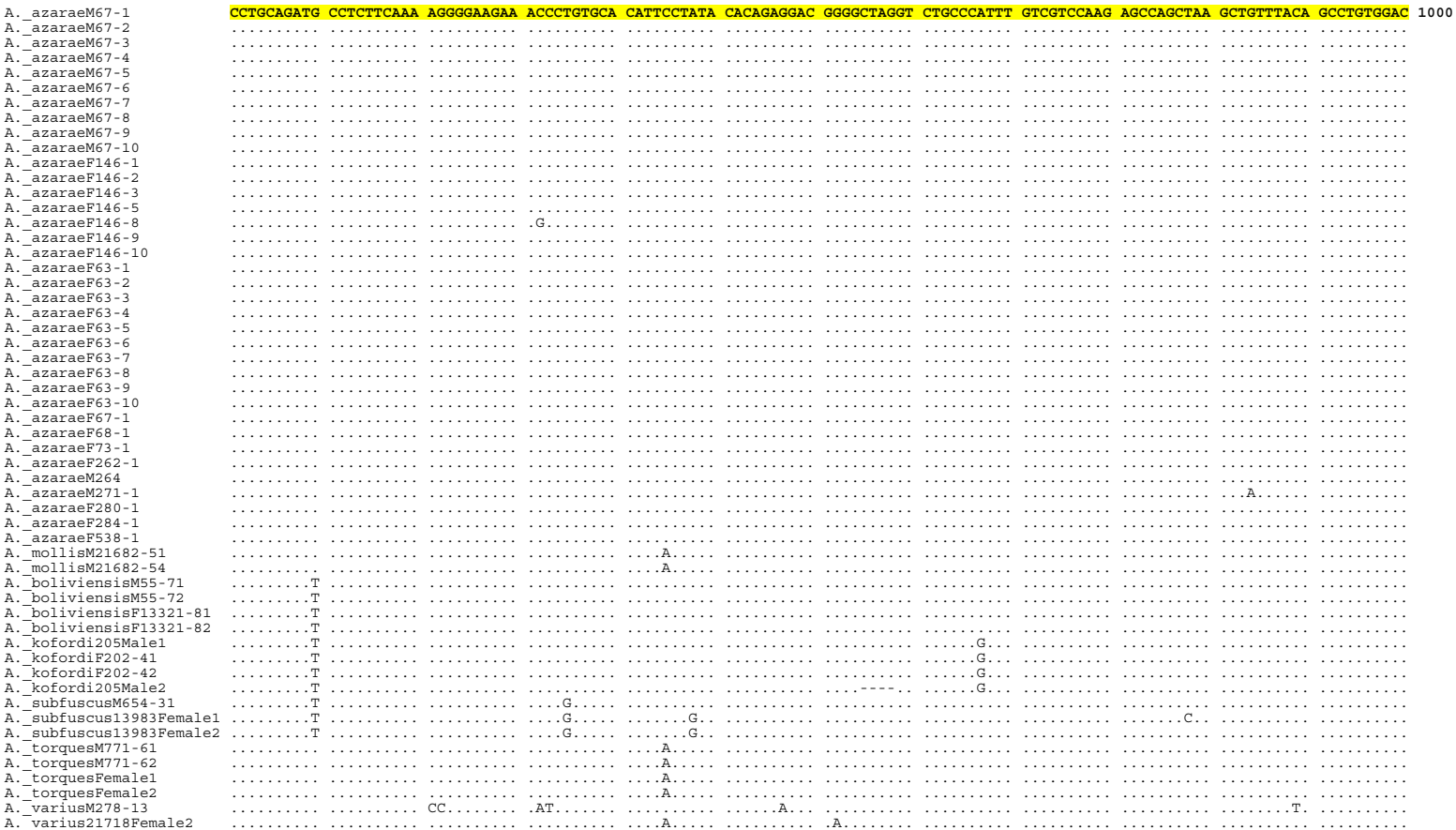
```
A._azaraeM67-1           AAGTTACATT AGTTGGGGCT GGGCTAAG-- -GGAGGGCTG AGGGAGGAGA AATGAATATT TTGTTACACA GTTTAAATAA CAAGATCTTT ATGACAGACA CATTTTGGAT AGTAGCTTTG 600
A._azaraeM67-2           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-3           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-4           .......... .......... .......A-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-5           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-6           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-7           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-8           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-9           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-10          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-1          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-2          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-3          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-5          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-8          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-9          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-10         .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-1           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-2           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-3           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-4           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-5           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-6           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-7           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-8           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-9           ...C...... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-10          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF67-1           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF68-1           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF73-1           .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF262-1          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM264            ...C...... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM271-1          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF280-1          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF284-1          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF538-1          .......... .......... ........-- ---------- .......... .......... .......... .......... .......... .......... .......... ..........
A._mollisM21682-51       .......... .......... ........-- ---------- .......... G......... .......... .......... .......... G......... .......... ..........
A._mollisM21682-54       .......... .......... ........-- ---------- .......... G......... .......... .......... .......... G......... .......... ..........
A._boliviensisM55-71     .......... .......... ........-- ---------- .......... .......... .......... .......... .......... G......... .......... .......G..
A._boliviensisM55-72     .......... .......... ........-- ---------- .......... .......... .......... .......... .......... G......... .......... .......G..
A._boliviensisF13321-81  .......... .......... ........-- ---------- .......... .......... .......... .......... .......... G......... .......... .......G..
A._boliviensisF13321-82  .......... .......... ........-- ---------- .......... .......... .......... .......... .......... G......... .......... .......G..
A._kofordi205Male1       .......... .......... ........-- ---A...... .......... .......... .......... .......... .......... G......... .......... ..........
A._kofordiF202-41        .......... .......... ........-- ---A...... .......... .......... .......... .......... .......... G......... .......... .......G..
A._kofordiF202-42        .......... .......... ........-- ---A...... .......... .......... .......... .......... .......... G......... .......... ..........
A._kofordi205Male2       .......... .......... ........-- ---A...... .......... .......... .......... .......... .......... G......... .......... ..........
A._subfuscusM654-31      .......... .......... ........-- ---------- .......... .......... .......... .......... .......... G......... ........T. ..........
A._subfuscus13983Female1 .......... .......... ........-- ---------- .......... .......... .......... .......... ...C...... G......... ........T. ..........
A._subfuscus13983Female2 .......... .......... ........-- ---------- .......... .......... .......... .......... ...C...... G......... ........T. ..........
A._torquesM771-61        .......... .......... ........-- ---------- .......... G......... .......... .......... .......... G.........G......... ..........
A._torquesM771-62        .......... .......... ........-- ---------- .......... G......... .......... .......... .......... G.........G......... ..........
A._torquesFemale1        .......... .......... ........-- ---------- .......... G......... .......... .......... .......... G......... .......... ..........
A._torquesFemale2        .......... .......... ........-- ---------- .......... G......... .......... .......... .......... G......... .......... ..........
A._variusM278-13         .......... .......... .......AA G......... .......... G......... .....C.... .......... .......... G....T.... .......... ..........
A._varius21718Female2    .......... .......... ........-- ---------- .......... G......... .......... .......... .......... G......... .......... ..........
```


A._azaraeM67-1 AGCTGTTACA CTTTAGTTTC CTCCTGTTCC TCC-ACATTT AGTTTCCTTC TCTCACTCCC CTTCTCTCTC TCTCT----- -GGTTATAAT TTATCCACTA TGTTCAGCAC ATTGAATCAT 720
A._azaraeM67-2
A._azaraeM67-3
A._azaraeM67-4
A._azaraeM67-5
A._azaraeM67-6
A._azaraeM67-7
A._azaraeM67-8
A._azaraeM67-9
A._azaraeM67-10
A._azaraeF146-1
A._azaraeF146-2
A._azaraeF146-3
A._azaraeF146-5
A._azaraeF146-8
A._azaraeF146-9
A._azaraeF146-10
A._azaraeF63-1
A._azaraeF63-2
A._azaraeF63-3
A._azaraeF63-4
A._azaraeF63-5
A._azaraeF63-6
A._azaraeF63-7
A._azaraeF63-8
A._azaraeF63-9
A._azaraeF63-10
A._azaraeF67-1
A._azaraeF68-1
A._azaraeF73-1
A._azaraeF262-1
A._azaraeM264
A._azaraeM271-1
A._azaraeF280-1
A._azaraeF284-1
A._azaraeF538-1
A._mollisM21682-51
A._mollisM21682-54
A._boliviensisM55-71
A._boliviensisM55-72
A._boliviensisF13321-81
A._boliviensisF13321-82
A._kofordi205Male1
A._kofordiF202-41
A._kofordiF202-42
A._kofordi205Male2
A._subfuscusM654-31
A._subfuscus13983Female1
A._subfuscus13983Female2
A._torquesM771-61
A._torquesM771-62
A._torquesFemale1
A._torquesFemale2
A._variusM278-13
A._varius21718Female2

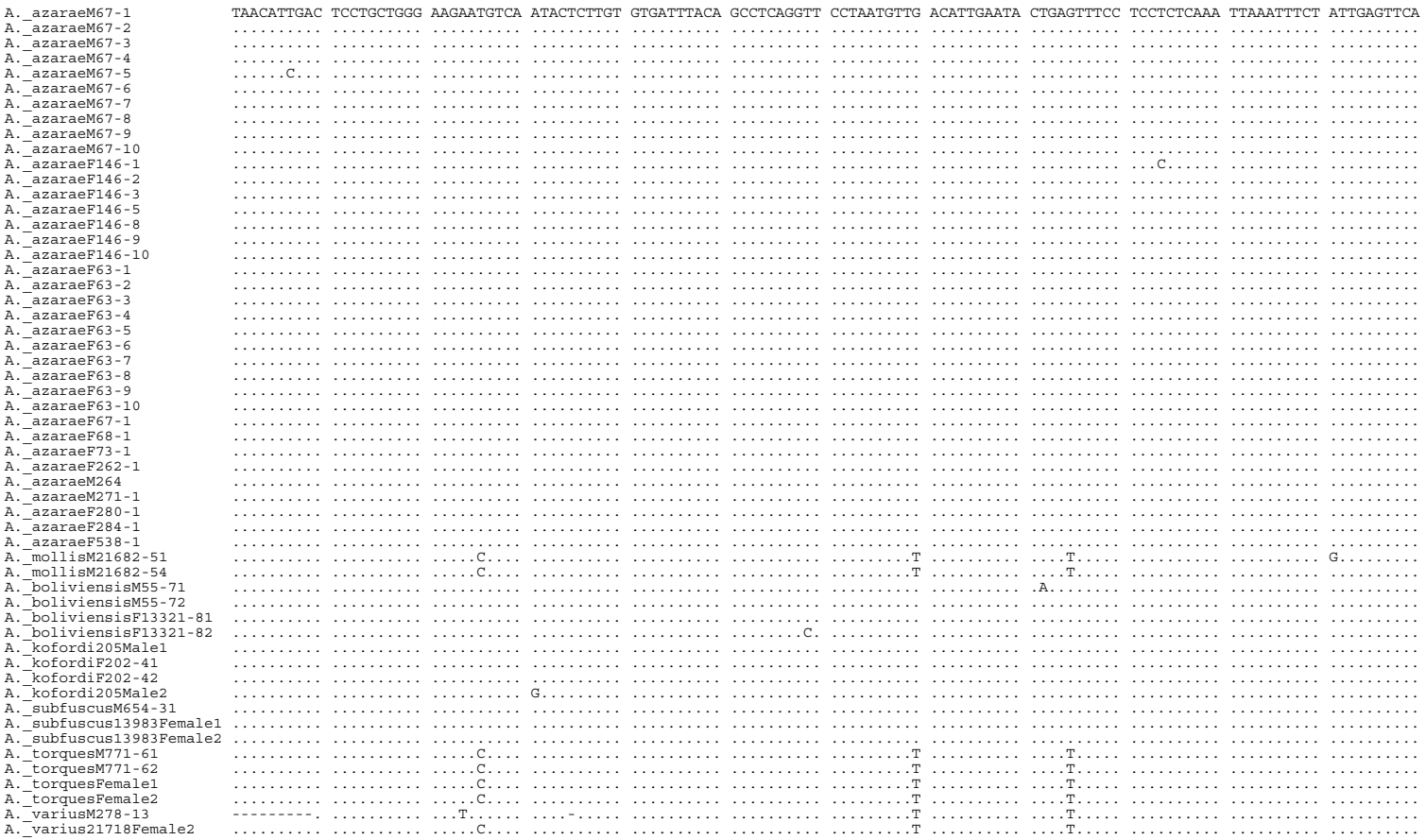
```
A._azaraeM67-1          GATTACTTTA ATTCAGCCTT ACAGCCACAG AATATCTTCG CCTCTGGAGA AAAGACATGC TTTGGGGCTG GCGACAATCA TATAAAGGGC ATCGATGGGC ACATCAAACG CCCCATGAAT 840
A._azaraeM67-2          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeM67-3          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeM67-4          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeM67-5          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeM67-6          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeM67-7          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeM67-8          .......... .......... ...A...... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeM67-9          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeM67-10         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF146-1         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF146-2         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF146-3         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF146-5         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF146-8         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF146-9         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF146-10        .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF63-1          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF63-2          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF63-3          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF63-4          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF63-5          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF63-6          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF63-7          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF63-8          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF63-9          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF63-10         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF67-1          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... T.........
A._azaraeF68-1          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF73-1          .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF262-1         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeM264           .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeM271-1         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF280-1         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF284-1         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._azaraeF538-1         .......... .......... .......... .......... .......... .......... ......A... .......... .......... .......... .......... ..........
A._mollisM21682-51      .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._mollisM21682-54      .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._boliviensisM55-71    .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._boliviensisM55-72    .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._boliviensisF13321-81 .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._boliviensisF13321-82 .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._kofordi205Male1      .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._kofordiF202-41       .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._kofordiF202-42       .......A.. .......... .......... .......... .......... .......... ......A... ...T..G... .......... ..A....... .......... ..........
A._kofordi205Male2      .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._subfuscusM654-31     .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._subfuscus13983Female1 .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._subfuscus13983Female2 .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._torquesM771-61       .......A.. .......... .......... T......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._torquesM771-62       .......A.. .......... .......... T......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._torquesFemale1       .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._torquesFemale2       .......A.. .......... .......... .......... .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
A._variusM278-13        .......A.. .......... .......... .......... .......... ....G..C.. .C....A... ...T...... .G........ ..A....... .......... ..........
A._varius21718Female2   .......A.. .......... .......... .......C.. .......... .......... ......A... ...T...... .......... ..A....... .......... ..........
```

```
A._azaraeM67-1           GCATTTATGG TGTGGTCTCG TGGTCAGAGG CGCAAGTTGG CTCTGGAGAA TCCCGGCATG CAAAATTCTG AGATCAGCAA GCAACTGGGA TGCCAGTGGA AAAGCCTTAC AGAAACTGAC 960
A._azaraeM67-2           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-3           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-4           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-5           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-6           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-7           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-8           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-9           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-10          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._azaraeF146-2          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._azaraeF146-3          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._azaraeF146-5          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-8          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._azaraeF146-9          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-10         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._azaraeF63-1           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-2           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-3           .......... .......C.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-4           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-5           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-6           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-7           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-8           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-9           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-10          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF67-1           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF68-1           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF73-1           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._azaraeF262-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._azaraeM264            .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._azaraeM271-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._azaraeF280-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF284-1          .......... .......... .........  .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._azaraeF538-1          .......... .......... .......C.. .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._mollisM21682-51       .......... .......... .......... .......... .......... .......... .....CGGC. .......... .......... .......... .......... ....G.....
A._mollisM21682-54       .......... .......... .......... .......... .......... .......... .....CGGC. .......... .......... .......... .......... ....G.....
A._boliviensisM55-71     .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._boliviensisM55-72     .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._boliviensisF13321-81  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._boliviensisF13321-82  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._kofordi205Male1       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._kofordiF202-41        .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._kofordiF202-42        .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._kofordi205Male2       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._subfuscusM654-31      .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..G....... ....G.....
A._subfuscus13983Female1 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._subfuscus13983Female2 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ....G.....
A._torquesM771-61        .......... .......... .......... .......... .......... .......... .....CGGC. .......... .......... .......... .......... ....G.....
A._torquesM771-62        .......... .......... .......... .......... .......... .......... .....CGGC. .......... .......... .......... .......... ....G.....
A._torquesFemale1        .......... .......... .......... .......... .......... .......... .....CGGC. .......... .......... .......... .......... ....G.....
A._torquesFemale2        .......... .......... .......... .......... .C........ .......... .....CGGC. .......... .......... .......... .......... ....G.....
A._variusM278-13         .......... .......... .........T .......... .......... .......... .....CG.C. .......... .......... .......... .......... ....G.....
A._varius21718Female2    .......... .......... .......... .......... .......... .......... ......GGC. .......... .......... .......... ...A...... ....G.....
```

```
A._azaraeM67-1          AAAAGGCCAT TTTTCGAAGA GGCACAGAGA CTGAAGAATC TACACAAAGA GAAATATCCA AACTATAAGT ATCAACCTCA CCGGAGAGCT AAAGTGCCAC AGAGGACTGA CCCTTTGCTG 980
A._azaraeM67-2          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-3          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-4          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-5          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-6          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-7          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-8          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-9          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM67-10         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-1         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-2         ..........C......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-3         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-5         .......... .......... .......... .......... .......... ......C... .......... .......... .......... .......... .......... ..........
A._azaraeF146-8         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-9         ..G....... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF146-10        .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-1          ..G....... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-2          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-3          .......... .......... .......... .......... .......... ......C... .......... .......... .......... .......... .......... ..........
A._azaraeF63-4          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-5          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-6          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-7          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-8          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-9          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF63-10         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF67-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF68-1          ..G....... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF73-1          .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF262-1         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM264           .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeM271-1         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF280-1         .......... .......... .......... .......... .....C.... ......C... .......... .......... .......... .......... .......... ..........
A._azaraeF284-1         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._azaraeF538-1         .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._mollisM21682-51      .......... .....A.... .......... .......... .......... .......... .......... .......... .......G T........ .......... ..........
A._mollisM21682-54      .......... .....A.... .......... .......... .......... .......... .......... .......... .......G T........ .......... ..........
A._boliviensisM55-71    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._boliviensisM55-72    .......... ..C....... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._boliviensisF13321-81 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._boliviensisF13321-82 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._kofordi205Male1      .......... .......... .......... ....G..... .......... .......... .......... .......... .......... .......... .......... ..........
A._kofordiF202-41       .......... .......... .......... ....G..... .......... .......... .......... .......... .......... .......... .......... ..........
A._kofordiF202-42       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .........G ..........
A._kofordi205Male2      .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
A._subfuscusM654-31     .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .........C ..........
A._subfuscus13983Female1 .......... .......... .......... .......... .......--.. .......... .......... .......... .......... G......... .........C ..........
A._subfuscus13983Female2 .......... .......... .......... .......... .......--.. .......... .......... .......... .......... .......... .........C ..........
A._torquesM771-61       .......... .......... .......... .......... .......... .......... .......... .......... .......G T........ .......... .......... ..........
A._torquesM771-62       .......... .......... .......... .......... .......... .......... .......... .......... .......G T........ .......... .......... ..........
A._torquesFemale1       .......... .......G.. .......... .......... .......... .......... .......... .......... .......G T........ .......... .......... ..........
A._torquesFemale2       .......... .......G.. .......... .......... .......... .......... .......... .......... .......G T........ .......... .......... ..........
A._variusM278-13        .......... .......... .....T.... ....G..... .......... ..CG...... .......... .......... .T........ .......... .......... ..........
A._varius21718Female2   .......... .....A.... .......... .......... .......... .......... .......... .......... T......... .......... .......... ..........
```


A._azaraeM67-1 CCTGCAGATG CCTCTTCAAA AGGGGAAGAA ACCCTGTGCA CATTCCTATA CACAGAGGAC GGGGCTAGGT CTGCCCATTT GTCGTCCAAG AGCCAGCTAA GCTGTTTACA GCCTGTGGAC 1000
A._azaraeM67-2
A._azaraeM67-3
A._azaraeM67-4
A._azaraeM67-5
A._azaraeM67-6
A._azaraeM67-7
A._azaraeM67-8
A._azaraeM67-9
A._azaraeM67-10
A._azaraeF146-1
A._azaraeF146-2
A._azaraeF146-3
A._azaraeF146-5
A._azaraeF146-8
A._azaraeF146-9
A._azaraeF146-10
A._azaraeF63-1
A._azaraeF63-2
A._azaraeF63-3
A._azaraeF63-4
A._azaraeF63-5
A._azaraeF63-6
A._azaraeF63-7
A._azaraeF63-8
A._azaraeF63-9
A._azaraeF63-10
A._azaraeF67-1
A._azaraeF68-1
A._azaraeF73-1
A._azaraeF262-1
A._azaraeM264
A._azaraeM271-1
A._azaraeF280-1
A._azaraeF284-1
A._azaraeF538-1
A._mollisM21682-51
A._mollisM21682-54
A._boliviensisM55-71
A._boliviensisM55-72
A._boliviensisF13321-81
A._boliviensisF13321-82
A._kofordi205Male1
A._kofordiF202-41
A._kofordiF202-42
A._kofordi205Male2
A._subfuscusM654-31
A._subfuscus13983Female1
A._subfuscus13983Female2
A._torquesM771-61
A._torquesM771-62
A._torquesFemale1
A._torquesFemale2
A._variusM278-13
A._varius21718Female2


A._azaraeM67-1 ATTCCAACTG AGCACTCGGT ACAGCGGCCG CAGCAGCAAT AACACTTTAG CACCAACTGA CTTCACCAGT GAACACAGAG CATAGCAGCA GTGCCTCAGC AAAACCACGG GGCAGTGTCA 1120
A._azaraeM67-2
A._azaraeM67-3
A._azaraeM67-4
A._azaraeM67-5
A._azaraeM67-6
A._azaraeM67-7
A._azaraeM67-8
A._azaraeM67-9
A._azaraeM67-10
A._azaraeF146-1
A._azaraeF146-2
A._azaraeF146-3
A._azaraeF146-5
A._azaraeF146-8
A._azaraeF146-9
A._azaraeF146-10
A._azaraeF63-1
A._azaraeF63-2
A._azaraeF63-3
A._azaraeF63-4
A._azaraeF63-5
A._azaraeF63-6
A._azaraeF63-7
A._azaraeF63-8
A._azaraeF63-9
A._azaraeF63-10
A._azaraeF67-1
A._azaraeF68-1
A._azaraeF73-1
A._azaraeF262-1
A._azaraeM264
A._azaraeM271-1
A._azaraeF280-1
A._azaraeF284-1
A._azaraeF538-1
A._mollisM21682-51
A._mollisM21682-54
A._boliviensisM55-71
A._boliviensisM55-72
A._boliviensisF13321-81
A._boliviensisF13321-82
A._kofordi205Male1
A._kofordiF202-41
A._kofordiF202-42
A._kofordi205Male2
A._subfuscusM654-31
A._subfuscus13983Female1
A._subfuscus13983Female2
A._torquesM771-61
A._torquesM771-62
A._torquesFemale1
A._torquesFemale2
A._variusM278-13
A._varius21718Female2


A._azaraeM67-1 TAACATTGAC TCCTGCTGGG AAGAATGTCA ATACTCTTGT GTGATTTACA GCCTCAGGTT CCTAATGTTG ACATTGAATA CTGAGTTTCC TCCTCTCAAA TTAAATTTCT ATTGAGTTCA 1240
A._azaraeM67-2
A._azaraeM67-3
A._azaraeM67-4
A._azaraeM67-5 ......C...
A._azaraeM67-6
A._azaraeM67-7
A._azaraeM67-8
A._azaraeM67-9
A._azaraeM67-10
A._azaraeF146-1 ...C.....
A._azaraeF146-2
A._azaraeF146-3
A._azaraeF146-5
A._azaraeF146-8
A._azaraeF146-9
A._azaraeF146-10
A._azaraeF63-1
A._azaraeF63-2
A._azaraeF63-3
A._azaraeF63-4
A._azaraeF63-5
A._azaraeF63-6
A._azaraeF63-7
A._azaraeF63-8
A._azaraeF63-9
A._azaraeF63-10
A._azaraeF67-1
A._azaraeF68-1
A._azaraeF73-1
A._azaraeF262-1
A._azaraeM264
A._azaraeM271-1
A._azaraeF280-1
A._azaraeF284-1
A._azaraeF538-1
A._mollisM21682-51 .....C.... .....T..... ....T..... G.....
A._mollisM21682-54 .....C.... .....T..... ....T.....
A._boliviensisM55-71 .....A.....
A._boliviensisM55-72
A._boliviensisF13321-81
A._boliviensisF13321-82 .....C.....
A._kofordi205Male1
A._kofordiF202-41
A._kofordiF202-42
A._kofordi205Male2 .....G.....
A._subfuscusM654-31
A._subfuscus13983Female1
A._subfuscus13983Female2
A._torquesM771-61 .....C.... .....T..... ....T.....
A._torquesM771-62 .....C.... .....T..... ....T.....
A._torquesFemale1 .....C.... .....T..... ....T.....
A._torquesFemale2 .....C.... .....T..... ....T.....
A._variusM278-13 ----------. ...T..... ....-..... .....T..... ....T.....
A._varius21718Female2 .....C.... .....T..... ....T.....

```
A._azaraeM67-1           AAAATGAGGT ATGTAAAGAT GTGTTGTGTA TGAGAAAGAA GAATGGTTGT AG-GCTTTT- ACCTAC-- 1508
A._azaraeM67-2           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeM67-3           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeM67-4           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeM67-5           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeM67-6           .......... .......... .......... .......... .......... ..T......- ....C---
A._azaraeM67-7           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeM67-8           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeM67-9           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeM67-10          .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF146-1          .......... .......... .......... .......... .......... ..T......- .....AC-
A._azaraeF146-2          .......... .......... .......... .......... .......... ..T......- ...C..--
A._azaraeF146-3          .......... .......... .......... .......... .......... ..T.....A- C.TAC---
A._azaraeF146-5          .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF146-8          .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF146-9          .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF146-10         .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF63-1           .......... .......... .......... .......... .......... ..T......- ......C-
A._azaraeF63-2           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF63-3           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF63-4           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF63-5           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF63-6           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF63-7           .......... .......... .......... .......... .......... ..T.....-- ......--
A._azaraeF63-8           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF63-9           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF63-10          .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF67-1           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF68-1           .......... .......... .......... .......... .......... ..T......T ......--
A._azaraeF73-1           .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF262-1          .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeM264            .......... .......... .......... .......... .......... ..T......- ......C-
A._azaraeM271-1          .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF280-1          .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF284-1          .......... .......... .......... .......... .......... ..T......- ......--
A._azaraeF538-1          .......... ..A....... .......... .......... .......... ..T......- ......--
A._mollisM21682-51       .......... G......... .......... .......... .......... ..T......- ......--
A._mollisM21682-54       .......... G......... .......... .......... .......... ..T......- ......--
A._boliviensisM55-71     .......... .......... ....-----. .......... .......... ..T......- ......--
A._boliviensisM55-72     .......... .......... ....-----. .......... .....-.... ..T......- ......--
A._boliviensisF13321-81  .......... .......... ....-----. .......... .......... ..T......- ......--
A._boliviensisF13321-82  .......... .......... ....-----. .......... .......... ..T......- ......C-
A._kofordi205Male1       .......... .......... .......... .......... .......... ..T......- ......--
A._kofordiF202-41        .......... .......... .......... .......... .......... ..T......- ......--
A._kofordiF202-42        .......... .......... .......... .......... .......... ..T......- ......--
A._kofordi205Male2       .......... .......... .......... .......... .......... ..-......- ......--
A._subfuscusM654-31      .......... .......... .......... .......... .......... ..T......- ......--
A._subfuscus13983Female1 .......... .......... .......... .......... .......... ..T......- ......--
A._subfuscus13983Female2 .......... .......... .......... .......... .......... ..T......- ......--
A._torquesM771-61        .......... G......... .......... .......... .......... ..T......- ......--
A._torquesM771-62        .......... G......... .......... .......... .......... ..T......- ......--
A._torquesFemale1        .......... G......... .......... .......... .......... ..T......- C.....--
A._torquesFemale2        .......... G......... .......... .......... .......... ..T......- ......--
A._variusM278-13         .......... G......... .......... ....---... .......... ..T......- ......--
A._varius21718Female2    .......... G......... .......... .......... .......... ..T......- ......--
```

APPENDIX C

Amino acid sequence alignment of *Sry* from *Akodon azarae* male and female and taxa representing 4 other major mammalian orders. This alignment was generated using CLUSTALW with default parameters. The dots represent amino acid identical to the first sequence and dash represents alignment gap. Highlighted region corresponds to the High Mobility Group domain.

```
A._azaraeM67-1          MFSTLN-HDY FNSALQPQNI FASGEKTCFG ---------- --AGDNHIKG IDGHIKRPMN AFMVWSRGQR RKLALENPGM QNSEISKQLG CQWKSLTETD KRPFFEEAQR 110
A._azaraeF146-1         ......-... .......... .......... ---------- --T....... .......... .......... .......... .......... ........A. .......... 
Human                   .L.VF.SD.. -SP.V.E-.. P.LRRSSS.L CTESCNSKYQ CET.E.SKGN VQDRV..... ..I....D.. ..M.....R. R......... Y...M...AE .W...Q...K
Chimpanzee              .L.VF.SD.. -SP.V.Q-.. P.LRRSSS.L CTESYNSKYQ RET.E.SKDS VQDRV..... ..F....D.. ..M.....R. R......... Y...M...AE .W...Q...K
Rat                     ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. ....QQ..S. .......H.. Y.......AE .....Q....
Target_Rat              ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. ....QK..S. ..T....... YR......AE .....Q....
Rattus_exulans          ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. ....QQ..S. .......... Y.......AE .....Q....
Mus_m._domesticus       ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. H...QQ..S. ..T....... .R......AE .....Q....
Mus_m._musculus         ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. H...QQ..S. ..T....... .R......AE .....Q....
Mus_spretus             ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. H...QQ..S. ..T....... .R......AE .....Q....
Mus_spicilegus          ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. H...QQ..S. ..T....... .R......AE .....Q....
Mus_cookii              ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. H...QQ..S. ..T....... .R......DE .....Q....
Mus_caroli              ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. H...QQ..S. ..T....... .R......AE .....Q....
Mus_cervicolor          ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. H...QQ..S. ..T....... .R......DE .....Q....
Mus_macedonicus         ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. H...QQ..S. ..T....... .R......AE .....Q....
Mus_minutoides          ---------- ---------- ---------- ---------- ---------- ME..V..... ......C.E. Q...QQ..SL ..T....L.. .R......DE .....Q....
Mus_pahari              ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. H...QQ..S. ..T....... .R......AE .....Q....
Mastomys_hildebrandti   ---------- ---------- ---------- ---------- ---------- ME..V..... ...A....E. ....QQ..S. ..T....... YR......AE .....Q....
Holomyscus_alleni       ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. ....QQ..S. ..T....... YR......AE .....Q....
Stochomys_longicaudatus ---------- ---------- ---------- ---------- ---------- ME..V..... ........E. ....QK..S. ..T....... YR......AE .....Q....
Cow                     ..RV..DDV. SPAVV.Q.TT L.FRKDSSLC -TDSHSANDQ CER.EHVRES SQD.V..... ..I....ER. ..V.....K. K..D...... YE..R..DAE ..........
Pig                     ..RV.KAD.. SPA.-.Q... L.L.KGSSLF PTDNHSSNDG RETRGSGRES GQDRV..... ..I....D.. ..V.....Q. .......W.. .K..M...AE ..........
Rabbit                  ..GA.YSDA. TPAK-.QS.S ..L------- ----VSTNHQ CNT.GTRKVS GQERV..... ......QH.. .QV..Q..K. R..D...... H...M.S.AE .W...Q....
Brown_Bear              ..GV..SD.H -CA.V.QR.. L.F.RTFSEF WMNNPTSNYR CETEG.SRDS GQNRVR.... ...L...D.. ..V.....Q. .......... Y..EM...AE .W........

A._azaraeM67-1          LKNLHKEKYP NYKYQPHRRA KVPQRTDPLL PADASSKG-- ----EETLCT FLYT-EDGAR SAHLSSKSQL S-CLQPVDIP TEHSVQRPQQ Q*-------- ------ 216
A._azaraeF146-1         .......... .......... .......... ---------- ----..S... ....-..... .......... .-........ .......... ..-------- ------
Human                   .QAM.R.... ....R.R.K. .MLPKNCS.. ...PA.VLCS EVQLDNR.Y- ---R-D.CTK AT.SRMEH.. GHL-P.INAA SSPQQRDRYS HWTKL*---- ------
Chimpanzee              .QAM.R.... ....R.R.K. NMLPKNCS.. ...PA.VLCS EVQLDNR.Y- ---R-D.CTK AT.SRMEH.. GHL-P.INAA SSPQQRDRYS HWTKL*---- ------
Rat                     ..T..R.... .........V .....SYT.Q REV..T.LYN LLQWDNN.H. II.G-Q.W.. A..Q...N.K .IY....... .GYPL.QK.. ---------- ------
Target_Rat              ..T..R.... .......... .....SGT.K .TA..T.LYN LLQWDMNPH. ITNR-QVR.. A......N.Q .FY..LMN.. .G.LLLQW.. R--------- ------
Rattus_exulans          ..T..R.... .........V .....SYT.Q REV..T.LYN LLQWDNN.H. II.G-Q.W.. A..R...N.K .IY....... .GYP------ ---------- ------
Mus_m._domesticus       ..T..R.... .......... ..S..SGI.Q ..V..T.LYN LLQWDRNPHA IT.R-Q.WS. A...Y..N.Q .FY....... .G.PQ.QQ.. .---QQFHNH HQQQQQ
Mus_m._musculus         ..T..R.... .......... ..S..SGI.Q ..V..T.LYN LLQWDRNPHA IT.R-Q.WS. A...Y..N.Q .FYW...... .G.LQ.QQ.. .QQQQQFHNH HQQQQQ
Mus_spretus             ..T..R.... .......... ..S..SGI.Q .RV..T.LYN LLQWDRNPHA IT.R-Q.WS. A...Y..N.Q .FY....... .G.P------ ---------- ------
Mus_spicilegus          ..T..R.... .......... ..S..SGI.Q .RV..T.LYN LLQWDRNPHA IT.R-Q.WS. A...Y..N.Q .FY....... .G.PQ.QQ.. .-----FHNH HQQKQQ
Mus_cookii              ..T..R.... .......... .....SGT.Q ..V..T.LYN LLQWDRNPHA IT.R-Q.WS. A......N.Q .FY.....N. .G.P------ ---------- ------
Mus_caroli              ..T..R.... .......... .....SGT.Q ..V..T.LYN LLQWDRNPHA IT.R-Q.WS. A......N.Q .FY....... .GQP------ ---------- ------
Mus_cervicolor          ..T..R.... .......... .....SGT.Q ..V..T.LYN LQQWDRNPHA IT.R-Q.WS. A....F.N.Q .FY....... .G.P------ ---------- ------
Mus_macedonicus         ..T..R.... .......... ..S..SGI.Q .RV..T.LYN LLQWDRNPHA IT.R-Q.WS. A...Y..N.Q .FY....... .G.P------ ---------- ------
Mus_minutoides          ..T..R.... ......R... .....SGT.Q ..VT.T.LYN LLQW------ ---------- ---------- ---------- ---------- ---------- ------
Mus_pahari              ..T..R.... ........K. .....SGA.Q R.V..TNLYN LLQWDR.PH. IT.R-Q.WT. A......N.Q .FY.RS.G.. .G.P------ ---------- ------
Mastomys_hildebrandti   ..T..R.... ........K. ...E.SGT.Q ..V..T.LYN LLQCDRNPHA IP.R-QEW.. A......N.Q .FY....... HWAPT----- ---------- ------
Holomyscus_alleni       ..T..R.... ........KV ...E.SGT.Q .MV..T.LYN LLQWDRNPHA IT.R-Q.W.. T..Q...NHQ .FY......S HWVPT----- ---------- ------
Stochomys_longicaudatus ..T..R.... .......... .....SGT.K .TA..T.LYN LLQWDMNPH. ITNR-QVR.. A......N.Q .FY..LMN.. .G.LLLQW.. R--------- ------
Cow                     .LAI.RD... G...R.R... ---K.PQKS. ...-..ILCN PMHV.T-.HP .T.R-DGC.K TTYSQME... .RS-.S.I.T NSLLQKEHHS SWTSLGHNKV TLATRI
Pig                     .QAV.RD... G...R.R.KG ---E.AQN.. ..E-AAVLCS QVRV..RMYP .T---YTV.K AKCSGTE... .HS-..MN.T SSLLQ.EDRC NWTGLCTVG* ------
Rabbit                  .QAM..... D...R.R.KV .IL.KS.S.. L.QPT.TLCS EVHMD.G.Y- ---------- -TCTGM.E.. ICS-...NTG SSLPQ.QCHS NWTSWQENRV TLAAQT
Brown_Bear              .QAM.RQ... D...R.R.K. -T..KD.K.. .SAS..TLCR QVRVD..WYP .T.R-NSHT. A..SGMED.. .SS-...NVA SSLLQ.EQHC SSTSFRDSRE TLATQL
```

APPENDIX D

Information for *Akodon azarae* tissue samples.

Species	Sex	UMMZ* ID	S	W	Country	Region	Date
Akodon azarae	female	GD063	26*35.511'	58*08.329'	Paraguay	Neembucu	15/9/1999
Akodon azarae	male	GD067	26*35.511'	58*08.329'	Paraguay	Neembucu	26/9/1999
Akodon azarae	female	GD068	26*36.014'	58*09.131'	Paraguay	Neembucu	26/9/1999
Akodon azarae	female	GD069	26*36.014'	58*09.131'	Paraguay	Neembucu	27/9/1999
Akodon azarae	female	GD073	26*34.718'	58*08.220'	Paraguay	Neembucu	27/9/1999
Akodon azarae	female	GD076	26*34.571'	58*08.370'	Paraguay	Neembucu	27/9/1999
Akodon azarae	female	GD079	26*34.971'	58*08.673'	Paraguay	Neembucu	30/9/1999
Akodon azarae	female	GD140	26*35.070'	58*08.324'	Paraguay	Neembucu	2/11/1999
Akodon azarae	female	GD142	26*35.070'	58*08.324'	Paraguay	Neembucu	3/11/1999
Akodon azarae	female	GD146	26*35.070'	58*08.324'	Paraguay	Neembucu	5/11/1999
Akodon azarae	female	GD262	26*24.050'	57*02.340'	Paraguay	Paraquari	15/12/1999
Akodon azarae	male	GD264	26*24.050'	57*02.340'	Paraguay	Paraquari	15/12/1999
Akodon azarae	male	GD271	26*24.615'	57*02.461'	Paraguay	Paraquari	15/12/1999
Akodon azarae	female	GD280	26*24.464	57*02.752'	Paraguay	Paraquari	16/12/1999
Akodon azarae	female	GD282	26*24.465	57*02.752'	Paraguay	Paraquari	16/12/1999
Akodon azarae	female	GD283	26*24.466	57*02.752'	Paraguay	Paraquari	16/12/1999
Akodon azarae	female	GD284	26*24.467	57*02.752'	Paraguay	Paraquari	16/12/1999
Akodon azarae	female	GD298	26*24.464'	57*02.752'	Paraguay	Paraquari	17/12/1999
Akodon azarae	female	GD299	26*24.464'	57*02.752'	Paraguay	Paraquari	17/12/1999
Akodon azarae	male	GD300	26*24.464'	57*02.752'	Paraguay	Paraquari	18/12/1999
Akodon azarae	female	GD308	26*.24.050'	57*02.340'	Paraguay	Paraquari	18/12/1999
Akodon azarae	female	GD538	26*30.816'	57*14.444'	Paraguay	Paraquari	22/5/2001
Akodon azarae	female	GD553	26*30.816'	57*14.444'	Paraguay	Paraquari	24/5/2001
Akodon azarae	female	GD562	26*30.598'	57*14.000'	Paraguay	Misiones	26/5/2001

* University of Michigan Museum of Zoology

At the time of this study, above *A. azarae* tissue samples were not catagued yet. ID numbers listed here correspond to frozen tissue samples.

LITERATURE CITED

Adkins, R. M., E. L. Gelke, D. Rowe, and R. L. Honeycutt. 2001. Molecular phylogeny and divergence time estimates for major rodent groups: Evidence from multiple genes. Molecular Biology and Evolution **18**:777-791.

Assumpcao, J. G., L. F. C. Ferraz, C. E. Benedetti, A. T. Maciel-Guerra, G. Guerra, A. P. Marques-de-Faria, T. M. Baptista, and M. P. de Mello. 2005. A naturally occurring deletion in the SRY promoter region affecting the Sp1 binding site is associated with sex reversal. Journal of Endocrinological Investigation **28**:651-656.

Balzar, M., I. H. Briaire-de Bruijn, H. A. Rees-Bakker, F. A. Prins, W. Helfrich, L. de Leij, G. Riethmuller, S. Alberti, S. O. Warnaar, G. J. Fleuren, and S. V. Litvinov. 2001. Epidermal growth factor-like repeats mediate lateral and reciprocal interactions of Ep-CAM molecules in homophilic adhesions. Mol Cell Biol **21**:2570-2580.

Barbara, P. D., C. Mejean, B. Moniot, M. H. Malcles, P. Berta, and B. Boizet-Bonhoure. 2001. Steroidogenic factor-1 contributes to the cyclic-adenosine monophosphate down-regulation of human SRY gene expression. Biology of Reproduction **64**:775-783.

Berta, P., J. R. Hawkins, a. H. Sinclair, a. Taylor, B. L. Griffiths, P. N. Goodfellow, and M. Fellous. 1990. Genetic-Evidence Equating Sry and the Testis-Determining Factor. Nature **348**:448-450.

Bezanilla, F., and C. M. Armstrong. 1977a. Inactivation of the sodium channel. I. Sodium current experiments. J Gen Physiol **70**:549-566.

Bezanilla, F., and C. M. Armstrong. 1977b. Inactivation of Sodium Channel.1. Sodium Current Experiments. Journal of General Physiology **70**:549-566.

Bianchi, N. O. 2002. Akodon sex reversed females: the never ending story. Cytogenetic and Genome Research **96**:60-65.

Bianchi, N. O., M. S. Bianchi, G. Bailliet, and a. Delachapelle. 1993. Characterization and Sequencing of the Sex-Determining Region-Y Gene (Sry) in Akodon (Cricetidae) Species with Sex Reversed Females. Chromosoma **102**:389-395.

Bianchi, N. O., and Contrera.Jr. 1967. Chromosomes of Field Mouse Akodon Azarae (Cricetidae Rodentia) with Special Reference to Sex Chromosome Anomalies. Cytogenetics **6**:306-&.

Bianchi, N. O., and S. Merani. 1984. Cytogenetics of South-American Akodont Rodents (Cricetidae).10. Karyological Distances at Generic and Intergeneric Levels. Journal of Mammalogy **65**:206-219.

Bianchi, N. O., O. a. Reig, O. J. Molina, and F. N. Dulout. 1971. Cytogenetics of South American Akodont Rodents (Cricetidae).1. Progress Report of Argentinian and Venezuelan Forms. Evolution **25**:724-&.

Birkhead, T. R., J. G. Martinez, T. Burke, and D. P. Froman. 1999. Sperm mobility determines the outcome of sperm competition in the domestic fowl. Proc R Soc Lond B Biol Sci **266**:1759-1764.

Bishop, C. E., D. J. Whitworth, Y. J. Qin, A. I. Agoulnik, I. U. Agoulnik, W. R. Harrison, R. R. Behringer, and P. A. Overbeek. 2000. A transgenic insertion

upstream of Sox9 is associated with dominant XX sex reversal in the mouse. Nature Genetics **26**:490-494.

Bor, Y. C., J. Swartz, A. Morrison, D. Rekosh, M. Ladomery, and M. L. Hammarskjold. 2006. The Wilms' tumor 1 (WT1) gene (+KTS isoform) functions with a CTE to enhance translation from an unspliced RNA with a retained intron. Genes & Development **20**:1597-1608.

Brennan, J., and B. Capel. 2004. One tissue, two fates: Molecular genetic events that underlie testis versus ovary development. Nature Reviews Genetics **5**:509-521.

Britten, R. J. 2002a. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. Proceedings of the National Academy of Sciences of the United States of America **99**:13633-13635.

Britten, R. J. 2002b. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. Proc Natl Acad Sci U S A **99**:13633-13635.

Brunet, M., F. Guy, D. Pilbeam, H. T. Mackaye, A. Likius, D. Ahounta, A. Beauvilain, C. Blondel, H. Bocherens, J. R. Boisserie, L. De Bonis, Y. Coppens, J. Dejax, C. Denys, P. Duringer, V. R. Eisenmann, G. Fanone, P. Fronty, D. Geraads, T. Lehmann, F. Lihoreau, A. Louchart, A. Mahamat, G. Merceron, G. Mouchelin, O. Otero, P. P. Campomanes, M. P. De Leon, J. C. Rage, M. Sapanet, M. Schuster, J. Sudre, P. Tassy, X. Valentin, P. Vignaud, L. Viriot, A. Zazzo, and C. Zollikofer. 2002. A new hominid from the Upper Miocene of Chad, central Africa. Nature **418**:145-151.

Burgos, M., R. Jimenez, D. M. Olmos, and R. D. Delaguardia. 1988. Heterogeneous Heterochromatin and Size Variation in the Sex-Chromosomes of Microtus-Cabrerae. Cytogenetics and Cell Genetics **47**:75-79.

Camerino, G., P. Parma, O. Radi, and S. Valentini. 2006. Sex determination and sex reversal. Current Opinion in Genetics & Development **16**:289-292.

Carlson, A. E., R. E. Westenbroek, T. Quill, D. J. Ren, D. E. Clapham, B. Hille, D. L. Garbers, and D. F. Babcock. 2003. CatSper1 required for evoked Ca2+ entry and control of flagellar function in sperm. Proceedings of the National Academy of Sciences of the United States of America **100**:14864-14868.

Civetta, a., and R. S. Singh. 1995. High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in Drosophila melanogaster and Drosophila virilis group species. Journal of Molecular Evolution **41**:1085-1095.

Cooper, G. M., M. Brudno, E. A. Stone, I. Dubchak, S. Batzoglou, and A. Sidow. 2004. Characterization of evolutionary rates and constraints in three mammalian genomes. Genome Research **14**:539-548.

Courtois, S. J., D. a. Lafontaine, F. P. Lemaigre, S. M. Durviaux, and G. G. Rousseau. 1990. Nuclear Factor-I and Activator Protein-2 Bind in a Mutually Exclusive Way to Overlapping Promoter Sequences and Trans-Activate the Human Growth-Hormone Gene. Nucleic Acids Research **18**:57-64.

D'Elia. 2003. Phylogenetics of Sigmodontinae (Rodentia, Muroidea, Cricetidae), with special reference to the akodont group, and with additional comments on historical biogeography. Cladistics:307-323.

Ensslin, M. A., and B. D. Shur. 2003. Identification of mouse sperm SED1, a bimotif EGF repeat and discoidin-domain protein involved in sperm-egg binding. Cell **114**:405-417.

Espinosa, M. B., and a. D. Vitullo. 1996. Offspring sex-ratio and reproductive performance in heterogametic females of the south American field mouse Akodon azarae - Reproduction in heterogametic Akodon azarae females. Hereditas **124**:57-62.

Fernandez, R., M. J. L. Barragan, M. Bullejos, J. A. Marchal, S. Martinez, R. D. de la Guardia, and A. Sanchez. 2002. Mapping the SRY gene in Microtus cabrerae: a vole species with multiple SRY copies in males and females. Genome **45**:600-603.

Fiddler, M., B. Abdelrahman, D. a. Rappolee, and E. Pergament. 1995. Expression of Sry Transcripts in Preimplantation Human Embryos. American Journal of Medical Genetics **55**:80-84.

Fitch, W. M., J. M. E. Leiter, X. Q. Li, and P. Palese. 1991. Positive Darwinian Evolution in Human Influenza-a Viruses. Proceedings of the National Academy of Sciences of the United States of America **88**:4270-4274.

Fredga, K. 1983. Aberrant Sex-Chromosome Mechanisms in Mammals - Evolutionary Aspects. Differentiation **23**:S23-S30.

Fredga, K., a. Gropp, H. Winking, and F. Frank. 1976. Fertile Xx-Type and Xy-Type Females in Wood Lemming Myopus-Schisticolor. Nature **261**:225-227.

Galindo, B. E., V. D. Vacquier, and W. J. Swanson. 2003. Positive selection in the egg receptor for abalone sperm lysin. Proceedings of the National Academy of Sciences of the United States of America **100**:4639-4643.

Giese, K., J. Pagel, and R. Grosschedl. 1994. Distinct DNA-Binding Properties of the High-Mobility Group Domain of Murine and Human Sry Sex-Determining Factors. Proceedings of the National Academy of Sciences of the United States of America **91**:3368-3372.

Glazko, G. V., and M. Nei. 2003. Estimation of divergence times for major lineages of primate species. Molecular Biology and Evolution **20**:424-434.

Golay, J., L. Basilico, L. Loffarelli, S. Songia, V. Broccoli, and M. Introna. 1996. Regulation of hematopoietic cell proliferation and differentiation by the myb oncogene family of transcription factors. International Journal of Clinical & Laboratory Research **26**:24-32.

Goldberg, A., D. E. Wildman, T. R. Schmidt, M. Huttemann, M. Goodman, M. L. Weiss, and L. I. Grossman. 2003. Adaptive evolution of cytochrome c oxidase subunit VIII in anthropoid primates. Proceedings of the National Academy of Sciences of the United States of America **100**:5873-5878.

Gomendio, M., A. H. Harbourt, and E. R. S. Roldan. 1998. Pp. 667-751 *in* T. R. Birkhead, and A. P. Moller, eds. Sperm Competition and Sexual Selection. Academic Press, London.

Goncalves, P. R., P. Myers, J. F. Vilela, and J. A. de Oliviera. 2006. Systematics of species of the genus Akodon (Rodentia: Sigmodontinae) in southeastern Brasil and implications for the biogeography of the Campos de Altitude. University of Michigan, Micellaneous Publications **MP 197**:1-24.

Gong, Q. H., J. Stern, and a. Dean. 1991. Transcriptional Role of a Conserved Gata-1 Site in the Human Epsilon-Globin Gene Promoter. Molecular and Cellular Biology **11**:2558-2566.

Goodman, M., L. I. Grossman, and D. E. Wildman. 2005. Moving primate genomics beyond the chimpanzee genome. Trends Genet **21**:511-517.

Goodman, M., C. A. Porter, J. Czelusniak, S. L. Page, H. Schneider, J. Shoshani, G. Gunnell, and C. P. Groves. 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Molecular Phylogenetics and Evolution **9**:585-598.

Gubbay, J., J. Collignon, P. Koopman, B. Capel, a. Economou, a. Munsterberg, N. Vivian, P. Goodfellow, and R. Lovellbadge. 1990. A Gene-Mapping to the Sex-Determining Region of the Mouse Y-Chromosome Is a Member of a Novel Family of Embryonically Expressed Genes. Nature **346**:245-250.

Hammes, a., J. K. Guo, G. Lutsch, J. R. Leheste, D. Landrock, U. Ziegler, M. C. Gubler, and A. Schedl. 2001. Two splice variants of the Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. Cell **106**:319-329.

Hancook, J. M. 1999a. Microsatellites and other simple sequences: genomic context and mutational mechanisms. Pp. 1-9 *in* D. B. Goldstein, and C. Schlotterer, eds. Microsatellites, Evolution, and Applications. Oxford Univ. Press, New York.

Hancook, J. M. 1999b. Pp. 1-9 *in* D. B. Goldstein, and C. Schlotterer, eds. Microsatellites, Evolution and Applications. Oxford University Press, New York.

Hannon, R., T. Evans, G. Felsenfeld, and H. Gould. 1991. Structure and Promoter Activity of the Gene for the Erythroid Transcription Factor Gata-1. Proceedings of the National Academy of Sciences of the United States of America **88**:3004-3008.

Harley, V. R., M. J. Clarkson, and A. Argentaro. 2003. The molecular action and regulation of the testis-determining factors, SRY (sex-determining region on the Y chromosome) and SOX9 [SRY-related high-mobility group (HMG) box 9]. Endocrine Reviews **24**:466-487.

Harley, V. R., D. I. Jackson, P. J. Hextall, J. R. Hawkins, G. D. Berkovitz, S. Sockanathan, R. Lovellbadge, and P. N. Goodfellow. 1992. DNA-Binding Activity of Recombinant Sry from Normal Males and Xy Females. Science **255**:453-456.

Hennig, L. 1999. WinGene/WinPep: User-friendly software for the analysis of amino acid sequences. Biotechniques **26**:1170-1172.

Hille, B. 2001. Ion Channels of Excitable Membranes. Sinauer, Sunderland, Mass.

Hoekstra, H. E., and S. V. Edwards. 2000. Multiple origins of XY female mice (genus Akodon): phylogenetic and chromosomal evidence. Proceedings of the Royal Society of London Series B-Biological Sciences **267**:1825-1831.

Hoshi, T., W. N. Zagotta, and R. W. Aldrich. 1990a. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. Science **250**:533-538.

Hoshi, T., W. N. Zagotta, and R. W. Aldrich. 1990b. Biophysical and Molecular Mechanisms of Shaker Potassium Channel Inactivation. Science **250**:533-538.

Hossain, a., and G. F. Saunders. 2001. The human sex-determining gene SRY is a direct target of WT1. Journal of Biological Chemistry **276**:16817-16823.

Hughes, A. L. 1999. Adaptive Evolution of Genes and Genomes. Oxford University Press, New York.

Hughes, A. L., and M. NEI. 1988a. Pattern of Nucleotide Substitution at Major Histocompatibility Complex Class-I Loci Reveals Overdominant Selection. Nature **335**:167-170.

Hughes, A. L., and M. Nei. 1988b. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335**:167-170.

Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature **409**:860-921.

Ishii, M., Y. Kanai, M. Kanai-Azuma, Y. Tajima, T. T. Wei, T. Kidokoro, Y. Sanai, M. Kurohmaru, and Y. Hayashi. 2005. Adhesion activity of fetal gonadal cells to EGF and discoidin domains of milk fat globule-EGF factor 8 (MFG-E8), a secreted integrin-binding protein which is transiently expressed in mouse early gonadogenesis. Anat Embryol (Berl) **209**:485-494.

Jager, R. J., V. R. Harley, R. a. Pfeiffer, P. N. Goodfellow, and G. Scherer. 1992. A Familial Mutation in the Testis-Determining Gene Sry Shared by Both Sexes. Human Genetics **90**:350-355.

Jones, D. H., and S. C. Winistorfer. 1997. Amplification of 4-9-kb human genomic DNA flanking a known site using a panhandle PCR variant. Biotechniques **23**:132-138.

Joseph, S. B., and M. Kirkpatrick. 2004. Haploid selection in animals. Trends in Ecology & Evolution **19**:592-597.

Just, W., W. Rau, W. Vogel, M. Akhverdian, K. Fredga, J. a. M. Graves, and E. Lyapunova. 1995. Absence of Sry in Species of the Vole Ellobius. Nature Genetics **11**:117-118.

Katoh, K., and T. Miyata. 1999. A heuristic approach of maximum likelihood method for inferring phylogenetic tree and an application to the mammalian SOX-3 origin of the testis-determining gene SRY. Febs Letters **463**:129-132.

Koopman, P., J. Gubbay, N. Vivian, P. Goodfellow, and R. Lovellbadge. 1991. Male Development of Chromosomally Female Mice Transgenic for Sry. Nature **351**:117-121.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244-1245.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform **5**:150-163.

Kvistgaard, A. S., L. T. Pallesen, C. F. Arias, S. Lopez, T. E. Petersen, C. W. Heegaard, and J. T. Rasmussen. 2004. Inhibitory effects of human and bovine milk constituents on rotavirus infections. J Dairy Sci **87**:4088-4096.

Kwok, C., C. TylerSmith, B. B. Mendonca, I. Hughes, G. D. Berkovitz, P. N. Goodfellow, and J. R. Hawkins. 1996. Mutation analysis of the 2 kb 5' to SRY in XY females and XY intersex subjects. Journal of Medical Genetics **33**:465-468.

Lawrence, N., T. Klein, K. Brennan, and A. Martinez Arias. 2000. Structural requirements for notch signalling with delta and serrate during the development and patterning of the wing disc of Drosophila. Development **127**:3185-3195.

Lee, Y. H., T. Ota, and V. D. Vacquier. 1995a. Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol Biol Evol **12**:231-238.
Lee, Y. H., T. Ota, and V. D. Vacquier. 1995b. Positive Selection Is a General Phenomenon in the Evolution of Abalone Sperm Lysin. Molecular Biology and Evolution **12**:231-238.
Lemaigre, F. P., S. J. Courtois, S. M. Durviaux, C. J. Egan, D. a. Lafontaine, and G. G. Rousseau. 1989. Analysis of Cis-Acting and Trans-Acting Elements in the Hormone-Sensitive Human Somatotropin Gene Promoter. Journal of Steroid Biochemistry and Molecular Biology **34**:79-83.
Lemaigre, F. P., D. a. Lafontaine, S. J. Courtois, S. M. Durviaux, and G. G. Rousseau. 1990. Sp1 Can Displace Ghf-1 from Its Distal Binding-Site and Stimulate Transcription from the Growth-Hormone Gene Promoter. Molecular and Cellular Biology **10**:1811-1814.
Li, B., N. B. Phillips, A. Jancso-Radek, V. Ittah, R. Singh, D. N. Jones, E. Haas, and M. A. Weiss. 2006. SRY-directed DNA bending and human sex reversal: Reassessment of a clinical mutation uncovers a global coupling between the HMG box and its tail. Journal of Molecular Biology **360**:310-328.
Liu, D. Y., G. N. Clarke, M. Martic, C. Garrett, and H. W. G. Baker. 2001. Frequency of disordered zona pellucida (ZP)-induced acrosome reaction in infertile men with normal semen analysis and normal spermatozoa-ZP binding. Human Reproduction **16**:1185-1190.
Lizarralde, M. S., N. O. Bianchi, and M. S. Merani. 1982. Cytogenetics of South-American Akodont Rodents (Cricetidae).7. Origin of Sex-Chromosome Polymorphism in Akodon-Azarae. Cytologia **47**:183-193.
Lobley, A., V. Pierron, L. Reynolds, L. Allen, and D. Michalovich. 2003. Identification of human and mouse CatSper3 and CatSper4 genes: characterisation of a common interaction domain and evidence for expression in testis. Reprod Biol Endocrinol **1**:53.
Lodish, H., A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell. 2000. Molecular Cell Biology. Freeman and Co., New York.
Lovell-Badge, R., and E. Robertson. 1990. Xy-Female Mice Resulting from a Heritable Mutation in the Primary Testis-Determining Gene, Tdy. Development **109**:635-646.
Lundrigan, B. L., and P. K. Tucker. 1997. Evidence for multiple functional copies of the mole sex-determining locus, Sry, in African murine rodents. Journal of Molecular Evolution **45**:60-65.
Lyon, J. D., and V. D. Vacquier. 1999. Interspecies chimeric sperm lysins identify regions mediating species-specific recognition of the abalone egg vitelline envelope. Developmental Biology **214**:151-159.
Maddison, W. P., and D. R. Maddison. 1992. MacClade. Sinauer, Sunderland, MA.
Makalowski, W., J. H. Zhang, and M. S. Boguski. 1996. Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. Genome Research **6**:846-857.
Makova, K. D., S. Yang, and F. Chiaromonte. 2004. Insertions and deletions are male biased too: a whole-genome analysis in rodents. Genome Res **14**:567-573.

Margarit, E., A. Guillen, C. Rebordosa, J. Vidal-Taboada, M. Sanchez, F. Ballesta, and R. Oliva. 1998. Identification of conserved potentially regulatory sequences of the SRY gene from 10 different species of mammals. Biochemical and Biophysical Research Communications **245**:370-377.

Mcdonald, J. H., and M. Kreitman. 1991. Adaptive Protein Evolution at the Adh Locus in Drosophila. Nature **351**:652-654.

McElreavey, K., E. Vilain, S. Barbaux, J. S. Fuqua, P. Y. Fechner, N. Souleyreau, M. DocoFenzy, R. Gabriel, C. Quereux, M. Fellous, and G. D. Berkovitz. 1996. Loss of sequences 3' to the testis-determining gene, SRY, including the Y pseudoautosomal boundary associated with partial testicular determination. Proceedings of the National Academy of Sciences of the United States of America **93**:8590-8594.

Mcelreavy, K., E. Vilain, N. Abbas, J. M. Costa, N. Souleyreau, K. Kucheria, C. Boucekkine, E. Thibaud, R. Brauner, F. Flamant, and M. Fellous. 1992. Xy Sex Reversal Associated with a Deletion 5' to the Sry Hmg Box in the Testis-Determining Region. Proceedings of the National Academy of Sciences of the United States of America **89**:11016-11020.

Metz, E. C., and S. R. Palumbi. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Molecular Biology and Evolution **13**:397-406.

Metz, E. C., R. Robles-Sikisaka, and V. D. Vacquier. 1998. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proceedings of the National Academy of Sciences of the United States of America **95**:10676-10681.

Mindell, D. P. 1996. Positive selection and rates of evolution in immunodeficiency viruses from humans and chimpanzees. Proceedings of the National Academy of Sciences of the United States of America **93**:3284-3288.

Murphy, W. J., P. A. Pevzner, and S. J. O'Brien. 2004. Mammalian phylogenomics comes of age. Trends Genet **20**:631-639.

Nef, S., S. Verma-Kurvari, J. Merenmies, J. D. Vassalli, A. Efstratiadis, D. Accili, and L. F. Parada. 2003. Testis determination requires insulin receptor family function in mice. Nature **426**:291-295.

Nei, M., and S. Kumar. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Osterberg, M. K., O. Shavorskaya, M. Lascoux, and U. Lagercrantz. 2002. Naturally occurring indel variation in the Brassica nigra COL1 gene is associated with variation in flowering time. Genetics **161**:299-306.

Page, S. L., and M. Goodman. 2001. Catarrhine phylogeny: Noncoding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. Molecular Phylogenetics and Evolution **18**:14-25.

Pamilo, P., and R. J. W. ONeill. 1997. Evolution of the Sry genes. Molecular Biology and Evolution **14**:49-55.

Pilon, N., I. Daneau, V. Paradis, F. Hamel, J. G. Lussier, R. S. Viger, and D. W. Silversides. 2003. Porcine SRY promoter is a target for steroidogenic factor 1. Biology of Reproduction **68**:1098-1106.

Piontkivska, H., and A. L. Hughes. 2003. Evolution of vertebrate voltage-gated ion channel alpha chains by sequential gene duplication. Journal of Molecular Evolution **56**:277-285.

Podlaha, O., D. M. Webb, P. K. Tucker, and J. Z. Zhang. 2005. Positive selection for indel substitutions in the rodent sperm protein Catsper1. Molecular Biology and Evolution **22**:1845-1852.

Podlaha, O., D. M. Webb, and J. Z. Zhang. 2006. Accelerated evolution and loss of a domain of the sperm-egg-binding protein SED1 in ancestral primates. Molecular Biology and Evolution **23**:1828-1831.

Podlaha, O., and J. Zhang. 2003. Positive selection on protein-length in the evolution of a primate sperm ion channel. Proc Natl Acad Sci U S A **100**:12241-12246.

Polanco, J. C., and P. Koopman. 2007. Sry and the hesitant beginnings of male development. Developmental Biology **302**:13-24.

Pontiggia, a., R. Rimini, V. R. Harley, P. N. Goodfellow, R. Lovellbadge, and M. E. Bianchi. 1994. Sex-Reversing Mutations Affect the Architecture of Sry-DNA Complexes. Embo Journal **13**:6115-6124.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817-818.

Poulat, F., S. Soullier, C. Goze, F. Heitz, B. Calas, and P. Berta. 1994. Description and Functional Implications of a Novel Mutation in the Sex-Determining Gene Sry. Human Mutation **3**:200-204.

Qin, Y. J., and C. E. Bishop. 2005. Sox9 is sufficient for functional testis development producing fertile male mice in the absence of Sry. Human Molecular Genetics **14**:1221-1229.

Quill, T. A., D. J. Ren, D. E. Clapham, and D. L. Garbers. 2001. A voltage-gated ion channel expressed specifically in spermatozoa. Proceedings of the National Academy of Sciences of the United States of America **98**:12527-12531.

Rat Genome Sequencing Consortium. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature **428**:493-521.

Ren, D. J., B. Navarro, G. Perez, A. C. Jackson, S. F. Hsu, Q. Shi, J. L. Tilly, and D. E. Clapham. 2001. A sperm ion channel required for sperm motility and male fertility. Nature **413**:603-609.

Rooney, A. P., and J. Zhang. 1999a. Rapid evolution of a primate sperm protein: relaxation of functional constraint or positive Darwinian selection? Mol Biol Evol **16**:706-710.

Rooney, A. P., and J. Z. Zhang. 1999b. Rapid evolution of a primate sperm protein: Relaxation of functional constraint or positive Darwinian selection? Molecular Biology and Evolution **16**:706-710.

Saffer, J. D., S. P. Jackson, and M. B. Annarella. 1991. Developmental Expression of Sp1 in the Mouse. Molecular and Cellular Biology **11**:2189-2199.

Sainudiin, R., W. S. W. Wong, K. Yogeeswaran, J. B. Nasrallah, Z. H. Yang, and R. Nielsen. 2005. Detecting site-specific physicochemical selective pressures: Applications to the class IHLA of the human major histocompatibility complex and the SRK of the plant sporophytic self-incompatibility system. Journal of Molecular Evolution **60**:315-326.

Saitou, N., and M. Nei. 1987a. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4**:406-425.
Saitou, N., and M. NEI. 1987b. The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. Molecular Biology and Evolution **4**:406-425.
Sanchez, a., M. Bullejos, M. Burgos, C. Hera, C. Stamatopoulos, R. D. DelaGuardia, and R. Jimenez. 1996. Females of four mole species of genus Talpa (Insectivora, Mammalia) are true hermaphrodites with ovotestes. Molecular Reproduction and Development **44**:289-294.
Shi, J., and G. E. Gilbert. 2003. Lactadherin inhibits enzyme complexes of blood coagulation by competing for phospholipid-binding sites. Blood **101**:2628-2636.
Silva, J. C., and A. S. Kondrashov. 2002a. Patterns in spontaneous mutation revealed by human-baboon sequence comparison. Trends Genet **18**:544-547.
Silva, J. C., and A. S. Kondrashov. 2002b. Patterns in spontaneous mutation revealed by human-baboon sequence comparison. Trends in Genetics **18**:544-547.
Silvestre, J. S., C. Thery, G. Hamard, J. Boddaert, B. Aguilar, A. Delcayre, C. Houbron, R. Tamarat, O. Blanc-Brude, S. Heeneman, M. Clergue, M. Duriez, R. Merval, B. Levy, A. Tedgui, S. Amigorena, and Z. Mallat. 2005. Lactadherin promotes VEGF-dependent neovascularization. Nat Med **11**:499-506.
Singer, S. S., J. Schmitz, C. Schwiegk, and H. Zischler. 2003. Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates). Molecular Phylogenetics and Evolution **26**:490-501.
Steiper, M. E., and M. Ruvolo. 2003. New World monkey phylogeny based on X-linked G6PD DNA sequences. Molecular Phylogenetics and Evolution **27**:121-130.
Suzuki, Y., and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. Molecular Biology and Evolution **16**:1315-1328.
Suzuki, Y., and M. Nei. 2001. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. Mol Biol Evol **18**:2179-2185.
Swanson, W. J., C. F. Aquadro, and V. D. Vacquier. 2001. Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive Darwinian selection of sperm lysin. Molecular Biology and Evolution **18**:376-383.
Swanson, W. J., R. Nielsen, and Q. Yang. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol **20**:18-20.
Swanson, W. J., and V. D. Vacquier. 2002a. The rapid evolution of reproductive proteins. Nat Rev Genet **3**:137-144.
Swanson, W. J., and V. D. Vacquier. 2002b. The rapid evolution of reproductive proteins. Nature Reviews Genetics **3**:137-144.
Swanson, W. J., and V. D. Vacquier. 1995. Extraordinary Divergence and Positive Darwinian Selection in a Fusagenic Protein Coating the Acrosomal Process of Abalone Spermatozoa. Proceedings of the National Academy of Sciences of the United States of America **92**:4957-4961.
Tanaka, T., and M. NEI. 1989. Positive Darwinian Selection Observed at the Variable-Region Genes of Immunoglobulins. Molecular Biology and Evolution **6**:447-459.

Tevosian, S. G., K. H. Albrecht, J. D. Crispino, Y. Fujiwara, E. M. Eicher, and S. H. Orkin. 2002. Gonadal differentiation, sex determination and normal Sry expression in mice require direct interaction between transcription partners GATA4 and FOG2. Development **129**:4627-4634.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997a. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research **25**:4876-4882.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997b. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25**:4876-4882.

Tsaur, S. C., C. T. Ting, and C. I. Wu. 2001. Sex in Drosophila mauritiana: a very high level of amino acid polymorphism in a male reproductive protein gene, Acp26Aa. Mol Biol Evol **18**:22-26.

Tsaur, S. C., and C. I. Wu. 1997. Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of Drosophila. Molecular Biology and Evolution **14**:544-549.

Tucker, P. K., and B. L. Lundrigan. 1993. Rapid Evolution of the Sex-Determining Locus in Old-World Mice and Rats. Nature **364**:715-717.

Tucker, P. K., S. A. Sandstedt, and B. L. Lundrigan. 2005. Phylogenetic relationships in the subgenus Mus (genus Mus, family Muridae, subfamily Murinae): examining gene trees and species trees. Biological Journal of the Linnean Society **84**.

Turner, M. E., C. Martin, A. S. Martins, J. Dunmire, J. Farkas, D. L. Ely, and A. Milsted. 2007. Genomic and expression analysis of multiple Sry loci from a single Rattus norvegicus Y chromosome. Bmc Genetics **8**:-.

Veitia, R. A., M. Fellous, and K. McElreavey. 1997. Conservation of Y chromosome-specific sequences immediately 5' to the testis determining gene in primates. Gene **199**:63-70.

Vidal, V. P. I., M. C. Chaboissier, D. G. de Rooij, and A. Schedl. 2001. Sox9 induces testis development in XX transgenic mice. Nature Genetics **28**:216-217.

Vitullo, a. D., M. S. Merani, O. a. Reig, a. E. Kajon, O. Scaglia, M. B. Espinosa, and a. Perezzapata. 1986. Cytogenetics of South-American Akodont Rodents (Cricetidae) - New Karyotypes and Chromosomal Banding-Patterns of Argentinean and Uruguayan Forms. Journal of Mammalogy **67**:69-80.

Wang, X., and J. Zhang. 2004. Rapid evolution of mammalian X-linked testis-expressed homeobox genes. Genetics **167**:879-888.

Watanabe, T., R. Totsuka, S. Miyatani, S. Kurata, S. Sato, I. Katoh, S. Kobayashi, and Y. Ikawa. 2005. Production of the long and short forms of MFG-E8 by epidermal keratinocytes. Cell Tissue Res **321**:185-193.

Werner, M. H., M. E. Bianchi, a. M. Gronenborn, and G. M. Clore. 1995. Nmr Spectroscopic Analysis of the DNA Conformation Induced by the Human Testis-Determining Factor Sry. Biochemistry **34**:11998-12004.

Whitfield, L. S., R. Lovellbadge, and P. N. Goodfellow. 1993. Rapid-Sequence Evolution of the Mammalian Sex-Determining Gene Sry. Nature **364**:713-715.

Wilhelm, D., F. Martinson, S. Bradford, M. J. Wilson, A. N. Combes, A. Beverdam, J. Bowles, H. Mizusaki, and P. Koopman. 2005. Sertoli cell differentiation is

induced both cell-autonomously and through prostaglandin signaling during mammalian sex determination. Developmental Biology **287**:111-124.
Wolfe, K. H., and W. H. Li. 2003a. Molecular evolution meets the genomics revolution. Nature Genetics **33**:255-265.
Wolfe, K. H., and W. H. Li. 2003b. Molecular evolution meets the genomics revolution. Nat Genet **33 Suppl**:255-265.
Wong, W. S., Z. Yang, N. Goldman, and R. Nielsen. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics **168**:1041-1051.
Wyckoff, G. J., W. Wang, and C. I. Wu. 2000a. Rapid evolution of male reproductive genes in the descent of man. Nature **403**:304-309.
Wyckoff, G. J., W. Wang, and C. I. Wu. 2000b. Rapid evolution of male reproductive genes in the descent of man. Nature **403**:304-309.
Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci **13**:555-556.
Yang, Z., S. Kumar, and M. Nei. 1995a. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics **141**:1641-1650.
Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen. 2000a. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**:431-449.
Yang, Z. H., and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends in Ecology & Evolution **15**:496-503.
Yang, Z. H., S. Kumar, and M. Nei. 1995b. A New Method of Inference of Ancestral Nucleotide and Amino-Acid-Sequences. Genetics **141**:1641-1650.
Yang, Z. H., R. Nielsen, N. Goldman, and A. M. K. Pedersen. 2000b. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**:431-449.
Yang, Z. H., W. J. Swanson, and V. D. Vacquier. 2000. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Molecular Biology and Evolution **17**:1446-1455.
Zagotta, W. N., T. Hoshi, and R. W. Aldrich. 1990a. Restoration of inactivation in mutants of Shaker potassium channels by a peptide derived from ShB. Science **250**:568-571.
Zagotta, W. N., T. Hoshi, and R. W. Aldrich. 1990b. Restoration of Inactivation in Mutants of Shaker Potassium Channels by a Peptide Derived from Shb. Science **250**:568-571.
Zhang, J. 2000a. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J Mol Evol **50**:56-68.
Zhang, J. 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. Mol Biol Evol **21**:1332-1339.
Zhang, J., S. Kumar, and M. Nei. 1997. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. Mol Biol Evol **14**:1335-1338.
Zhang, J., R. Nielsen, and Z. Yang. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol **22**:2472-2479.

Zhang, J., and H. F. Rosenberg. 2002. Diversifying selection of the tumor-growth promoter angiogenin in primate evolution. Mol Biol Evol **19**:438-445.
Zhang, J., H. F. Rosenberg, and M. Nei. 1998a. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci U S A **95**:3708-3713.
Zhang, J., and D. M. Webb. 2004. Rapid evolution of primate antiviral enzyme APOBEC3G. Hum Mol Genet **13**:1785-1791.
Zhang, J., Y. P. Zhang, and H. F. Rosenberg. 2002a. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat Genet **30**:411-415.
Zhang, J. Z. 2000b. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. Journal of Molecular Evolution **50**:56-68.
Zhang, J. Z., K. D. Dyer, and H. F. Rosenberg. 2000. Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. Proceedings of the National Academy of Sciences of the United States of America **97**:4701-4706.
Zhang, J. Z., and M. Nei. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. Journal of Molecular Evolution **44**:S139-S146.
Zhang, J. Z., H. F. Rosenberg, and M. Nei. 1998b. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proceedings of the National Academy of Sciences of the United States of America **95**:3708-3713.
Zhang, J. Z., Y. P. Zhang, and H. F. Rosenberg. 2002b. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nature Genetics **30**:411-415.
Zhang, Z. L., N. Carriero, and M. Gerstein. 2004. Comparative analysis of processed pseudogenes in the mouse and human genomes. Trends in Genetics **20**:62-67.
Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas at Austin, Austin.