ABSTRACT

Title of Dissertation: MOLECULAR EVOLUTION OF MELANISM IN THE FELIDAE (MAMMALIA, CARNIVORA)

Eduardo Eizirik, Doctor of Philosophy, 2002

Dissertation Directed by: Adjunct Professor Stephen J. O'Brien
Department of Biology, and National Cancer Institute, NIH

and

Professor Gerald S. Wilkinson
Department of Biology

Coat color phenotypes are readily apparent and highly diverse mammalian traits, whose variability may be associated with adaptation to different environments. The cat family (Mammalia, Felidae) displays extensive variation in coat color, including several cases of polymorphic pigmentation. Among these, melanism (dark background coloration) has been confirmed to occur in at least eleven cat species, in some cases reaching appreciable population frequencies. The genetic basis and evolutionary history of melanism in the Felidae have not been studied in detail, and the molecular mechanisms responsible for this phenotype have not been characterized in this group.

The present study addresses the molecular genetic basis and evolutionary history of melanism in the cat family, aiming to identify and characterize genes involved in this phenotype in multiple related species. The domestic cat homologues of two candidate genes for melanism, *Agouti Signaling Protein* (*ASIP*) and the *Melanocortin-1 receptor* (*MC1R*), were mapped, cloned and sequenced, and genomic tools were developed to study the involvement of these loci in melanism in several felid species. Comparative analyses of the domestic cat *ASIP* and *MC1R* genes relative to available homologous sequences were used to characterize their genomic structure and patterns of molecular evolution of both loci across mammals (and other vertebrates, in the case of *MC1R*). Both genes were found to exhibit short conserved segments interspersed with highly variable regions, leading to overall moderate to fast rates of molecular evolution. An analysis of conserved sequence blocks in non-coding genomic regions revealed several segments of potential regulatory relevance, and a detailed characterization of these motifs in the *MC1R* promoter region is presented.

Three molecular variants associated with melanism were identified in three different cat species. In the domestic cat, black coloration is associated with a 2 bp deletion in *ASIP*, whereas two 'in-frame' deletions at adjacent locations in the *MC1R* gene are implicated in melanism in the jaguar and the jaguarundi. These variants were absent from all other surveyed felids, including melanistic individuals from five additional species. These findings indicate that melanism arose independently at least four times in the history of extant Felidae lineages, and reveal that dark coloration in the jaguarundi (more common in the wild) represents a derived condition, likely increased in frequency in an expansion process influenced by natural selection.

MOLECULAR EVOLUTION OF MELANISM IN THE FELIDAE (MAMMALIA, CARNIVORA)

by

Eduardo Eizirik

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2002

Advisory Committee:

Adjunct Professor Stephen J. O'Brien, Co-Chair, Advisor
Professor Gerald S. Wilkinson, Co-Chair
Professor Eric Baehrecke
Professor Matthew P. Hare
Professor Stephen M. Mount
Professor Sarah A. Tishkoff

DEDICATION

To my wife Laura,

for the love, understanding, and support during

the development of this and all my other projects.

ACKNOWLEDGMENTS

I would like to thank the following people and institutions for their help during the development of this project, and also in other related areas of my life.

Dr. Stephen O'Brien, for the opportunity to develop this research and many other exciting projects at the Laboratory of Genomic Diversity, for the advice, support and friendship, and for the encouragement to continue pursuing my scientific interests with energy and enthusiasm.

Dr. Gerald Wilkinson, for the advice and support throughout this project, and for the opportunity to maintain a more frequent interaction with the College Park campus community through the participation in his laboratory's activities. I also thank my Committee members Dr. Eric Baehrecke, Dr. Matthew Hare, Dr. Stephen Mount and Dr. Sarah Tishkoff, as well as past Committee members Dr. Soichi Tanda, Dr. Ulrich Mueller and Dr. Wolfgang Stephan for constructive suggestions, helpful advice and encouragement to pursue this research.

My wife Laura Utz, for the emotional and logistic support during this project and over the years, and particularly for the advice and help in formatting and proofreading this dissertation, as well in the preparation of some of the included figures and tables.

My parents, Cláudio and Marisa, and my sister Mariana, for their support and love throughout my life, which have been of fundamental importance for me to become who I am.

All my other family members, for the support and encouragement over the years.

Drs. Naoya Yuhki, Marilyn Menotti-Raymond, Warren Johnson, Bill Murphy, and Mark Springer, for fruitful scientific collaborations and stimulating discussions on topics related to this and other ongoing research projects.

My friends and colleagues at the Laboratory of Genomic Diversity, for helping create a joyful and stimulating research environment. In particular, I thank Al Roca, Gila Bar-Gal, Melody Roelke, Mike Dean, George Nelson, Carlos Driscoll, Agostinho Antunes, Taras Oleksyk, Colm O'hUigin, William Nash, Roscoe Stanyon and Jill Pecon-Slattery for helpful discussions related to this project, and Victor David, Tom Beck, Tarmo Anillo, Yoko Nishigaki, Kym Newmann, Bob Stephens, Jan Martenson, Leslie Wachter, Ali Wilkerson, John Page, John Arthur, Mike Malasky, Guo-Kuo Pei and Stan Cevario for technical assistance.

Steven Hannah, Solveig Pflueger, Cindy Bolte, Neil Wiegand, Leslie Lyons, Lyn Colenda and Kevin Cogan for helpful discussions and/or technical assistance in projects related to coat color genetics in the domestic cat.

My Brazilian friends and colleagues, many of whom have shared my interest and love for wild cats, evolutionary biology, biodiversity conservation and/or science in general over many years, in particular Cibele Indrusiak, Jan Mähler Jr., Dênis Sana,

Milton Mendonça Jr., Tatiane Trigo, Sandro Bonatto, Francisco Salzano, Thales Freitas, Loreta Freitas, Aldo Araújo, and my uncle Décio Eizirik.

Lois Reid, Joan Boxell, Laraine Main, Ginny Frye, Lori Larson and Barb Holder for administrative help at several moments during the development of this and other research projects.

I thank Al Roca, Marilyn Menotti-Raymond, Warren Johnson, Naoya Yuhki, Bill Murphy, Agostinho Antunes, Emma Teeling and Mike Dean for critically reading and providing helpful suggestions to different portions of this dissertation.

I am grateful to Joe Maynard, EFBC's Feline Conservation Center (USA), D. Sana, R. Morato, R. Gasparini-Morato, Associação Pró-Carnívoros (Brazil), P. Crawshaw Jr., L. Cullen, Instituto de Pesquisas Ecológicas (Brazil), T. Trigo, T. Freitas, C. Hilton, E. Salomão, R. Spindler, W. Swanson, C. Driscoll, G. Bar-Gal, M. Brown, M. Culver, A. Boldo, N. Sullivan, E. Rodrigues, M. Gomes, K. Nalewaik, N. Huntzinger, P. Menotti, L. Lyons, L. Garvey, and all the other people and institutions listed in Appendix 4 for help in obtaining biological samples required for this project.

The first four years of this work were supported by a fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil. The fifth year was supported by a Pre-Doctoral fellowship from the National Institutes of Health, USA. Support to attend scientific meetings during the development of this project was provided by the National Institutes of Health and the University of Maryland, College Park.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction to Dissertation

Molecular genetics of phenotypic variation

The interest in understanding the genetic bases of phenotypic variation is as old as the fields of genetics and evolutionary biology themselves (Darwin 1859, 1883; Mendel 1865; Galton 1889; Bateson 1913; Fisher 1930; Wright 1968), and reflects our scientific curiosity with respect to the mechanisms underlying natural phenomena. From the late 19th century through the 20th century, numerous studies investigated the heritable nature of phenotypic variation, its interaction with environmental factors and the possible mechanisms involved in its origin and evolutionary dynamics (e.g. Muller 1922; McClintock 1950; Mayr 1973; Wright 1977, 1978). These studies were based on laboratory experiments as well as field research on natural populations, and explored a diverse array of organisms concentrating on well-characterized model systems including bacteria, fungi, plants, insects and rodents. These pioneering efforts laid the foundations of modern genetics and led to the development of major theories on the nature and dynamics of genetic processes. However, their attempts to understand the physical (molecular) bases of heritable phenotypes were limited by a lack of suitable technologies to address these issues directly.

Until the advent of the era of molecular genetics (especially with respect to the ability to clone and sequence DNA segments) in the second half of the 20th century, knowledge on the genetic basis of phenotypic variation consisted of indirect inferences

on the number, nature and interaction among the genes (or 'factors') that influenced aspects of morphology, physiology, ecology and behavior (e.g. Muller 1922; Fisher 1930; Wright 1968, 1977). These studies led to the development of numerous hypotheses concerning the structure, function and interaction among genes, which could start to be investigated in detail and evaluated mechanistically as the direct molecular analysis of genes became feasible.

The molecular genetic basis of phenotypic traits was initially investigated in the context of human pathological conditions and experiments with model organisms (e.g. Lewis 1978; Robson et al. 1982; Struhl 1983; Gitschier et al. 1985; Burghes et al. 1987; review by Echols 2001). Evolutionary studies performing direct molecular analyses of genes determining natural phenotypic diversity are in their infancy, though some initial examples illustrate their contribution to our understanding of the interactions of genomic, cellular and physiological processes with organismal ecology and adaptation to varying environments (reviewed by Li 1997; Golding & Dean 1998; Hughes 1999). Some of the earliest examples include the study of the molecular signature of adaptive evolution at the loci coding for hemoglobins in several groups of vertebrates (reviewed by Perutz 1983; Li 1997); lysozyme in primates, ruminants and birds (Stewart & Wilson 1987; Kornegay et al. 1994); alcohol dehydrogenase in *Drosophila* (McDonald & Kreitman 1991); visual pigments in several vertebrates (e.g. Yokoyama & Yokoyama 1989; Shyue et al. 1995; Yokoyama 1997); genes of the vertebrate Major Histocompatibility Complex (Hughes & Nei 1988); and genes involved in reproductive traits of several organisms (reviewed by Swanson & Vacquier 2002). In the last few years, studies addressing adaptive aspects of

molecular evolution have multiplied rapidly (e.g. Yang & Bielawski 2000; Fay et al. 2002; Liberles & Wayne 2002; Smith & Eyre-Walker 2002), in some cases identifying clear connections between molecular genotypes and ecologically-relevant phenotypes (e.g. Hough et al. 2002; Zhang et al. 2002). The vast majority of these studies have focused on inter-specific molecular differences that implied adaptive evolution along the lineages leading to present-day species. So far, few studies have investigated the molecular basis of intra-specific phenotypic diversity, applying such information to analyze the adaptive significance of these traits in natural populations. Initial examples include work on the molecular basis of natural variation in disease resistance in humans and plants (e.g. Hill et al. 1992; Bergelson et al. 2001; Tishkoff et al. 2001; Dean et al. 2002), and also in a gene associated with coloration phenotypes in several species (*MC1R*), which will be described in the next two sections.

Coat color genetics in mammals

The genetic basis and evolutionary significance of coat color variation in mammals have attracted the attention of scientists since the late 19th century (e.g. Darwin 1883; Beddard 1895; Bateson 1913; Castle & Wright 1916; Wright 1917, 1918; Haldane 1927; Fisher 1930; Cott 1940; reviewed by Searle 1968; Robinson 1970a; Silvers 1979). Natural phenotypic variation in these traits has been used to propose classic theories of mammalian adaptation (e.g. Cott 1940), and their ecological and behavioral relevance have been explored in various contexts (e.g. Cott 1940; Ortolani & Caro 1996). At the same time, natural and induced variants in these traits became very important genetic markers for model animals in the pre-molecular

era, and were crucial for initial efforts in gene mapping and characterization of genomic processes (reviewed by Silvers 1979). For example, coat color variants were the first mammalian traits analyzed with Mendelian genetics (Castle & Allen 1903); they were used in the first demonstration of genetic linkage in a vertebrate (Haldane et al. 1915); and they formed the basis for the hypothesis of X-chromosome inactivation in mammals (Lyon 1961).

The use of these traits as genetic markers led to the accumulation of a comprehensive body of knowledge on the genetics, biochemistry, physiology and molecular biology of the processes involved in coat color determination of the house mouse (*Mus musculus*) (e.g. Silvers 1979; Jackson 1994; Barsh 1996). Several genes involved in the production and distribution of pigment in mice have now been characterized at the molecular level (Jackson 1994; Barsh 1995, 1996; He et al. 2001). They have been shown to be part of diverse cellular, developmental and physiological processes, and in some cases to be implicated in pathologies such as anemia, sterility and neurological disorders (e.g. Silvers 1979; Fleischman 1993; Jackson 1994). Homologues for most of these genes have been identified and characterized in humans, and in several cases are also associated with pathological conditions (e.g. hypopigmentation [Fleischman 1993] and oculocutaneous albinism [Manga et al. 1997]; reviewed by Barsh [1995] and Sturm et al. [2001]). Their role in human skin and hair color variation is still poorly understood, but available evidence indicates that complex interactions and different forms of selection involving some of these same genes could be responsible for the phenotypic variation observed in our species (Barsh 1996; Box et al. 1997; Rana et al. 1999; Harding et al. 2000; Rees 2000; Sturm et al.

2001). Homologues for some of these genes have also been identified and characterized in other species such as the cow (*Bos taurus*), pig (*Sus scrofa*) and horse (*Equus caballus*), and often implicated in coat color and/or pathological phenotypes similar to those observed in mice and humans (e.g. Klungland et al. 1995; Joerg et al. 1996; Kijas et al. 1998, 2001; Metallinos et al. 1998; Marklund et al. 1999).

In spite of this significant progress, detailed knowledge about the structure, function, regulation and interactions of the genes involved in pigmentation remains in its infancy for virtually all vertebrate groups. Moreover, little has been attempted so far in terms of integrating the available knowledge from model organisms with the phenotypic diversity observed in wild species and natural populations, in the context of directly addressing and testing evolutionary hypotheses involving these traits. The first examples of such attempts in vertebrates involve the characterization of molecular variants of the *MC1R* gene (described in the next section) and their association with coloration phenotypes in natural populations of humans (*Homo sapiens*), bananaquits (*Coereba flaveola*) and black bears (*Ursus americanus*). In humans *MC1R* variants are associated with red hair and fair skin (Valverde et al. 1995; Box et al. 1997), and evolutionary studies indicate the occurrence of contrasting patterns of natural selection acting on this gene in different geographic populations (Rana et al. 1999; Harding et al. 2000; Rees 2000; Smith et al. 2001; more details are provided in Chapter 3). Bananaquits are birds whose populations in different Caribbean islands show variable frequencies of melanism (defined in the next section) associated with a mutation at *MC1R*, whose dynamics may also be influenced by contrasting selective pressures (Theron et al. 2001). In the case of black bears, the molecular basis of the 'Kermode'

phenotype (white-phased individuals occurring in British Columbia, Canada) has been found to be a different type of mutation at *MC1R*, likely causing loss of function of the resulting protein (Ritland et al. 2001). Until now, no study has addressed the molecular basis of coat color phenotypes in multiple species of the same family of organisms, attempting to investigate aspects of their evolutionary history and adaptive significance.

Melanism

Among the diverse traits influencing mammalian coat color, melanism has received considerable attention over the last four decades, both in terms of experimental work in the mouse (Silvers 1979; Jackson 1993, 1994; Barsh 1995, 1996) and frequent reports of naturally occurring variants in many taxa (Searle 1968; Robinson 1970a). Melanism is a phenomenon present in many life forms, and has been broadly defined as 'any situation in which there is, on average, a general darkening of the ground color or patterning of an organism' (Majerus 1998). In the mouse, both dominant and recessive forms of this trait have been described, and primarily attributed to two different genetic loci (allelic series): *agouti* and *extension* (Silvers 1979). Mutations at these two loci were found to cause inverse phenotypic effects in the mouse: in the *agouti* series the most dominant alleles lead to lighter (yellow) pigmentation, whereas recessive mutants are associated with melanism; the opposite is observed in the *extension* series, in which both dominant melanistic and recessive yellow mutants have been described (Silvers 1979; Jackson 1994; Barsh 1996). Based on inheritance patterns of similar coloration mutants observed in other

species, it has been hypothesized that variants at the *agouti* and *extension* series were also implicated in pigmentation diversity in other groups (Haldane 1927; Searle 1968; Robinson 1970a, 1976). Additional loci at which mutant alleles also led to melanistic phenotypes were identified through experimental crosses in mice (e.g. mahogany [*mg*], mahoganoid [*md*], umbrous [*U*] and dark [*da*]; Silvers 1979); however, these genes have not yet been as well characterized in mice as *agouti* and *extension*, and have not been proposed to alter coloration phenotypes in other mammalian species.

Both *agouti* and *extension* have now been well characterized at the molecular level in mice (Bultman et al. 1992; Mountjoy et al. 1992; Robbins et al. 1993; Perry et al. 1996). *Extension* corresponds to the *MC1R* (*Melanocortin-1 receptor*) gene, also called *MSHR*, for *α-Melanocyte Stimulating Hormone* [α-MSH] *Receptor*. *MC1R* encodes a seven-transmembrane G-protein-coupled receptor expressed in skin and hair follicle melanocytes, and also in some immune system cells (Mountjoy et al. 1992; Smith et al. 2001). It responds to extra-cellular α-MSH binding and activates the synthesis of eumelanin (dark pigment: black or brown) by means of G-protein coupling and cAMP signaling (Robbins et al. 1993; Jackson 1994; Barsh 1996; Lu et al. 1998).

Agouti encodes a unique peptide whose human homologue has been called *ASIP* (for *Agouti Signaling Protein;* the human protein has been called ASP [Wilson et al. 1995]). The *ASIP* designation for the gene (ASIP for the protein) will be used preferentially throughout this dissertation. *ASIP* codes for a paracrine peptide that is produced in hair follicles (dermal papilla cells) and behaves as an inverse agonist to the *MC1R*, preventing its activation by α-MSH and thus inducing a switch from

eumelanin to pheomelanin (light pigment: yellow or reddish) synthesis (Lu et al. 1994; Barsh 1996; Abdel-Malek et al. 2001). The wild-type pattern of regulation of these two genes in dorsal hairs of mice is an initial and terminal synthesis of eumelanin, interrupted by a switch to pheomelanin production caused by a pulse of *agouti* expression (Jackson 1994; Barsh 1996). This produces a banded hair shaft, called the 'agouti' phenotype, which can also be observed in many other mammals including the domestic cat (Searle 1968). Melanistic (black or mostly black) mouse phenotypes can be due to mutants producing either a constitutively active MC1R protein, with dominant inheritance, or a defective (or defectively expressed) agouti peptide, with recessive inheritance (Silvers 1979; Jackson 1993, 1994).

A third gene involved in the causation of melanism in the mouse, *mahogany* (*mg*), has also been recently cloned (Gunn et al. 1999; Nagle et al. 1999). It was found to encode a transmembrane form of *attractin* (ATRN) which plays an important role in stabilizing the molecular interaction between *ASIP* and *MC1R* on the melanocyte plasma membrane, and also participates in other biological processes including activities in immune and neurological systems (Gunn et al. 2001; He et al. 2001). So far, *mahogany/ATRN* has not been found to play a role in pigmentation phenotypes in any species other than the mouse, and therefore the remainder of this section will focus on providing further background on the *MC1R* and *ASIP* genes.

The mouse *MC1R* coding region is intronless and spans 945 bp (315 amino-acids), while the human homologue has two additional codons (Mountjoy et al. 1992). The promoter region has so far been characterized in humans and mice (Moro et al. 1999; Adachi et al. 2000; Makova et al 2001; Smith et al 2001), and critical elements

for *MC1R* expression have been identified *ca.* 500 bp upstream of the translation initiation site. The mouse *agouti* (*ASIP*) gene is composed of three coding exons (containing 170 bp, 65 bp and 385 bp, respectively) that span a genomic region of approximately 5 kb, as well as three or four more distant upstream non-coding exons, which were found to be variably included in alternatively transcribed/spliced mRNA transcripts (Siracusa 1994; Vrieling et al. 1994). The overall genomic region that includes all these exons consists of over 110 kb (Bultman et al. 1994). The resulting agouti (ASIP) peptide is 131 amino acids long, and includes a signal sequence, an N-linked glycosylation site, a central basic domain and a C-terminal tail that includes 10 cysteine residues (Perry et al. 1996; Miltenberger et al. 2002; more details will be provided in Chapter 2).

In addition to mice and humans, *ASIP* and *MC1R* have now been sequenced in a few other mammal species (and some other vertebrates, in the case of *MC1R*), in several instances leading to the identification of sequence variants in one or both genes that are associated with melanism or other coat color phenotypes (e.g. Joerg et al. 1996; Marklund et al. 1996; Våge et al. 1997; Everts et al. 2000; Newton et al. 2000). *MC1R* variants have been implicated in melanism in the cow (Klungland et al. 1995), red fox (Våge et al. 1997), pig (Kijas et al. 1998), sheep (Våge et al. 1999), and also in chickens (Takeuchi et al. 1996) and bananaquit birds (Theron et al. 2001). *ASIP* variants have been implicated in melanistic phenotypes in the red fox (Våge et al. 1997); rat (Kuramoto et al. 2001a) and horse (Rieder et al. 2001), in addition to the originally described mouse mutations. This abundance of examples of coat-color

altering mutations in these two interacting loci makes them attractive candidate genes for studies of the genetic basis of melanism in other mammalian systems.

Genetics of coat color variation in the Felidae

Among mammals, felids are an extremely interesting group for the study of coat colors and their evolution. Extensive intra-specific variation in these traits is observed in domestic and wild cats, and conspicuous diversity across species has been the basis for proposed hypotheses of adaptation, biogeography and ecological associations (Beddard 1895; Cott 1940; Weigel 1961; Kitchener 1991; Ortolani & Caro 1996). Polymorphic pigmentation is common in wild cat species, including variation in background color (e.g. light tan or yellow to gray, reddish or dark brown) and also in the presence, shape, coloration and distribution of markings (spots, stripes, rosettes). In several felid species the variable components of polymorphic color are known to segregate geographically, and have been used to describe differentiated 'subspecies' or local populations (e.g. Pocock 1940; Garcia-Perea 1994). It is conceivable that these variants may be associated with adaptation to local environments, and may be important life history components in these species. Nonetheless until now there has been no direct study of their ecological significance or underlying genetic mechanisms, both of which may aid in the elucidation of their adaptive relevance.

The domestic cat (*Felis catus*) can serve as a useful model for evolutionary genetic studies in the Felidae. Available genomic resources in this species facilitate the mapping and cloning of candidate genes (e.g. Menotti-Raymond et al. 1999;

Murphy et al. 2000; Beck et al. 2001), and a framework of evolutionary studies at the phylogenetic and population levels (e.g. Johnson & O'Brien 1997; Pecon-Slattery & O'Brien 1998; Eizirik et al. 1998, 2001) provide useful information for comparative studies within and among different species. The domestic cat can also be an important model in the specific case of coat color phenotypes, as this species exhibits ample variation in several pigmentation traits. At least nine different genetic loci involved in cat coloration diversity have been identified on the basis of breeding experiments (Doncaster 1904; Whiting 1918; Wright 1918; Castle 1919; Robinson 1959, 1976, 1991). These include the following (mutant allele described; nomenclature following Robinson [1991]): (i) *a* (non-agouti), causing black color (melanism); (ii) *b* (brown), changing black coloration to brown; (iii) *c* (color), causing the Siamese (c^s) and Burmese (c^b) phenotypes; (iv) *d* (dilute), causing dilution from black to bluish gray; (v) *I* (inhibitor of melanin), causing the 'silver' or 'smoke' grayish phenotypes; (vi) *O* (X-linked Orange), changing black pigment to orange, and causing the 'tortoiseshell' and 'calico' phenotypes in heterozygote females; (vii) *S* (white spotting); (viii) *T* (tabby), affecting the striping/spotting patterns; and (ix) W (dominant white), causing all-white coloration. So far only one of these loci (the *c* [Siamese/Burmese] locus, likely a homologue of the *albino*/*Tyrosinase* gene) has been mapped (O'Brien et al. 1986), and none has been characterized at the molecular level.

The occurrence of melanism has been at least anecdotally reported in as many as 20 of the 37 felid species (Ulmer 1941; Robinson 1976; Dittrich 1979) belonging to all of the eight major evolutionary lineages identified in the family Felidae (Johnson & O'Brien 1997; Pecon-Slattery & O'Brien 1998; Mattern & McLennan 2000) (Figure 1-

1). In eleven of these species (representing at least five of the eight major lineages) the occurrence of melanism has been confirmed by photographic or other direct evidence (Table 1-1; Figure 1-1). Interestingly, although melanistic variants seem to occur at considerably high frequencies in some cat species, in no case has this phenotype reached fixation in any extant felid (Robinson 1970a, 1976; Kitchener 1991; Nowell & Jackson 1996). This is intriguing in the context of hypothesizing about the evolutionary origin and maintenance of this phenotype in multiple species of the same family, and the potential implications for the adaptive relevance of such a trait in different ecological contexts.

In the domestic cat, melanism exhibits recessive inheritance (Whiting 1918; Robinson 1959; see Table 1-1), suggesting *agouti/ASIP* as a candidate gene. The same inheritance pattern has been observed in the leopard (*Panthera pardus*) (Robinson 1969, 1970b) and suggested for the pampas cat (*Oncifelis colocolo*) (Dittrich 1979), whereas in the jaguar (*Panthera onca*) and possibly also in the jungle cat (*Felis chaus*) melanism is inherited as a dominant trait (Dittrich 1979; see Table 1-1). Such inheritance pattern would suggest *extension/MC1R* as a potential candidate gene for melanism in these species, given the available knowledge from the mouse system.

Scope of this Study

The goal of this project was to investigate the molecular basis and evolutionary history of melanism in the Felidae. For this purpose, it was necessary to identify molecular variants associated with melanistic phenotypes in one or more felids, and to investigate their occurrence in multiple species of the cat family. Using a candidate-

gene approach based on previous studies in the mouse and other mammals (see above), I sought to (i) characterize the domestic cat homologues of two of these genes, *ASIP* and *MC1R*; (ii) describe their patterns of molecular evolution in mammals; (iii) assess their potential for evolutionary studies in the Felidae; (iv) investigate their involvement in melanistic phenotypes in multiple felid species; and (v) draw evolutionary inferences on their structure, function, and influence on coat color diversity in domestic and wild cats.

In Chapter 2, I describe the mapping, cloning and characterization of the *ASIP* gene in the domestic cat, and investigate its genomic structure and patterns of sequence evolution across mammals. An adjacent gene (*AHCY*) contained in the same domestic cat genomic clone is also partially characterized, and used to perform comparative analyses relative to *ASIP*. In addition, I investigate the utility of a segment of the *ASIP* gene for phylogenetic and population-genetic studies in the Felidae, and identify novel microsatellite (STR) markers tightly linked to this gene with potential applicability for mapping and evolutionary studies in cat species.

In Chapter 3, I describe the mapping, cloning and characterization of the domestic cat *MC1R* gene, and perform evolutionary analyses addressing patterns of sequence conservation and variability in its coding region and in adjacent genomic segments. This chapter includes the characterization of constrained *versus* relaxed portions of the *MC1R* gene, an investigation of substitution rates and patterns in this locus in mammals, and the assessment of molecular variation in *MC1R* among multiple species of the Felidae. I also perform genomic comparisons of non-coding

DNA segments adjacent to the *MC1R* gene, and identify novel conserved sequence blocks that are likely involved in regulatory functions of *MC1R* or adjacent loci.

In Chapter 4, I employ linkage analysis (using STR markers developed in Chapter 2) to assess the potential implication of the *ASIP* gene in melanism in the domestic cat, and then describe the identification of molecular variants at this locus and also at *MC1R* that are associated with melanistic phenotypes in three different felid species. Based on these results I draw inferences on the evolution of melanism in the Felidae and the functional implications of the identified molecular variants.

In Chapter 5, I review and integrate the main findings described in Chapters 2, 3 and 4, and discuss their implications for the study of the *MC1R-ASIP* system and its relevance for the evolution of coat color diversity in felids and other mammals. I include discussions on the phylogeny of the Felidae and the evolution of melanism in this group. I conclude with a general discussion of future perspectives and applications for molecular genetic analyses of phenotypic polymorphisms in natural populations, and their potential contribution for evolutionary studies in the Felidae and in other groups of organisms.

Table 1-1. Available information on the occurrence of melanism in felid species; bold types indicate species for which reliable evidence of melanism exists. Common species names are given in Figure 1-1.

Species	Strongest evidence and original references	Proposed inheritance	No. of offspring analyzed in the original literature source
Felis catus	**Visual [2, 11, 18]**	**Recessive [2, 11, 19]**	**1 black offspring from a pair of wild type parents [2, 11]**
Felis chaus	**Photograph [15]**	**Dominant [15]**	**1 wild-type offspring from a pair of melanistic parents [15]**
Felis silvestris, F. lybica	Anecdotal [9,15]	-	-
Panthera pardus	**Visual [13,18]**	**Recessive [12, 13, 19]**	**Total of 439 offspring [12, 13]**
Panthera onca	**Visual [15, 18]**	**Dominant [15, 19]**	**Total of 81 offspring [15]**
Panthera leo	Anecdotal [15]	-	-
Panthera tigris	Anecdotal [3,4]	-	-
Panthera uncia	Anecdotal [1]	-	-
Lynx rufus	**Photograph [4]**	-	-
Leptailurus serval	**Video [1, 4, 9, 18]**	-	-
Oncifelis geoffroyi	**Visual [8, 18]**	-	-
Oncifelis guigna	**Photograph [10, 15, 16, 17]**	-	-
Leopardus tigrinus	**Visual [7, 15, 18]**	-	-
Lynchailurus colocolo	**Photograph [15]**	**Recessive [15]**	**2 black offspring from a pair of wild type parents [15]**
Puma concolor	Anecdotal [6]	-	-
Prionailurus bengalensis	Anecdotal [4, 14]	-	-
Caracal caracal	Anecdotal [4]	-	-
Profelis aurata	Anecdotal [5]	-	-
Catopuma temmincki	**Photograph [4, 14]**	-	-

References: [1]Lönnberg 1898; [2]Whiting 1918; [3]Burton 1928; [4]Ulmer 1941; [5]Lamotte 1942; [6]Young & Goldman 1946; [7]Weigel 1961; [8]Coleman 1974; [9]Angwin 1975; [10]Junge 1975; [11]Robinson 1959; [12]Robinson 1969; [13]Robinson 1970b; [14]Robinson 1976; [15]Dittrich 1979; [16]Sunquist & Sanderson 1998; [17]Dunstone et al. 1998; [18]Personal observation; [19]Results confirmed by this study.

Figure Legend

Figure 1-1. Phylogeny of the Felidae showing the occurrence of melanism in the family. Asterisks indicate nodes supported by congruence among various analyses reported in recent phylogenetic studies (Johnson & O'Brien 1997, Pecon-Slattery & O'Brien 1998, Mattern & McLennan 2000), defining eight major evolutionary lineages (identified by brackets). Additional phylogenetic resolution within the Ocelot and *Panthera* lineages is derived from Johnson et al. (1998, 1999) and Kim et al. (2001). The basal relationships among these major lineages and the position of three unaligned species (marbled cat [*Pardofelis marmorata*], serval [*Leptailurus serval*] and rusty-spotted cat [*Prionailurus rubiginosus*]) are depicted here as a polytomy. A fourth unaligned species (Pallas cat [*Otocolobus manul*]) is shown as a relative of the Domestic cat lineage, based on previous studies (Collier & O'Brien 1985; Salles 1992). The felid taxonomy used here and throughout this study follows Wozencraft (1993). Black circles indicate felid species for which definitive evidence of melanism has been confirmed (see Table 1-1); starting at the top of the figure, these are: leopard (*Panthera pardus*), jaguar (*Panthera onca*), bobcat (*Lynx rufus*), serval (*Leptailurus serval*), Asian golden cat (*Catopuma temmincki*), domestic cat (*Felis catus*), jungle cat (*Felis chaus*), Geoffroy's cat (*Oncifelis geoffroyi*), kodkod (*Oncifelis guigna*), oncilla (*Leopardus tigrinus*) and pampas cat (*Oncifelis colocolo*). Open circles indicate additional cat species for which anecdotal evidence of melanism has been reported (see Table 1-1); from the top, these are: lion (*Panthera leo*), tiger (*P. tigris*), snow leopard (*P. uncia*), leopard cat (*Prionailurus bengalensis*), caracal (*Caracal caracal*), African golden cat (*Profelis aurata*), puma (*Puma concolor*), European wildcat (*Felis*

silvestris), and African wildcat (*F. lybica*, currently considered to be conspecific with *F. silvestris* [Nowell & Jackson 1996]). In all cases melanism exists as a polymorphic trait, observed to occur in one or more geographic areas. The triangle shows the phylogenetic position of the jaguarundi (*Herpailurus yaguarondi*), which shows conspicuous coloration polymorphism, but has not been referred to as exhibiting melanism. A molecular analysis of jaguarundi color phenotypes will be included in Chapter 4.

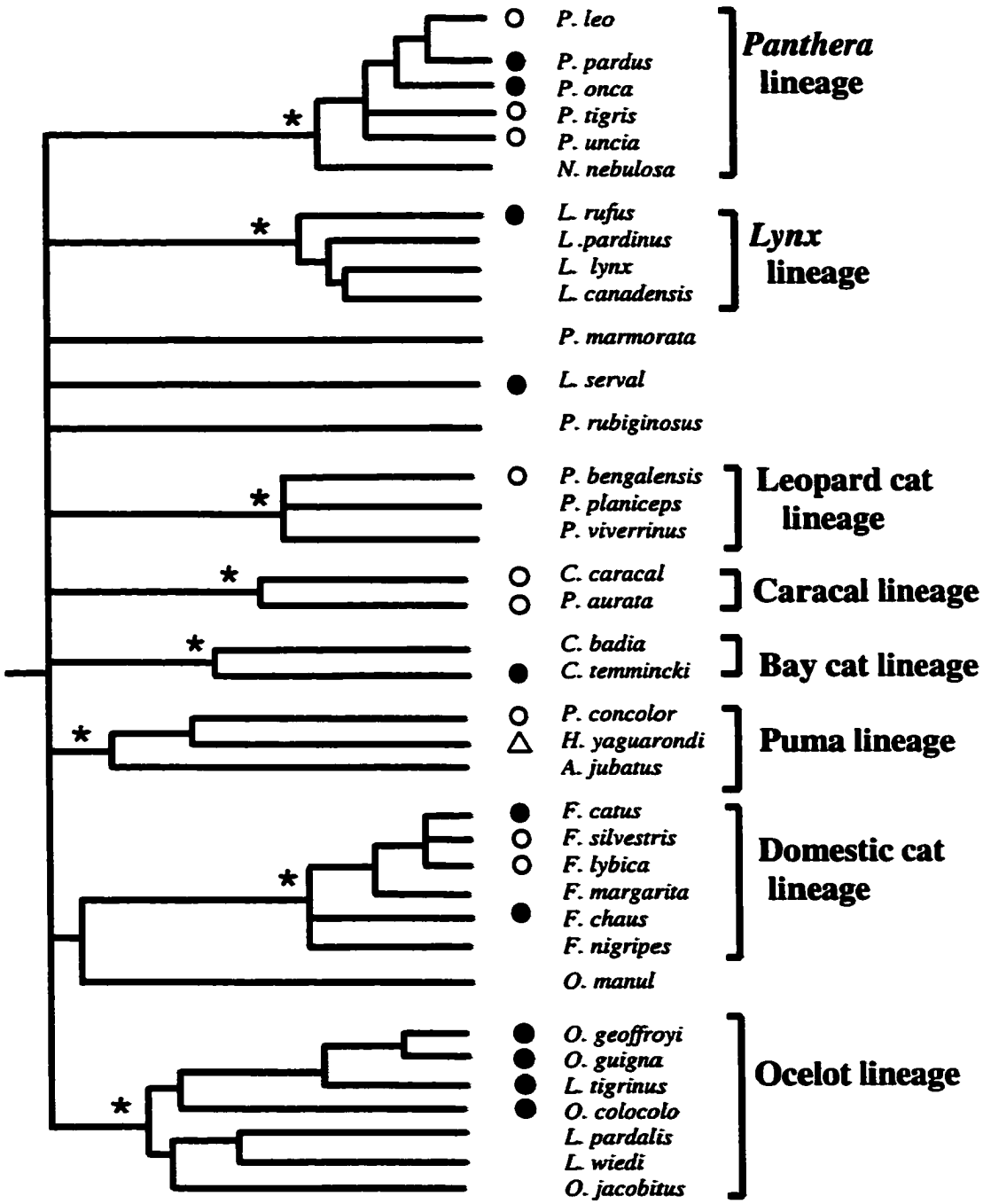


Figure 1-1

CHAPTER 2

Evolutionary Characterization of the Feline *ASIP* (*agouti*) Gene

Abstract

The *agouti* / *ASIP* (*Agouti Signaling Protein*) gene is a classical mouse coat color locus, and known to be involved in pigmentation phenotypes in three additional mammalian species. So far *ASIP* has not been characterized in cats, and its patterns of molecular evolution in mammals have not been studied in detail. In the present study I describe the mapping, cloning and sequencing of *ASIP* in the domestic cat, and evolutionary analyses of this gene among different mammalian groups and within the Felidae. *ASIP* maps to domestic cat chromosome A3p, and is located *ca.* 15 kb away from neighboring gene *AHCY*. Comparative analyses revealed that *ASIP* is considerably variable, exhibiting similar synonymous substitution rates and over 6-fold higher nonsynonymous rates relative to *AHCY*. This and other comparisons suggest that several portions of *ASIP* are subjected to relaxed functional constraints, leading to a relatively high rate of molecular evolution. Amino acid sites that have been experimentally demonstrated to be critical for ASIP activity were found to be completely conserved across mammals, suggesting maintenance of common function in these species, and the potential to apply evolutionary comparisons to identify important residues in this peptide. In this study I also evaluate the utility of *ASIP* sequences for evolutionary studies in the Felidae, by analyzing a 1.75 kb segment encompassing intron 2 in multiple cat species. This segment displays considerable variability, providing novel insights into felid relationships.

Introduction

Variation at the classical *agouti* locus has been known for decades to influence coat color variation in mice, and possibly also in other mammals (Silvers 1979; Searle 1968; Robinson 1970a). Cloning of this gene in the early 1990's revealed that it encoded a unique peptide secreted by dermal papilla cells and acting as a paracrine antagonist to the melanocortin-1 receptor (MC1R) (Bultman et al. 1992; Miller et al. 1993; Lu et al. 1994; see Chapter 1 for more details). The occurrence of gain-of-function yellow *agouti* mouse mutants that exhibit pleiotropic effects such as obesity, increased size, diabetes and susceptibility to tumors (Bultman et al. 1992; Duhl et al. 1994) contributed to promote a wave of studies on the molecular biology of these variants, producing a comprehensive body of knowledge on the structure and function of this gene and its peptide product (e.g. Duhl et al. 1994; Michaud et al. 1994; Miller et al. 1994; Vrieling et al. 1994; Siracusa 1994; Hustad et al. 1995; Millar et al. 1995; Perry et al. 1996). The human homologue of *agouti* (named *ASIP* for *Agouti Signaling Protein*) was mapped and cloned soon after the characterization of the mouse gene (Kwon et al. 1994; Wilson et al. 1995), however its functional role in our species has still not been fully understood (Sturm et al. 2001; Voisey et al. 2001; Voisey & van Daal 2002). In addition to potentially influencing human pigmentation (Kanetsky et al. 2002), current knowledge indicates that *ASIP* is implicated in lipid metabolism, being expressed at highest levels in adipocytes (Voisey & van Daal 2002).

In addition to mouse and human, the *agouti/ASIP* gene has so far been sequenced in five species: red fox, cow, pig, rat and horse (Våge et al. 1997; Leeb et al. 2000; Kuramoto et al. 2001a; Rieder et al. 2001). *ASIP* molecular variants

involved in pigmentation phenotypes (recessive melanism) have been identified in the red fox, rat and horse. None of the previous studies of *ASIP* involved evolutionary analyses of its rates and patterns of sequence variation among multiple mammals, or direct comparisons of measures of diversity to other genomic loci. Some studies have referred to *ASIP* as being 'highly conserved' (or 'highly homologous') among mammals (e.g. Kwon et al. 1994; Miltenberger et al. 2002), however its levels of conservation have not been systematically compared to other genes, nor evaluated in terms of spatial homogeneity.

Many domestic and wild cats display banded (agouti) or yellowish hairs throughout most of their body, and often exhibit a marked distinction between a lighter ventrum and a darker dorsum, both of which are known to be influenced by the *agouti* gene in mice (Vrieling et al. 1994; Siracusa 1994). These observations suggest a similar role for *ASIP* in the background pigmentation of felids, and a potential involvement of this locus in recessively inherited melanism mutants, such as observed in the domestic cat (Robinson 1959, 1970a). The *ASIP* gene has not been mapped or sequenced in any felid species, and its characterization in cats is required to allow further investigation of its role in melanistic phenotypes observed in this group (see Chapter 1).

In this study I describe the mapping, cloning and characterization of the *ASIP* gene in the domestic cat, and the development of genomic tools for the evolutionary study of this locus in domestic and wild felids. Mapping of feline *ASIP* sequences confirmed a location on chromosome A3p homologous to the position of this locus in humans and mice, and analysis of a large-insert clone allowed the characterization of

the *ASIP* coding region and adjacent genomic segments in this species. Comparative analyses were used to infer patterns of sequence conservation in *ASIP* among mammals, and one segment of this gene was evaluated for its potential applicability in evolutionary studies within and among Felidae species.

Materials and Methods

PCR-based sequencing and mapping of the feline *ASIP* gene

Multiple PCR (Polymerase Chain Reaction: Mullis & Faloona 1987) primers for the *ASIP* gene were designed on the basis of previously available mouse, human, fox and cow sequences (Bultman et al. 1992; Kwon et al. 1994; Wilson et al. 1995; Våge et al. 1997; Oulmouden et al., unpublished [GenbankAccession X99692]), and tested in domestic and wild cats in all possible forward and reverse combinations with Applied BioSystems (ABI) *Ampli Taq* DNA polymerase or *Taq Gold* (Perkin Elmer) DNA polymerase, varying the annealing temperature and $MgCl_2$ concentration. Primers agoF1 and agoR1 (see Table 2-1 for list of all primers), located in exons 2 and 3, respectively (Figure 2-1), successfully amplified the *ASIP* intron 2 using *Ampli Taq*, a 57°C annealing temperature and 1.25 mM $MgCl_2$. This 1.8 kb segment was then sequenced using the PCR primers as well as internal oligonucleotides agoR3, agoF4, agoR4 and agoR7, designed by primer walking. A 177-bp segment of *ASIP* exon 4 was amplified with primers agoF3 and agoR5 (Figure 2-1), using a touchdown-PCR profile (annealing temperature decreasing from 60°C to 50°C in the 12 initial cycles, followed by 33 cycles at 50°C) with *Ampli Taq* and 1.5 mM $MgCl_2$. This fragment was sequenced with the PCR primers. Amplification products were purified with

Centricon-100 concentrators (Amicon) and sequenced using ABI Dye terminator chemistry. Sequencing products were isopropanol-precipitated and analyzed with an ABI 373 automated sequencer. Resulting sequences were checked with BLAST (Altschul et al. 1990) and used for initial genomic characterization and design of additional primers with the program Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). A complete list of computer programs used in this study (all chapters) is given in Appendix 1.

Based on the initial sequence of *ASIP* intron 2 in domestic and wild felids, two assays were developed to map the location of this gene in the cat genome. To utilize the domestic cat radiation hybrid (RH) panel (Murphy et al. 2000) for mapping this gene, primers agoRH-F1 and agoRH-R1 were designed to amplify a 206 bp product within *ASIP* intron 2 (Figure 2-1). PCR conditions were optimized for specific amplification of cat genomic DNA, with no product in the hamster background cell line. The PCR-based typing assay was performed in duplicate in 96-well format, with 31 cycles using 59.5°C annealing temperature, *AmpliTaq* Gold and 1.5 mM $MgCl_2$. The resulting data were analyzed using RHMAP (Boehnke et al. 1991) relative to markers and genes previously mapped by Murphy et al. (2000). The RH analysis uses the frequency at which different markers are retained in the same radiation-induced fragments to estimate their relative position and physical distance in the original genome.

To map the genomic location of *ASIP* relative to available STR (microsatellite) markers (Menotti-Raymond et al. 1999), another PCR-based assay was designed based on the presence of a B2 SINE element within *ASIP* intron 2 in domestic cats, but not

in Asian leopard cats (*Prionailurus bengalensis*) (see Figure 2-1C). These two species have been previously used to generate an inter-specific back-cross (ISB) pedigree for application in feline genetic mapping projects, with which 253 STR markers and 81 coding loci have been located in the domestic cat genome (Menotti-Raymond et al. 1999; Menotti-Raymond et al., in press). Primers agoISB-F1 and agoISB-R1 were designed to flank this SINE integration, producing a 521 bp product in domestic cats and 286 bp product in leopard cats. The large difference in product length allowed straightforward size-based typing of parental *versus* recombinant chromosomes in ethidium bromide-stained 1% agarose gels. LOD scores between *ASIP* and previously available STR loci (Menotti-Raymond et al. 1999) were estimated with the program FASTLINK (Cottingham et al. 1993).

Genomic cloning of the domestic cat *ASIP*

To obtain the full coding sequence of the domestic cat *ASIP* gene, explore its surrounding chromosomal region, and obtain new polymorphic markers closely linked to this locus for use in mapping studies (described below and in Chapter 4), a large-insert genomic clone containing *ASIP* was isolated and characterized. A domestic cat BAC library (RPCI 86; Beck et al. 2001) was screened using a homologous probe derived from a 177 bp PCR product amplified from exon 4. This segment was amplified with primers agoF3 and agoR5, and labeled with ^{32}P using a random priming kit (Boehringer-Manheim). The hybridization was performed simultaneously with probes for different genomic regions of interest (not shown), and positive clones for the *ASIP* gene were identified using direct PCR from bacterial colonies, confirmed

by sequencing. Two BAC clones containing the *ASIP* coding region were identified (139g8 and 188e3), grown in a 500-ml culture, purified with a Qiagen Large-Construct Kit, and analyzed by *Hind*III digestion for quality assessment. BAC clone 188e3 was selected for detailed characterization using a random shotgun sub-cloning approach (Yuhki et al., in press). This clone was randomly sheared by nebulization, and resulting DNA segments had their ends repaired and phosphorylated using Klenow fragment and T4 polynucleotide kinase. DNA fragments were size-selected (between 1.6 kb and 3 kb) in a 1% agarose gel, electro-eluted using a dialysis tube (Gibco BRL), and ligated to pBS KS^+ vectors (Stratagene) treated with *Eco*RV and CIAP (calf intestine alkaline phosphatase). The resulting shotgun library was electroporated into *E.coli* DH10B competent cells (Gibco BRL) and spread onto plates containing ampicillin and X-gal/IPTG. Nine-hundred and sixty positive colonies were transferred to liquid cultures in 96-well 2-mL deep plates. Cultures were grown overnight at 37°C in Superbroth, and plasmids were purified using a 96-well mini-preparation protocol modified from Ng et al. (1996) as described by Yuhki et al. (in press). Shotgun clones were sequenced from both vector flanks in 96-well format with BigDye chemistry, followed by purification using Sephadex G-50 plates (Amersham Pharmacia) and analysis in an ABI 3700 automated sequencer. Sequences were analyzed using a BioLIMS environment (PE informatics), including base calling, quality assessment, elimination of vector and *E. coli* sequences, and assembly into contigs using PHRED, PHRAP and CONSED (Ewing et al 1998; Gordon et al. 1998).

Sequence variation in *ASIP* intron 2 among felids

Initial characterization of *ASIP* intron 2 revealed considerable nucleotide variation among felid species. To further assess nucleotide variability in this segment among felid species, and to survey the occurrence of intra-specific diversity of potential use in population-level studies, primers agoEx2-F2 and agoEx3-R1 (Table 1) were employed to generate additional sequences from this region. These primers were designed based on the domestic cat genomic sequence obtained from BAC clone RPCI86-188e3, and amplified a 2-kb product that included exons 2 and 3, as well as the complete intron 2 (Figure 2-1). This PCR reaction used *AmpliTaq* Gold and 1.5 mM $MgCl_2$, with a touchdown profile consisting of 10' initial denaturation at 95°C, 12 cycles with decreasing annealing temperature [3 at 60°C, 3 at 58°C, 3 at 56°C, 3 at 54°C] followed by 30 cycles with 52°C annealing temperature, and a 7' final extension at 72°C. Amplification products were purified with exonuclease I (Amersham) and shrimp alkaline phosphatase (Amersham), and sequenced with BigDye terminator reactions using the PCR primers as well as internal oligonucleotides agoEx2-R1, agoF4, agoR1, agoR3, agoR4, agoR7, agoRH-F1, agoISB-F1 and agoISB-R1 and agoEx3-F1 (listed in Table 2-1). Sequencing products were purified with Sephadex G-50 plates or tubes (AutoSeq G-50) (Amersham Pharmacia) and analyzed with an ABI 3700 automated sequencer. Sequences were manually inspected and variants identified using SEQUENCHER (Gene Codes).

Sequence analysis

Genes and conserved non-coding segments contained in the domestic cat genomic sequences were identified through BLAST and mVISTA (Dubchak et al. 2000; Mayor et al. 2000) comparisons with sequences available in GenBank. BLAST analyses included both global searches of the non-redundant database and pairwise alignments with specific segments of the human genome. To position and orient the domestic cat genomic contigs, large-scale alignments with the human and mouse homologous regions were produced using mVISTA. These analyses were also employed to assess spatial patterns of sequence conservation in the *ASIP* genomic region. The expectation in such comparisons is that over time homologous genomic sequences will tend to accumulate mutational differences from their shared ancestral sequence, resulting in progressive erosion of their similarity. Thus, if sufficient time elapses, homologous genomic sequences will tend to become sufficiently different to preclude recognition or alignment of equivalent segments. It is usually hypothesized that regions under functional constraint (coding or regulatory segments) will accumulate changes more slowly than those evolving under neutrality, and thus the identification of conserved non-coding sequence blocks between divergent taxa may indicate the existence of relevant functional constraint (e.g. regulatory activity) in those genomic areas (Tagle et al. 1988; Gumucio et al. 1996). As two-species comparisons may identify spurious similarity due to stochastic lack of mutations in a given segment, it is often recommended that more taxa are compared in order to identify segments that are potentially relevant for regulatory functions (Dubchak et al.

2000). In the present study, three taxa (cat, human and mouse) were used in such comparisons for the *ASIP* genomic region.

Repeat elements contained in the domestic cat genomic sequence were identified and masked out (excluded from large-scale comparisons) using RepeatMasker (A.F.A. Smit & P. Green, unpublished data; available at http://ftp.genome.washington.edu/cgi-bin/RepeatMasker). Microsatellite loci consisting of perfect repeat arrays with at least 10 units (which are more likely to be polymorphic) were identified in these genomic segments using the Microsatellite Target Identification Program (R. Stephens and V.A. David [NCI-Frederick], unpublished), and resulting matches were verified manually against the original contig sequences.

The *ASIP* coding region (and also that of the neighboring gene *AHCY*) of multiple mammals and the set of *ASIP* intron 2 sequences among felids were aligned using ClustalX (Thompson et al. 1997) and visually corrected. Exploratory sequence analyses were conducted using MEGA 2.1 (Kumar et al. 2001), including evaluation of nucleotide and amino acid variability, calculation of synonymous (dS) and nonsynonymous (dN) differences among taxa, and estimation of phylogenetic relationships. Spatial variation in nucleotide variability along *ASIP* coding sequences was assessed using a sliding window approach as implemented in DnaSP (Rozas & Rozas 1999).

From the dS and dN estimates, the rates of synonymous (r_S) and nonsynonymous (r_N) substitutions among lineages were calculated for both *ASIP* and *AHCY* using mammalian evolutionary calibrations previously applied to other

genomic loci (Li 1997; Makalowski & Boguski 1998). Standard errors (S.E.) around these estimates were calculated using 1000 bootstrap replications as implemented in MEGA 2.1. Ninety-five percent confidence intervals (95 % C.I.) of computed substitution rates were derived by adding and subtracting two S.E. from the calculated point estimates. Estimates of r_N and r_S were statistically contrasted through a direct comparison of their confidence intervals, i.e. two point estimates were considered to be significantly different if their confidence intervals did not overlap. To correct for multiple comparisons, a Bonferroni adjustment was applied (Sokal & Rohlf 1995) by dividing the proposed significance level (0.05) by the number of comparisons. To meet the adjusted significance level, 99.7% confidence intervals comprising three S.E. above and below the mean were used for statistical testing (see Results).

Phylogenetic analyses were performed with PAUP 4.0 (Swofford 1998), including maximum-likelihood (ML), maximum-parsimony (MP) and minimum-evolution (ME) approaches. ML analyses included estimation of the best-fit model of DNA sequence evolution (and its required parameters) using a likelihood-ratio method (Whelan et al. 2001), and phylogenetic inference using a heuristic search starting from a Neighbor-Joining [NJ: Saitou & Nei 1987] tree followed by tree-bisection-reconnection (TBR) branch-swapping. MP analyses used heuristic searches with 50 replicates of random taxon-addition and TBR branch-swapping. ME phylogenies were built with NJ followed by TBR branch-swapping. Bootstrap support for each of the three methods was assessed using 100 replicates and the same search parameters listed above.

Results

Mapping of the domestic cat *ASIP* gene

The radiation hybrid (RH) analysis indicated that the domestic cat *ASIP* gene maps to the short arm of chromosome A3 (Figure 2-1A), very close (7.9 cR_{5000} distance) to the previously mapped *S-adenosylhomocysteine hydrolase* (*AHCY*) locus (Murphy et al. 2000). This position is homologous to the human (chromosome 20) and mouse (chromosome 2) location of the *ASIP* gene (Kwon et al. 1994; Miller et al. 1994; Wilson et al. 1995), and the proximity of *ASIP* and *AHCY* is observed in these species as well. The position of *ASIP* was compared to those of the closest available STR loci previously mapped onto the RH panel, FCA514 and FCA102 (Figure 2-1A). FCA514 mapped 118.6 cR_{5000} proximal from *ASIP*, which suggests a genetic distance of *ca.* 31.2 cM between these loci (based on the calculated correlation between physical and genetic distances in the cat genome; Murphy et al. 2000). FCA102 was found to be even more distant, mapping 178.3 cR_{5000} (estimated 46.9 cM) distally from *ASIP* (Figure 2-1A).

A separate mapping experiment for *ASIP* was performed with the inter-specific back-cross (ISB) pedigree between the domestic cat and the Asian leopard cat previously used to build a recombination map of the feline genome (Menotti-Raymond et al. 1999). The goals were to verify consistency with the RH results, obtain a direct estimate of the recombination distance between *ASIP* and STR loci FCA514 and FCA102, and also to allow a comparison with a third STR locus, FCA080, which had not been placed in the RH map. FCA080 has been previously estimated to be closely

linked to FCA514 in the ISB linkage map (Menotti-Raymond et al. 1999), and therefore should also be located in the same genomic region as *ASIP*.

A direct count of recombinants in the ISB experiment produced results consistent with the RH analysis. The observed recombinant fraction (θ) relative to *ASIP* was 0.33 for FCA514 (48 informative meioses), 0.32 for FCA080 (47 informative meioses), and 0.56 for FCA102 (25 informative meioses). Only one recombinant was observed between STR loci FCA514 and FCA080 (45 informative meioses; $\theta = 0.02$). Analyses using FASTLINK produced no significant linkage results (i.e. LOD score ≥ 3) between *ASIP* and any of these STR loci. Only negative LOD scores were obtained between *ASIP* and FCA102, consistent with the latter's distant chromosomal location from the gene of interest. Positive though non-significant LOD scores were observed between *ASIP* and the two other STR loci, with an identical peak at θ=0.3 (LOD score = 1.18) for both FCA514 and FCA080. A control analysis using the same data set showed significant linkage between the two adjacent STR loci FCA080 and FCA514 (peak LOD score=18.87 at $\theta = 0.05$). These results supported the RH estimate that *ASIP* is located *ca.* 30 cM away from the closest previously available STR loci (Figure 2-1A), and indicated that additional genomic markers more closely linked to *ASIP* had to be generated to allow reliable and efficient mapping of phenotypes relative to this gene.

Characterization of a domestic cat BAC clone containing the *ASIP* gene

A total of 1,920 shotgun reads (bi-directional sequencing of 960 sub-clones) was obtained from the domestic cat BAC clone RPCI86-188e3, representing *ca.* 7X

coverage based on its estimated insert size of 138 kb (using a comparison with the homologous human segment; Figures 2-1B and 2-2). After removal of low quality reads and sequences derived from the vector and the *E. coli* genome, 23 sequence contigs (comprising 124,730 bp) were assembled, and 925,993 bp remained as 626 singletons. Contigs 2 and 4 were short (< 1.7 kb) and contained only repetitive sequences, and were thus excluded from the analyses. The remaining 21 contigs (ranging in size from 505 bp to 30,564 bp after trimming) were surveyed for gene and repeat content (Table 2-2), and ordered on the basis of mVISTA and BLAST genomic comparisons with the homologous regions in human and mouse. Twelve contigs contained conserved sequences (after repeat masking) that allowed reliable positioning relative to the human segment (Figures 2-1B and 2-2). Much lower nucleotide similarity was observed with respect to the mouse segment (Figure 2-3), allowing the positioning of only five cat contigs (14, 22, 13, 3 and 16; see Figures 2-1B and 2-3).

The *ASIP* coding region was contained in contigs 14 (exon 2) and 22 (exons 3 and 4), which could be merged into a single contiguous segment (Figure 2-1B) by PCR-based sequencing of intron 2 from the same individual (Fca273) used for construction of the RPCI86 BAC library (Beck et al. 2001). The genomic region containing the domestic cat *ASIP* coding exons spans 4,387 bp, of which 1,651 bp are comprised by intron 2, and 2,328 bp by intron 3 (Figure 2-1C). The upstream non-coding exons of *ASIP* were not found to be contained in the examined contigs, and are probably beyond the genomic segment included in clone RPCI86-188e3 (see Figure 2-3). Two segments of *ASIP* intron 1 were included in the identified sequences: the complete contig 15 (2,987 bp) and the initial 1830 bp of contig 22 (Figure 2-1B, Table

2-2). The conserved 'GT' and 'AG' splice sites were observed in all *ASIP* intron-exon junctions.

The neighboring gene *AHCY* was found to be fully contained in this BAC clone (Figures 2-1B, 2-2 and 2-3), and positioned only 15.3 kb 3' of *ASIP*. *AHCY* is coded by the complementary DNA strand relative to *ASIP*, as also observed with the human and mouse homologues (Figures 2-2 and 2-3). Genomic segments homologous to seven of the nine *AHCY* exons could be identified (Figure 2-2), allowing the characterization of 381 of its 432 codons in the domestic cat (described in the next section). The 5' region (including the first two exons) of the *ITCH* gene (E3 ubiquitin protein ligase, homologous to the mouse *itchy* locus [Perry et al. 1998], and also named atrophin-1 interacting protein 4 [AIP4]) was also within this domestic cat BAC clone, based on alignment with the human genomic sequence (Figures 2-1B and 2-2). A comparison between the cat contigs and the mouse genomic sequence did not detect any significant similarity in the *ITCH* region (named LOC228810 in the mouse genomic annotation; see Figure 2-3). However, the mouse *itchy* locus does map to that location (Perry et al. 1998), and an alignment between the same human and mouse genomic sequences in this region revealed perfectly defined peaks of identity for all *ITCH* exons in this segment (Figure 2-4). The lack of alignment between cat and mouse *ITCH* regions is therefore likely due to the overall lower level of sequence conservation in this taxon comparison (see Figures 2-2 and 2-3), and to the increased difficulty in identifying significant similarity using an unordered cat contig bearing only short segments of high identity with the mouse homologous segment.

Several non-coding genomic segments exhibited high sequence similarity between cat and human (Figure 2-2). These included portions of the *ASIP* introns surrounding exons 2 and 3, a conserved segment in intron 1 (*ca.* 6 kb upstream of the *ASIP* coding region), and blocks located between *ASIP* and *AHCY*. Considerable sequence conservation was also detected in several blocks contained in *AHCY* and *ITCH* introns, and in the intergenic space between *AHCY* and *ITCH* (Figure 2-2). Very limited sequence similarity in these genomic segments was observed in the cat-mouse alignment, although some conserved blocks between *ASIP* and *AHCY* were identifiable in this comparison (Figure 2-3), and also in that between human and mouse (Figure 2-4).

Numerous genomic repeats were identified in the cat sequences, including SINEs, LINEs, LTR and DNA elements, and also short tandem repeat (STR) loci (Table 2-2). As a whole, repetitive sequences composed *ca.* 45 % of the cat genomic segment included in this BAC clone. A direct count of repeats (including only LINEs, SINEs and LTR and DNA elements) in the 20 contigs listed in Table 2-2 indicated a density of 1.9 integrations per kilobase, of which 56% were SINEs, 28% LINEs, 8% LTRs and 8 % DNA elements. The human and mouse homologous segments presented a similar density of repeats (2.2 to 2.3 insertions/ kb), slightly higher than that observed in the cat sequence. Both contained a higher proportion of SINE elements (72% in mouse, 66% in human) compared to the cat, and a lower proportion of LINE integrations (18% in human, 10% in mouse). The mouse segment exhibited the highest proportion of LTR elements (15%), and the lowest density of DNA elements (3%).

A targeted search for long, uninterrupted microsatellites (STRs) in the domestic cat *ASIP* genomic region yielded 20 loci, numbered FCA701 to FCA720 (Table 2-3). Sixteen of these loci contained dinucleotide repeats (AC [TG], GA/AG [CT/TC], and AT), three consisted of tetranucleotide repeats (CTTT [AAAG]), and one was a trinucleotide repeat (AAC). The most common STR motif in this region was GA/AG [CT/TC], comprising 11 of the 20 loci, followed by AC [TG], with four loci (Table 2-2). Interestingly all three of the identified tetranucleotide loci were composed of the motif CTTT [AAAG] (Table 2-3). Fifteen of the 20 identified STR loci could be ordered in the cat genomic sequence based on the position and orientation of the contigs in which they were located (Figure 2-1B). STR loci FCA701, 703, 707, 708 and 709 were located in sequence contigs whose location within the BAC clone could not be established (Table 2-3).

Evolutionary analyses of the *ASIP* coding region

The domestic cat *ASIP* coding region consists of 405 bp (135 codons), and is similar in structure to previously described mammalian homologues (Figures 2-5 and 2-6), though exhibiting many substitutions at nucleotide and amino acid levels, as well as a three-residue insertion at codon position 85 (position 86 in Figures 2-5 and 2-6). Overall nucleotide similarity was 83% to 84% with rat and mouse, 85% with human, 88% with pig and horse, and 90% with cow and fox. Amino acid identity was 77% relative to human, 78% with rat, 81% with mouse and horse, 85% with cow and pig, and 86% with fox.

The GC content in the cat *ASIP* coding region (59.3%) was similar to that computed for other mammals (overall mean 58.9%), and higher than in neighboring non-coding segments: 44 % in the 3' end of intron 1 (1,830 bp); 46% in intron 2; 41% in intron 3; and 43% in the genomic region between the *ASIP* and *AHCY* coding sequences (16,109 bp). A comparative analysis of GC content in the three codon positions was performed for *ASIP* and *AHCY*, using the average among cat, mouse, rat and human (the only mammalian sequences available for *AHCY*). *ASIP* contained 48.2% GC in the first, 52.8% in the second, and 75.8% in the third codon position. *AHCY* had 58.1% GC in the first, 37.9% in the second, and 72.5% in the third position. The overall GC content in the *AHCY* coding region was 56.2%, slightly lower than that of *ASIP*.

Considerable variability was observed among the eight mammalian *ASIP* sequences compared. Of 408 aligned nucleotide sites 143 (35%) were variable, as were 59 (43%) out of 136 amino acid residues. Variability was not distributed homogeneously along the *ASIP* coding region, which showed short conserved motifs interspersed with highly diverse segments (Figures 2-5, 2-6 and 2-7). Overall there were no continuous segments spanning more than 16 bp or 7 amino acids in the *ASIP* coding region that were completely conserved across the eight mammals. A highly variable region of *ASIP* was located between nucleotide coding positions 240 and 290 (Figures 2-5 and 2-7), at the boundary between the basic (lysine- and arginine-rich) and proline-rich central domains, and flanking the segment where most of the insertions / deletions were observed (Figure 2-5). At the amino acid level this region is also considerably variable, but equally high diversity was observed in portions of

the signal peptide and the 'mature N-terminus' as well (Figure 2-6). Interestingly, all amino acid residues in which replacements have been experimentally shown to cause loss of ASIP function are completely conserved across mammals (Figure 2-6). Most of the sites where changes have decreased (but not abolished) ASIP activity are also completely conserved (see Discussion).

To assess the effect of functional constraints on *ASIP* evolutionary rates, synonymous (dS) and nonsynonymous (dN) substitutions among four taxa (cat, human, rat and mouse) were compared in this gene and the nearby *AHCY* locus, which can be assumed to be subjected to a similar mutational environment. The estimated dS was similar for both genes in all taxon comparisons, whereas dN was markedly higher in *ASIP* (Figure 2-8). The mean dN/dS ratio was 0.34 for *ASIP* and 0.05 for *AHCY*.

Rates of synonymous (r_S) and nonsynonymous (r_N) substitutions were estimated for both loci using three evolutionary calibrations (Table 2-4), allowing a comparison between the two genes and also relative to other genomic loci (see Discussion). For each taxon pair shown in Table 2-4 (rat-human, mouse-human, rat-mouse), four statistical comparisons of substitution rates were performed: (i) *AHCY* r_N *versus* r_S; (ii) *ASIP* r_N *versus* r_S; (iii) *AHCY* r_S *versus ASIP* r_S; and (iv) *AHCY* r_N *versus ASIP* r_N. Rate comparisons between taxon pairs were not performed, as they are not phylogenetically independent and their difference can be strongly influenced by the applied calibration dates. The Bonferroni correction for 12 comparisons (four in each taxon pair) yielded an adjusted significance level of 0.004167. To assess the existence of significant differences in rate estimates while meeting this adjusted significance level, the 99.7% confidence interval (3 X S.E. above or below the mean) was used.

Non-overlapping 99.7% C.I.'s were considered to indicate strong evidence for significant difference (at $\alpha = 0.05$) using a conservative test (as this requirement is slightly more stringent than the Bonferroni-corrected significance level).

Using this statistical approach, *AHCY* exhibited significantly higher ($P<0.05$) r_S compared to r_S in all three taxon comparisons (Table 2-4). In contrast, r_S and r_N were not significantly different in *ASIP* after the Bonferroni correction. In the rat-human comparison, *ASIP* r_S and r_N were not different even considering the uncorrected 95% C.I. of each estimate. For the mouse-human and mouse-rat comparisons, the uncorrected 95% test indicated significant differences between r_N and r_S in *ASIP*, while the more stringent adjusted 99.7% C.I. test did not. Estimates of r_S were not significantly different between *AHCY* and *ASIP*. Estimates of r_N were significantly lower in *AHCY* versus *ASIP* in the rat-human and mouse-human comparisons, but not in the rat-mouse comparison (likely due to the higher coefficient of variation observed in this taxon comparison).

Molecular evolution of *ASIP* sequences in the Felidae

The genomic sequence from *ASIP* intron 2 and exons 2 and 3 (total segment size: 1.8 kb to 2 kb) was obtained for 33 individuals representing 12 felid species: domestic cat (*Felis catus*, n=11); Pallas cat (*Otocolobus manul*, n=1); caracal (*Caracal caracal*, n=1 [incomplete]), African golden cat (*Profelis aurata*, n=1); puma (*Puma concolor*, n=3); cheetah (*Acinonyx jubatus*, n=1); Geoffroy's cat (*Oncifelis geoffroyi*, n=5); pampas cat (*O. colocolo*, n=2); oncilla (*Leopardus tigrinus*, n=1); leopard (*Panthera pardus*, n=4); Asian golden cat (*Catopuma temmincki*, n=2); and marbled

cat (*Pardofelis marmorata*, n=1). The intron-2 SINE element insertion used above for mapping *ASIP* (Figure 2-1C) was found to be present in all surveyed domestic cats, but was absent from all examined individuals from other felid species. A remnant from a second SINE insertion was present in all surveyed species, adjacent to a GT-rich repetitive region showing extensive variability among taxa. After exclusion of the domestic cat-specific SINE insertion and other segments of ambiguous alignment, a final data set containing 1,750 bp was used to assess levels of variability and phylogenetic information present in this *ASIP* segment.

High levels of variation were observed across felid species, including 147 variable sites and 106 parsimony-informative characters. Pairwise divergence among species ranged from 0% (between the caracal and African golden cat, which had identical sequences) to 3.4% (between the pampas cat and the Pallas cat). Intra-specific variation was observed in the domestic cat (3 single-nucleotide polymorphisms [SNPs]), puma (2 SNPs), Geoffroy's cat (5 SNPs), leopard (2 SNPs), and Asian golden cat (4 SNPs). Phylogenetic analyses using three different approaches produced identical topologies (Figure 2-9), and provided a good resolution of the relationships among the included felid species. All species represented by multiple individuals formed monophyletic clusters (not shown), and previously defined lineages of the Felidae (*C. caracal* + *P. aurata*; *P. concolor* + *A. jubatus*; *L. tigrinus* + *O. geoffroyi* + *O colocolo*; see Chapter 1) were recovered with high support. This data set provided strong evidence for the relative placement of *O. manul* near the Domestic cat lineage (represented by *F. catus*), and *P. marmorata* near the Bay cat

lineage (represented by *C. temminckii*), and also for the relationships among six of the eight major felid clades (see Discussion).

A 14 bp insertion / deletion (indel) occurred at position 123 of intron 2, which was shared by all surveyed domestic cats and also by a Pallas cat (Figure 2-10). To assess the polarity of this indel and survey its presence throughout the Felidae, partial sequences of this segment of intron 2 were examined for one individual of each of nine additional felid species (European wild cat [*Felis silvestris*], black-footed cat [*F. nigripes*], leopard cat [*Prionailurus bengalensis*], flat-headed cat [*P. planiceps*], bobcat [*Lynx rufus*], serval [*Leptailurus serval*], ocelot [*Leopardus pardalis*], jaguarundi [*Herpailurus yaguarondi*] and jaguar [*Panthera onca*]), as well as a feloid carnivore outgroup (binturong [*Arctictis binturong*], Carnivora: Viverridae). This analysis showed that the absence of the indel segment is the derived state (i.e. it is a deletion), and that it is shared by all surveyed members of the Domestic cat lineage (*F. catus*, *F. silvestris*, *F. nigripes*), Leopard cat lineage (*P. bengalensis*, *P. planiceps*) and Lynx lineage (*L. rufus*), in addition to the Pallas cat (*O. manul*). Two additional informative deletions were identified in this segment (not shown), one of them (6 bp) exclusive to all surveyed members of the Puma lineage (three pumas and one cheetah), and another (1 bp) shared by all domestic cats, the Pallas cat, caracal and African golden cat (only species shown in Figure 2-9 were available for this segment).

Discussion

Development of genomic tools for the study of *ASIP* in cats

One goal of the present study was to characterize the *ASIP* locus in the domestic cat as a candidate gene for involvement in melanistic phenotypes in this and other felid species (see Chapter 1). Mapping of *ASIP* using the RH panel and the ISB pedigree confirmed that the identified sequences corresponded to the characterized human and mouse homologues. Since there were no feline STR markers available in the immediate vicinity of this locus, as would be required for efficient mapping of the *ASIP* region in domestic cat pedigrees segregating for melanism, 20 new markers were identified and characterized in this segment, providing several potential candidates for polymorphism screening and linkage studies (described in Chapter 4). In addition to their use as markers for linkage analysis of melanism, the identification of multiple ordered STR loci in a short genomic region allows the prospect of haplotype-based analyses of segment evolution in the domestic cat and other felids (discussed in Chapter 5).

In order to screen the *ASIP* coding region for variants potentially involved in cat melanism (Chapter 4), a full characterization of the three included exons and adjacent genomic segments was required. Sequences flanking each coding exon were needed to design primers suitable for PCR-based screening of cat *ASIP* variants, since most available samples for melanism studies in felids consist of genomic DNA. The sequences obtained from BAC clone RPCI86-188e3 encompassed the complete *ASIP* coding region, and allowed the design of primers flanking each of its three exons. Three of these primers have been used in this chapter to amplify and/or sequence an

ASIP segment including intron 2 and exons 2 and 3. The remaining exon-specific primers will be described in Chapter 4.

Comparative analysis of the *ASIP* genomic region

Assembled shotgun sequences from the feline BAC clone RPCI86-188e3 produced 20 final contigs that were used to characterize the domestic cat genomic region containing the *ASIP* coding exons. Genomic alignments with human and mouse homologous regions indicated that the upstream non-coding exons that have been implicated in differential *ASIP* transcription in ventral and dorsal regions of the body (Bultman et al. 1994; Vrieling et al. 1994) were not contained in this BAC clone, precluding their characterization (and that of *ASIP* promoter elements) in the domestic cat. The sequencing of this upstream *ASIP* region in the cat requires the identification of one or more additional BAC clones, and will be pursued in future studies.

Although the promoter elements located 5' of the *ASIP* gene could not be characterized in this study, conserved sequence blocks identified 3' of this gene may have regulatory functions with respect to this locus. Miller et al. (1994) reported that the lethal nonagouti (a^{x}) mouse mutation was caused by a 100-kb deletion 3' of *ASIP* that removed the *AHCY* locus (which was the inferred cause of homozygous lethality). The hypomorphic effect of this mutation (darker coat color when heterozygous with loss-of-function *agouti* alleles) was inferred to be caused by loss of regulatory elements located 3' of the *ASIP* coding region, which would be required for normal *ASIP* expression. However, these regulatory motifs have so far not been identified, and no further study on the genomic region between the *ASIP* and *AHCY* genes has

been reported. This potential regulatory segment is located 3' of both *ASIP* and *AHCY* transcriptional units, as they are coded in opposite directions (Figures 2-2 and 2-3). In this study, nine sequence blocks (ranging in size from 87 bp to 163 bp) conserved between cat and human (75% to 82% nucleotide identity) were identified in this region (Figure 2-2), providing candidate regions for studies of transcriptional regulation of *ASIP* and possibly also *AHCY*. Comparisons of this genomic segment in cat *versus* mouse (Figure 2-3) and human *versus* mouse (Figure 2-4) identified fewer conserved blocks, which is likely due to the well-established accelerated rate of nucleotide substitution in the mouse and other murid rodents (e.g. Li 1997; Eizirik et al. 2001b; Chapter 3). The same murid acceleration is also suggested by comparisons of coding region substitutions for both *ASIP* and *AHCY* (Table 2-4). Two of the identified conserved blocks (located *ca.* 6 kb 3' of the cat *ASIP* gene) were observed in both human-cat and human-mouse comparisons, suggesting that these segments are likely to be functionally relevant.

Multiple conserved segments were also observed between the *AHCY* and *ITCH* loci (Figure 2-2). Since this region is located 5' of both genes, it is likely that these conserved sequence blocks include the majority of the *cis*-acting elements required for regulation of these loci. Conserved non-coding segments were also observed in introns of *ASIP*, *AHCY* and *ITCH* (Figures 2-2, 2-3, 2-4), suggesting that these regions may also be involved in regulatory functions over these loci. These observations can be further refined by additional comparative genomic analyses using more taxa and varying search parameters, and ultimately by functional studies assessing the relevance of these DNA segments for the regulation of these genes.

Evolution of the *ASIP* gene in mammals

Abundant sequence variation was observed in the *ASIP* gene among eight mammal species, including synonymous and nonsynonymous substitutions and insertions / deletions that seemed to be concentrated in particular segments (Figures 2-5 and 2-6). The overall pattern is suggestive of a gene under relaxed functional constraints, and/or that experiences episodes of positive selection driving the diversification of particular segments. To evaluate these hypotheses, and to directly compare the levels of sequence conservation observed in the whole *ASIP* coding region to those of other loci, rates of synonymous (r_S) and nonsynonymous (r_N) substitutions at this gene were estimated for three well-characterized mammalian species-pairs (Table 2-4), and compared to previous estimates for other genes (Li 1997; Makalowski & Boguski 1998). Substitutions among available taxa were also computed for the neighboring locus *AHCY* (Figure 2-8), and derived rates were compared to those at *ASIP* for the three reference divergences (Table 2-4).

The estimated synonymous rates were similar between *ASIP* and *AHCY* for the three reference divergences (no significant difference detected, see Table 2-4), and were below the average reported for 47 human-rodent (mouse/rat) comparisons (Li 1997), as well as for 470 mouse-rat and 1,880 human-rodent pairs analyzed by Makalowski & Boguski (1998). In contrast, nonsynonymous rates varied dramatically between *ASIP* and *AHCY* (Table 2-4; see also Figure 2-8). *ASIP* exhibited rates that were slightly above the average computed among loci (mean human-rodent $r_N = 0.74$ in Li [1997]; mean $r_N = 0.49$, 0.55 and 1.05 in Makalowski &

Boguski [1998] for rat-human, mouse-human and rat-mouse, respectively). In contrast, *AHCY* showed r_N estimates that were at least 5-fold lower than the averages among loci (Table 2-4).

The expectation under neutral amino acid evolution is that the rates of synonymous and nonsynonymous substitutions will not be significantly different. Under negative selection (functional constraint), nonsynonymous changes will be maintained during evolution less frequently than synonymous substitutions, leading to significantly lower r_N relative to r_S. Relaxation of functional constraints leads to a pattern of nucleotide substitution that approaches neutral expectations, but where nonsynonymous changes may still be less frequent than synonymous substitutions in case portions of the segment in question are constrained under negative selection. Finally, under positive or diversifying selection the rate of replacement (nonsynonymous) changes is expected to be higher than that of silent (synonymous) substitutions. The data obtained here for *ASIP* and *AHCY* allowed an evaluation of these expected patterns for these two loci.

Since *ASIP* and *AHCY* are located only *ca.* 15 kb apart in the cat genome (*ca.* 9 kb in mouse and 11 kb in human) and do not exhibit strong differences in GC content or synonymous rates, it may be assumed that both genes are located in a similar mutational environment. This suggests that their difference in nonsynonymous rates is mostly due to differential effects of natural selection acting on these genes (i.e. relaxed functional constraints or positive selection in *ASIP*). The results shown in Table 2-4 and Figure 2-8 indicate that *AHCY* is under strong negative selection (functional constraint), supported by significantly lower nonsynonymous than synonymous

substitution rates. This is not surprising, as *AHCY* codes for an enzyme required for basic cellular metabolism (Miller et al. 1994). In contrast, *ASIP* exhibited a lower level of functional constraint, indicated by non-significant differences in r_N *versus* r_S. However, r_N was consistently lower than r_S in *ASIP* (Figure 2-8 and Table 2-4), indicating the existence of some level of functional constraint in this gene. No evidence of positive selection acting on *ASIP* was detected, since there was no instance in which r_N was higher than r_S. It is important to note that these estimates are based on the relationship between r_N and r_S averaged over the whole coding region, and therefore may include segments that fall under each of the three possible categories (negative selection, neutrality or near-neutrality, and positive selection). Further analyses comparing individual sites should be performed to evaluate this inference in more detail.

In summary, *ASIP* cannot be considered to be highly conserved in the context of evolutionary rates in mammals, as it exhibits rates of replacement substitutions that are similar to or faster than the average observed among many loci. The pattern observed in *ASIP* is in agreement with relaxation of functional constraints in many parts of the gene, although conservation of particular motifs was also apparent. Importantly, all amino acid residues that have been shown to be critical for maintenance of *ASIP* function are completely conserved among the surveyed mammals (Figure 2-6). This includes sites where loss-of-function replacements have been identified (Perry et al. 1996; Miltenberger et al. 2002), and also a core motif containing four residues within a loss-of-function deletion induced in the signal peptide (Perry et al. 1996). Furthermore, most other residues at which mutations have

induced only partial loss of ASIP function are also completely conserved in this data set (Figure 2-6), including a putative N-glycosylation site and adjacent serine residues, the valine at position 83 reported by Kiefer et al. (1997) to be important for MC1R inhibition, and the three additional cysteines mutated by Perry et al. (1996). These observations suggest that these sites are very likely under negative selection (constraint) in all included mammalian groups, and that evolutionary comparisons can be useful to guide further experimental investigations of residues potentially implicated in ASIP function.

An interesting feature of the *ASIP* alignments shown in Figures 2-5 and 2-6 is occurrence of multiple 'in-frame' insertion / deletion (indel) events among mammalian lineages, in most cases seeming to concentrate in particular segments. In the proline-rich central domain, suggested by Miltenberger et al. (2002) to potentially act as a "hinge" in ASIP folding, a deletion of two residues seems to have occurred in the murid ancestor, and an insertion of three codons was identified in the domestic cat sequence. Given the potential functional roles suggested for this portion of ASIP (Kiefer et al. 1997; Virador et al. 2000; Miltenberger et al. 2002), and the known effect of proline residues on protein structure (Li 1997), it would be important to investigate whether these indel variants actually modify ASIP activity. Most other observed indel variation occurred at one site in the basic central domain (positions 70-71 in Figure 2-6), where at least four separate evolutionary events are required to explain the observed pattern (given the mammalian phylogeny presented by Murphy et al. [2001]). Independent gains or losses of one or two additional basic residues (lysine/arginine) have occurred at this site, indicating either reduced functional

relevance of the specific number of basic amino acids or positively selected changes, perhaps in compensation for other events occurring in this domain. Deletion of this *ASIP* region (shown by a box in Figure 2-6) did not seem to cause any loss of function in transgenic mice (Perry et al. 1996), supporting the inference of relaxed functional constraint at these residues. On the other hand, particular substitutions occurring next to these indels (within the box) have been reported to cause reduced ASIP function in mice (Miltenberger et al. 2002). Additional work is required to further understand the biological role of this basic central domain (Miltenberger et al. 2002), and it is likely that comparative analyses of different taxa will help clarify the functional and evolutionary significance of observed patterns of variation.

Evolution of the Felidae inferred from *ASIP* intron 2 sequences

Previous studies based on DNA sequence data have identified eight major lineages in the Felidae (Johnson & O'Brien 1997; Pecon-Slattery & O'Brien 1998). However, their basal relationships have not been completely resolved, and the phylogenetic placement of four species (*Otocolobus manul*, *Pardofelis marmorata*, *Leptailurus serval* and *Prionailurus rubiginosus*) relative to these lineages has not been conclusively established (see Chapter 1, Figure 1-1). In this study, the utility of *ASIP* sequences to aid in the resolution of these questions was evaluated using a 1,750-bp segment of this gene analyzed in 33 individuals of 12 felid species, representing six of the eight major lineages and two of the unaligned taxa.

Phylogenetic analyses based on nucleotide substitutions in this segment produced a well-resolved tree that was consistent using different inferential methods

(Figure 2-9), and that provided evidence on the relationships among lineages and the placement of two additional species. The monophyly of the three lineages represented by more than one species (Caracal lineage: *C. caracal*, *P. aurata*; Puma lineage: *P. concolor*, *A. jubatus*; Ocelot lineage: *L. tigrinus*, *O. geoffroyi*, *O. colocolo*) was recovered with consistent support (Figure 2-9), in agreement with previous studies (Johnson & O'Brien 1997; Pecon-Slattery & O'Brien 1998; Mattern & McLennan 2000). The inferred topology suggested a close relationship between the Ocelot and Puma lineages, which has not been reported previously and should be investigated further using rooted analyses including additional species. Strong support was observed for the placement of the Pallas cat (*O. manul*) near the Domestic cat lineage, and for these two in turn to be related to the Caracal lineage. The proximity of the Pallas cat to the Domestic cat lineage has been inferred in previous studies (Collier and O'Brien 1985; Salles 1992; Mattern & McLennan 2000; Bininda-Emonds et al. 2001), however other analyses failed to find strong support for this relationship (e.g. Herrington 1986; Johnson & O'Brien 1997; Pecon-Slattery & O'Brien 1998). A shared-derived deletion identified in this segment (Figure 2-10) provided additional support linking the Pallas cat and the Domestic cat lineage, along with the Leopard cat and Lynx lineages. The marbled cat (*P. marmorata*) was found to be closely related to the Asian golden cat (*C. temmincki*, included in the Bay cat lineage), a relationship that has been suggested by some previous analyses (e.g. Pecon-Slattery & O'Brien 1998), albeit with lower support. These results provide some novel insights into the phylogeny of the Felidae, and indicate that sequence variation in this genomic segment should be investigated for all remaining cat species, as well as carnivore outgroups.

The domestic cat-specific SINE element identified in intron 2 (Figure 2-1C) may also be used as a phylogenetic character in future studies of the Felidae, with a more detailed investigation of the position in the felid tree where its integration has occurred. The present study suggests that it happened before the origin of current domestic cat populations, as it was found in all surveyed individuals (including random-bred animals from Brazil, Israel and the USA), but after the divergence of this species from the Pallas cat and leopard cat. A survey of the presence of this SINE element in other species of the domestic cat lineage will help determine the exact phylogenetic position of this genomic integration.

The identification of intra-specific variants (single nucleotide polymorphisms [SNPs]) in the *ASIP* intron 2 of several felids indicates that this genomic segment may also be useful for population-level studies in this group. A limited survey revealed the occurrence of well-defined SNPs in five of the surveyed cat species, so it is likely that further studies using expanded sample sizes will lead to the identification of additional variants in this segment for these and other felids. This may be particularly interesting in the context of utilizing genomic haplotypes to investigate the population history and adaptive significance of *ASIP* variants involved in coat color phenotypes (as described in Chapter 4).

Table 2-1. Primers utilized in this study.

Primer [a]	Sequence (5' – 3')
agoF1	GTCATCCGCCTACTCCTGGC
agoR1	TCCGCTTCTTTTCTGCTGATC
agoR3	AGTCCAGAGACCAGGGGTCA
agoF4	TTTTGGTTTAACTTGGCATT
agoR4	CATGCTGACAGTACTGCTTG
agoR7	TTCCCTGCTCCTTCCCTGAT
agoF3	AAAAAGGCTTCGATGAAGAA
agoR5	TCAGCAGGTGGGGTTGAG
agoRH-F1	GGGCCTGACATTGAACATCT
agoRH-R1	CACCTGGGTTTGAATTCTGG
agoISB-F1	CCGAGAGACCCTGAAGTCAA
agoISB-R1	CTCACTTCCCAGTGCCTAGC
agoEx2-F2	TTCTCTGTTCCACTCAGGCC
agoEx3-F1	TCCACTCCTCCCACTTTACTG
agoEx3-R1	CCCTTAGCTCTCTGGGCTTC

[a] For any primer, 'F' refers to forward, and 'R' to reverse.

Table 2-2. Features of the sequence contigs contained in the domestic cat BAC clone RPCI86-188e3. Asterisks indicate contigs whose location and orientation are identified in Figure 2-1B.

Contig	Size (bp) [a]	Genes included	Repeat elements [b]				
			SINE	LINE	LTR	DNA	STR
1	656		0	1	0	0	1
3 *	1,178	*AHCY*	1	0	0	0	0
5	505		0	0	1	0	0
6	625		0	0	0	0	0
7	700		0	0	0	0	0
8	618		2	0	0	0	2
9	840		1	2	0	0	0
10 *	637		1	0	0	0	1
11	740		3	1	0	0	0
12	1,144		1	1	1	0	1
13 *	2,422	*AHCY*	2	2	1	0	0
14 / 22 *	26,637	*ASIP*, *AHCY* (3' end)	24	3	2	6	7
15 *	2,987	*ASIP* (partial intron 1)	2	0	0	0	1
16 *	6,478	*AHCY*	7	7	0	0	0
17 *	5,537		7	1	2	1	1
18 *	6,621		10	7	1	0	7
19 *	8,883		3	5	3	2	1
20	13,323		16	14	2	0	3
21 *	13,775		16	7	3	2	4
23 *	30,564	*ITCH*	34	15	2	8	9
Total	124,870		130	66	18	19	38

[a] Size after trimming.

[b] Only SINEs, LINEs, LTR elements, DNA elements and STRs are listed. Small RNA elements were also identified in contigs 17 (1 insert), 18 (2 inserts) and 14/22 (1 insert).

Table 2-3. Microsatellite (STR) loci identified in shotgun sequence contigs of BAC clone RPCI86-188e3, containing the domestic cat *ASIP* coding region. Only uninterrupted repeat arrays with at least 10 units are included. Asterisks indicate loci mapped to a specific ordered location in the genomic clone (Figure 2-1B).

Contig	Locus	Repeat
1	FCA701	$(AC)_{18}$
10	FCA702 *	$(GA)_{15}$
12	FCA703	$(GA)_{18}$
18	FCA704 *	$(GA)_{12}$
18	FCA705 *	$(CTTT)_{16}$
18	FCA706 *	$(AG)_{12}$
20	FCA707	$(CT)_{12}$
20	FCA708	$(AC)_{22}$
20	FCA709	$(AAC)_{11}$
21	FCA710 *	$(AG)_{15}$
21	FCA711 *	$(GA)_{16}$
21	FCA712 *	$(TC)_{17}$
21	FCA713 *	$(AG)_{21}$
22	FCA714 *	$(AT)_{22}$
22	FCA715 *	$(GA)_{10}$
22	FCA716 *	$(TG)_{20}$
23	FCA717 *	$(AC)_{10}$
23	FCA718 *	$(AAAG)_{19}$
23	FCA719 *	$(AAAG)_{15}$
23	FCA720 *	$(TC)_{14}$

Table 2-4. Estimates of synonymous (r_S) and nonsynonymous (r_N) substitution rates for the *ASIP* and *AHCY* genes applying the method of Pamilo & Bianchi (1993) and Li (1993). Rates are shown as the number of estimated substitutions per site per 10^9 years. Rat-human and mouse-human rates were estimated assuming that primates and rodents diverged 80 million years ago (MYA) (to allow comparisons with Li [1997] and Makalowski & Boguski [1998]). Rat-mouse rates were estimated assuming a 15 MYA divergence between these species (as in Makalowski & Boguski [1998]; a similar date for this node is supported by recent molecular studies [Springer et al., in preparation]). Estimates were generated with MEGA 2.1, and standard errors were computed using 1000 bootstrap replications. All sites containing gaps or missing information were completely excluded from the data set.

	AHCY	*ASIP*
Rat – Human	$r_N = 0.0969 \pm 0.0281$	$r_N = 0.8244 \pm 0.1544$
	$r_S = 1.9256 \pm 0.2356$	$r_S = 2.0006 \pm 0.4363$
Mouse – Human	$r_N = 0.0956 \pm 0.0281$	$r_N = 0.6706 \pm 0.1319$
	$r_S = 2.0288 \pm 0.2413$	$r_S = 2.2013 \pm 0.4556$
Rat – Mouse	$r_N = 0.1833 \pm 0.0767$	$r_N = 1.1300 \pm 0.3967$
	$r_S = 4.8267 \pm 0.8200$	$r_S = 5.2533 \pm 1.4467$

Figure Legends

Figure 2-1. Schematic of the domestic cat *ASIP* genomic region. **A)** Mapping of the *ASIP* gene to cat chromosome A3, and relative location of the adjacent gene *AHCY* and previously published STR loci FCA080, FCA514 and FCA102. The dark circle represents the chromosomal centromere. **B)** Ordering of sequence contigs obtained from BAC clone RPCI86-188e3, showing the relative positions of the *ASIP*, *AHCY* and *ITCH* genes. Each horizontal arrow represents one contig (numbered underneath as in Table 2-2), and indicates its relative position and direction relative to the chromosomal orientation shown in [A]; contig length and spacing are not drawn to scale (see Figures 2-2 and 2-3). Vertical arrows represent domestic cat STR markers identified in this region (labeled above with number after the 'FCA' prefix; see Table 2-3). **C)** Schematic of the three *ASIP* coding exons (exons 2, 3 and 4, represented by shaded boxes) and intervening introns. The black circle represents a SINE element insertion identified only in the domestic cat. The triangle indicates the location of a 14-bp phylogenetically informative deletion shared by seven felid species (see text and Figure 2-10). The position of eight primers used for the sequence characterization of intron 2 and for *ASIP* mapping experiments is indicated by arrows labeled as follows: a: agoEx2-F2; b: agoF1; c: agoISB-F1; d: agoISB-R1; e: agoRH-F1; f: agoRH-R1; g: agoR1; h: agoEx3-R1. The remaining intron 2 primers are not shown. Primers labeled 'i' and 'j' are agoF3 and agoR5, respectively, used to amplify a segment of exon 4.

Figure 2-2. Comparison between the human and domestic cat *ASIP* genomic regions using mVISTA. The X-axis shows the nucleotide positions (in kilobases) of the

human sequence from chromosome 20 (GenBank accession number NT_028392, updated August 2002), including sites 3,000,000 (3,000 kb) to 3,150,000 (3,150 kb) of this genomic segment. The Y axis represents nucleotide similarity with the cat sequence, calculated using a 100-bp window size. Blue peaks in the plot indicate exonic regions, and peaks colored in pink represent conserved ($\geq$ 75% nucleotide identity) non-coding segments (intronic or inter-genic). The position and orientation of the *ASIP*, *AHCY* and *ITCH* genes are indicated by arrows on top, and exons are shown as blue boxes. Gray arrows at the bottom of each plot represent domestic cat genomic contigs, numbered as follows (order as in the Figure, 5' to 3'): **13:** contig 15; **12:** contig 14/22; **11:** contig 13; **14:** contig 16; **15:** contig 17; **19:** contig 21; **20:** contig 23. Unnumbered contigs were subsequently identified using pairwise BLAST comparisons between human genomic clones (accession numbers AL035458 and AL356299) and all remaining cat segments.

Figure 2-3. Comparison between the mouse and domestic cat *ASIP* genomic regions using mVISTA. The X-axis shows the nucleotide positions (in kilobases) of the mouse sequence from chromosome 2 (GenBank accession number NW_000179, updated June 2002), including sites 3,550,000 to 3,700,000 of this genomic segment (numbering for this plot was started at position 3,400,001 to simplify analysis, so the included segment is numbered 150 kb to 300 kb). The Y axis represents nucleotide similarity with the cat sequence contigs, calculated using a 100-bp window size. Labels and symbols are the same as in Figure 2-2. The arrow representing the *ASIP* gene is longer than in Figure 2-2 due to the inclusion of a distant 5' non-coding exon

of *ASIP* in the mouse genomic annotation; no sequence similarity with the cat contigs was identified in this upstream region of *ASIP*. The gene annotated as LOC228810 in the mouse corresponds to the *Itchy* locus (*ITCH* in humans, see Figure 2-4).

Figure 2-4. Comparison between the human and mouse *ASIP* genomic regions using mVISTA. The X-axis shows the nucleotide positions (in kilobases) of the same human sequence used in Figure 2-2, including sites 3,000,000 to 3,200,000 of this genomic segment. The Y axis represents nucleotide similarity with the mouse sequence from chromosome 2 (same as in Figure 2-3), calculated using a 100-bp window size. Labels and symbols are the same as in Figure 2-2.

Figure 2-5. Nucleotide alignment of the *ASIP* coding region including the domestic cat sequence and those of the red fox (*Vulpes vulpes*, GenBank accession number Y09877); horse (*Equus caballus*, AF288358); cow (*Bos taurus*, X99692); pig (*Sus scrofa*, AJ251837); human (*Homo sapiens*, NM_001672); mouse (*Mus musculus*, NM_015770); and rat (*Rattus rattus*, NM_052979.1). Dots indicate identity to top sequence; nucleotide positions are shown at the end of each line. Vertical lines demarcate the boundaries between exons 2 and 3 (after position 160), and between exons 3 and 4 (after position 225). Dashes (shaded) indicate insertions/deletions (indels).

Figure 2-6. Amino acid alignment of ASIP including the domestic cat sequence and those of seven other mammals (names and accession numbers as listed in Figure 2-5).

Dots indicate identity to top sequence; amino acid positions are shown at the end of each line. Vertical lines demarcate the boundaries among the five functional domains proposed for ASIP (Miltenberger et al. 2002), named above or below the sequence. Dashes represent insertions/deletions (indels). The dark circle indicates the putative N-glycosylation site; arrows at the bottom point to the 10 conserved cysteine residues. Shaded residues are sites at which amino acid substitutions abolished ASIP function in mice (Perry et al. 1996; Miltenberger et al. 2002), while asterisks mark sites where substitutions decreased ASIP function in mice (Perry et al. 1996; Miltenberger et al. 2002). Black diamonds mark sites where induced mutations decreased ASIP inhibition of MC1R, while open diamonds mark sites proposed to be important for selective inhibition of other melanocortin receptors (Kiefer et al. 1997, 1998). The horizontal line above the sequence demarcates a region whose deletion abolished ASIP function, whereas deletion of boxed residues did not affect its activity (Perry et al. 1996). Horizontal lines below the sequence mark small deletions reported to decrease ASIP function (Miltenberger et al. 2002).

Figure 2-7. Nucleotide variability in the *ASIP* gene measured using a sliding window approach. Shown is a plot of the number of variable sites in non-overlapping windows of 12 nucleotides each. Sites with alignment gaps were excluded from the assessment of window size or number of variable sites.

Figure 2-8. Graph comparing synonymous (dS) and nonsynonymous (dN) substitutions (per synonymous or nonsynonymous site, respectively) estimated for the

ASIP and *AHCY* genes in six different taxon comparisons, using the method of Pamilo & Bianchi (1993) and Li (1993). Domestic cat sequences for both genes were generated in this study; accession numbers for other *ASIP* sequences are listed in Figure 2-5; accession numbers for other *AHCY* sequences are as follows: mouse: NM_016661; rat: NM_017201; human: NM_000687. Codons that were not present in all four taxa (insertions/deletions or missing data) were excluded prior to the analysis, leaving a total of 130 codons for *ASIP* and 381 codons for *AHCY*.

Figure 2-9. Unrooted phylogenetic tree of twelve felid species constructed with 1,750 bp of nucleotide sequences from *ASIP* intron 2 and exons 2 and 3. The tree shown was constructed with a maximum likelihood (ML) heuristic search using a GTR+G+I model of nucleotide evolution and parameters estimated from the data set. Identical topologies were obtained with maximum parsimony (MP) and minimum evolution (ME) analyses. Numbers above branches are bootstrap support values for adjacent nodes (ML / ME / MP). The triangle indicates the branch in which the deletion shown in Figure 2-10 is inferred to have arisen. The main felid lineages represented in this data set are indicated by brackets and numbered as follows: 1. Caracal lineage; 2. Domestic cat lineage; 3. Puma lineage; 4. Ocelot lineage; 5. Panthera lineage; 6. Bay cat lineage. See text for species common names.

Figure 2-10. Alignment of a segment of *ASIP* intron 2 (positions 110-157 after end of exon 2) among 22 cat species, showing a 14-bp phylogenetically informative deletion, and the node it defines in the felid phylogeny (marked by a triangle). The number of

individuals sequenced for each species is given in parentheses. The tree on the left shows the phylogenetic relationships among the included species, based on Johnson & O'Brien (1997), Pecon-Slattery & O'Brien (1998), Mattern & McLennan (2000) and results from this study (Figure 2-9). Circles indicate branches representing the eight major lineages of the Felidae; dashed lines (for *O. manul*, *L. serval* and *P. marmorata*) indicate species that have not been conclusively placed in these lineages. An outgroup species (*Artictis binturong*, Carnivora: Viverridae) is also included to demonstrate the derived state represented by the deletion.

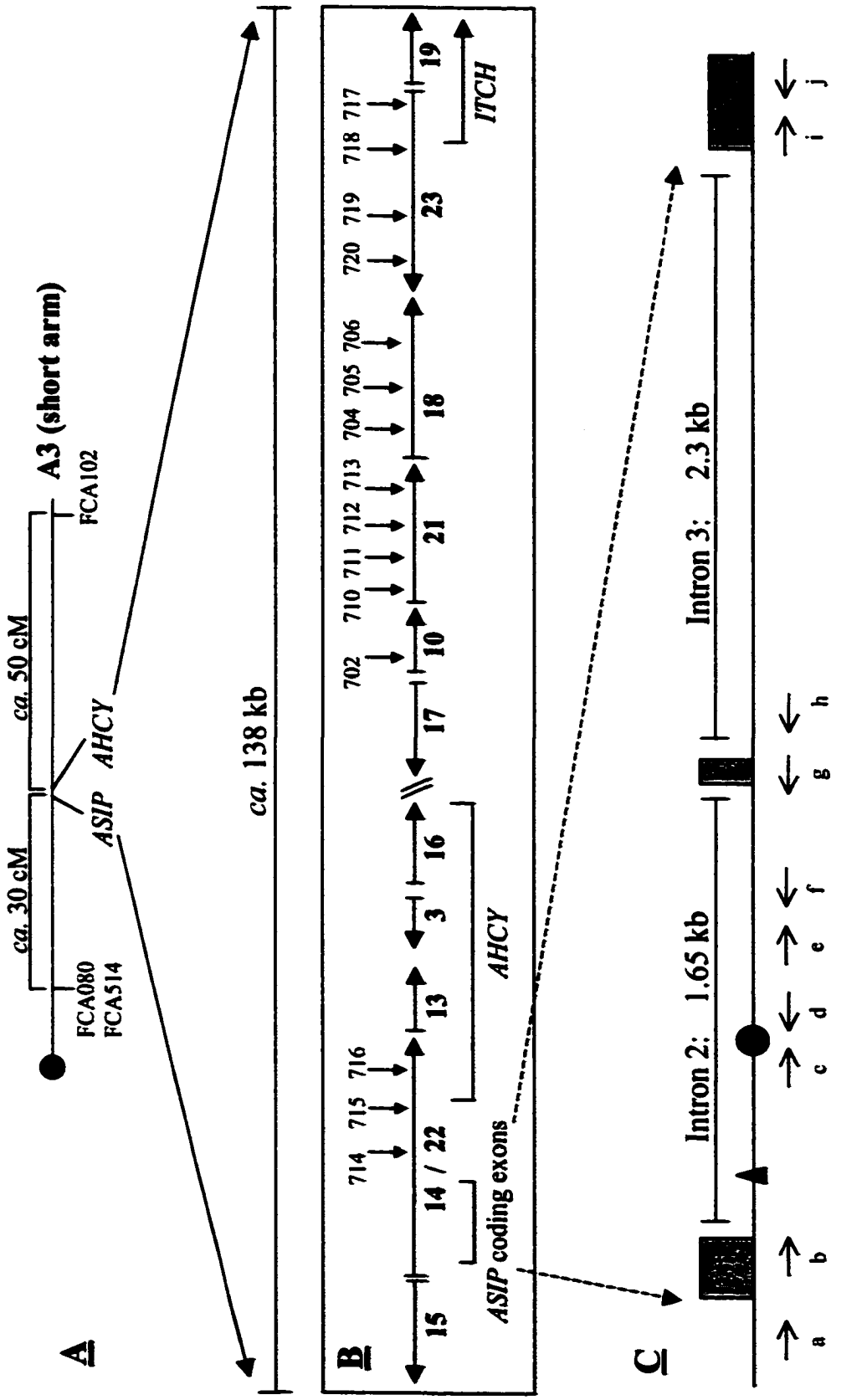


Figure 2-1

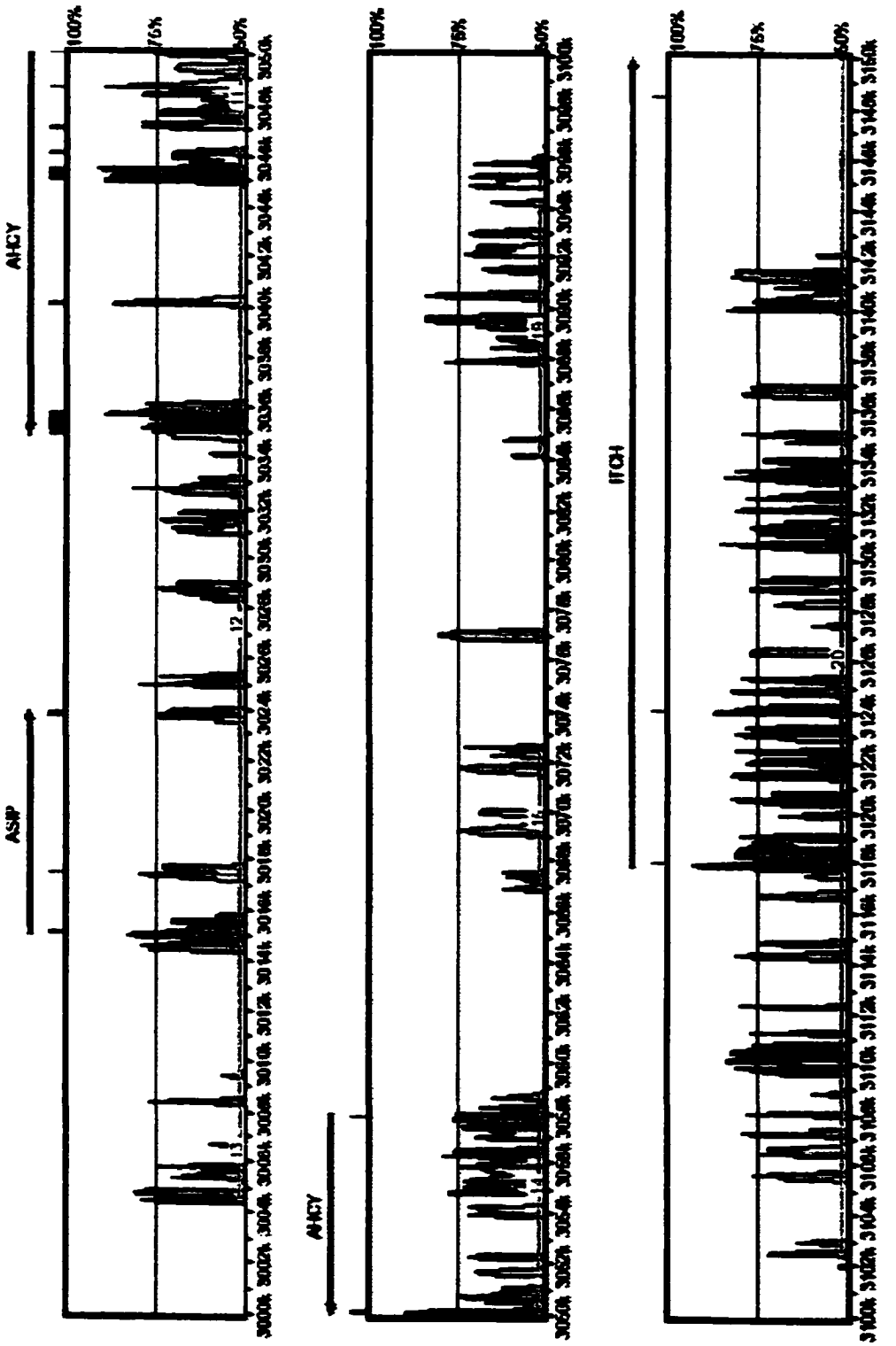


Figure 2-2

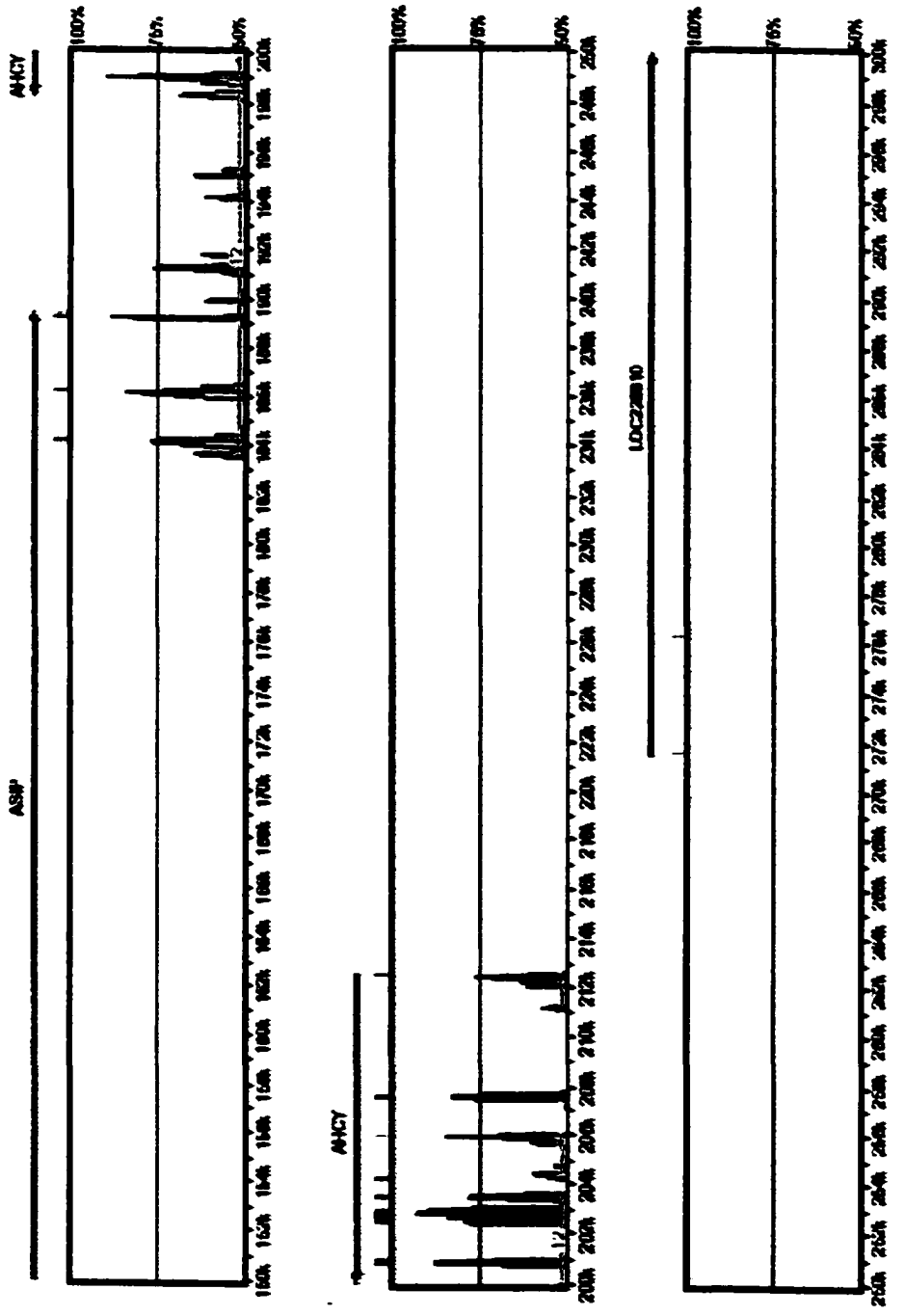


Figure 2-3

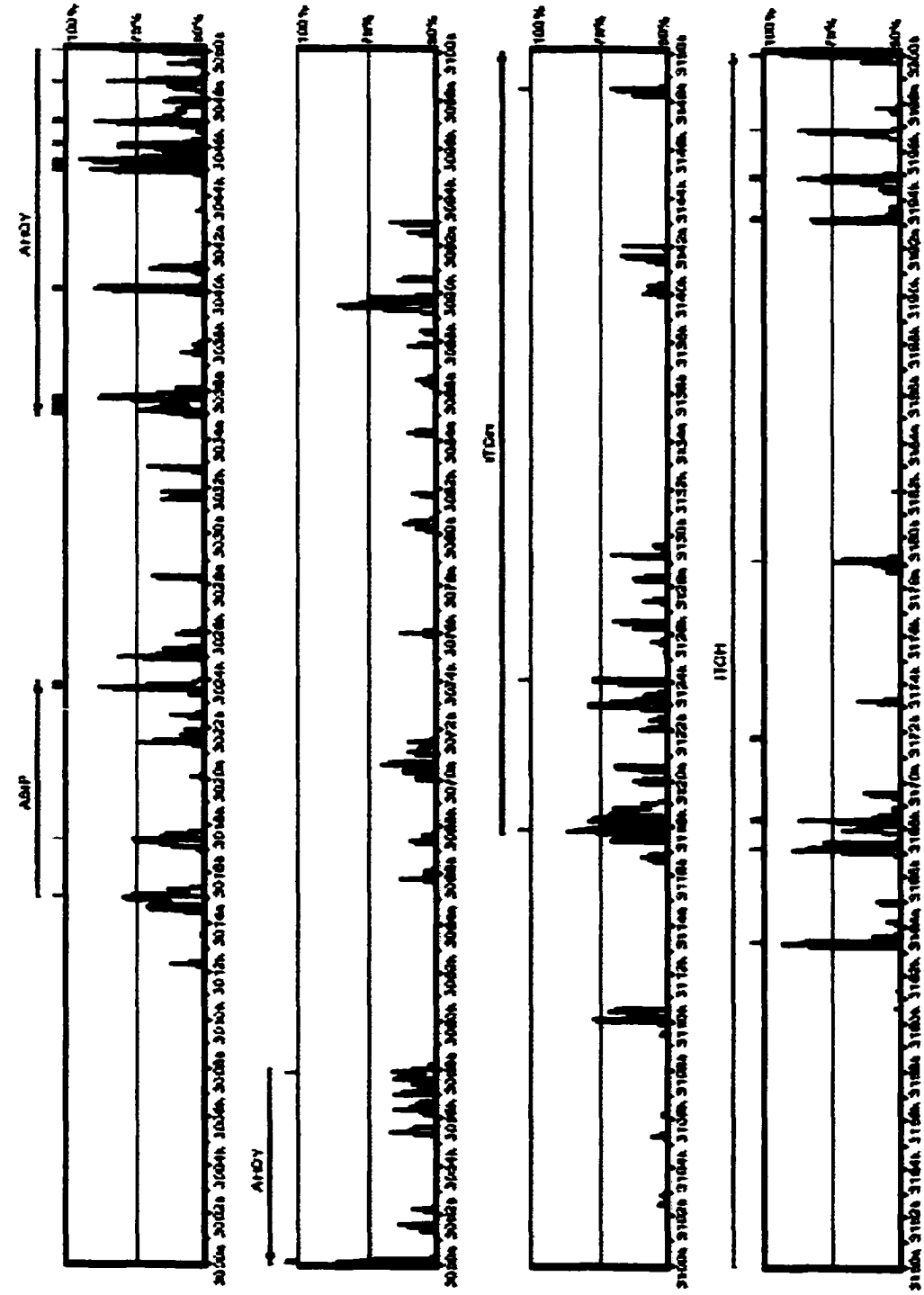


Figure 2-4

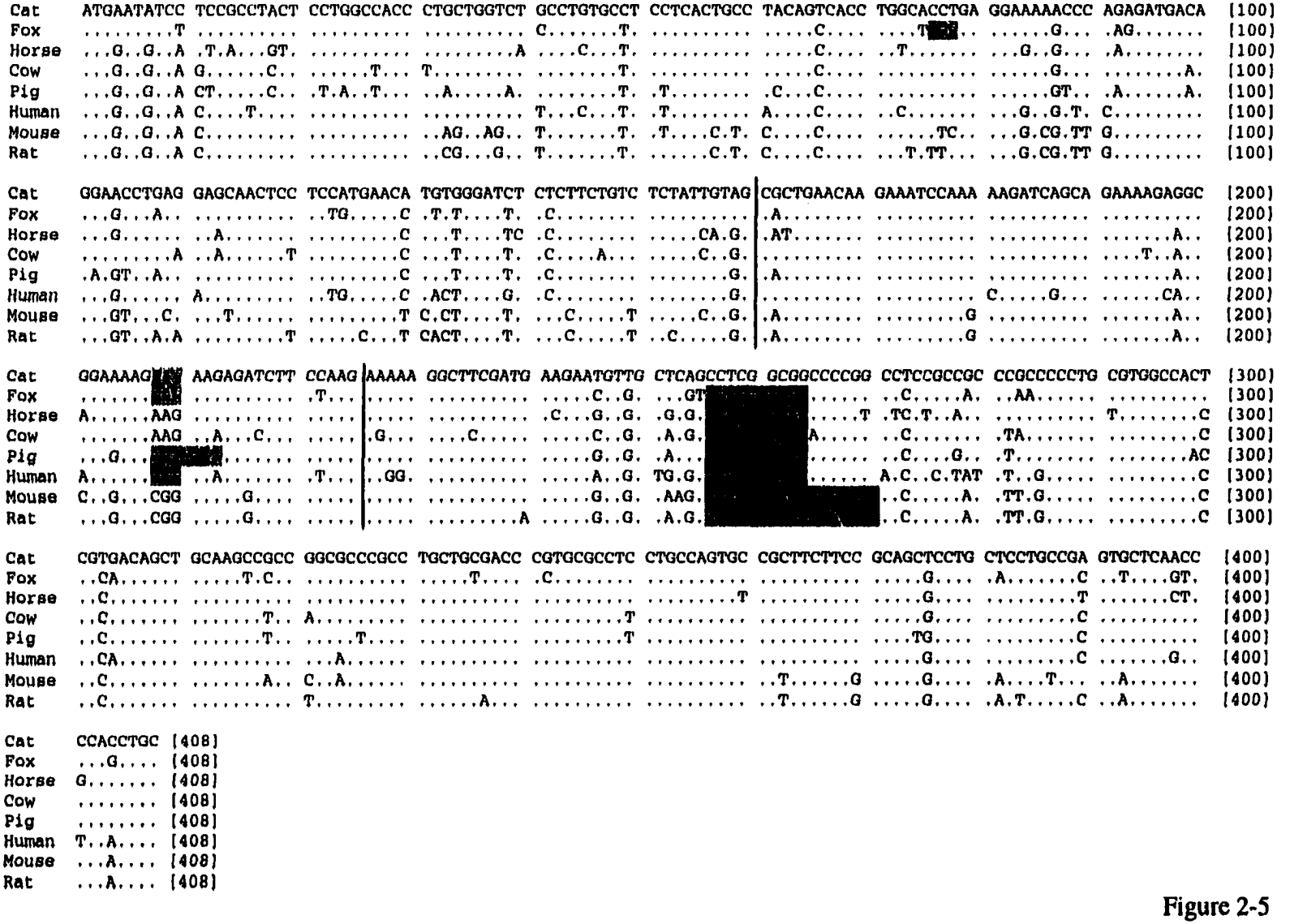


Figure 2-5

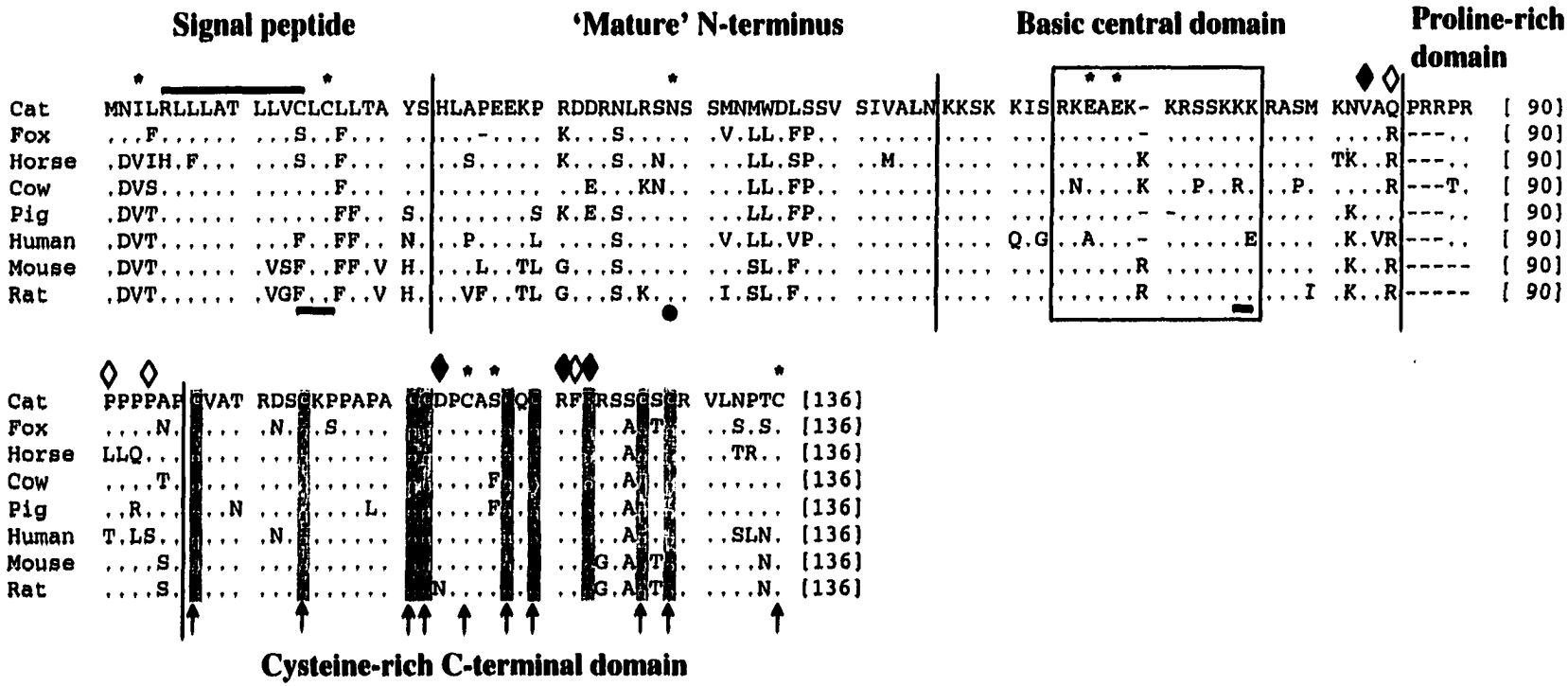


Figure 2-6

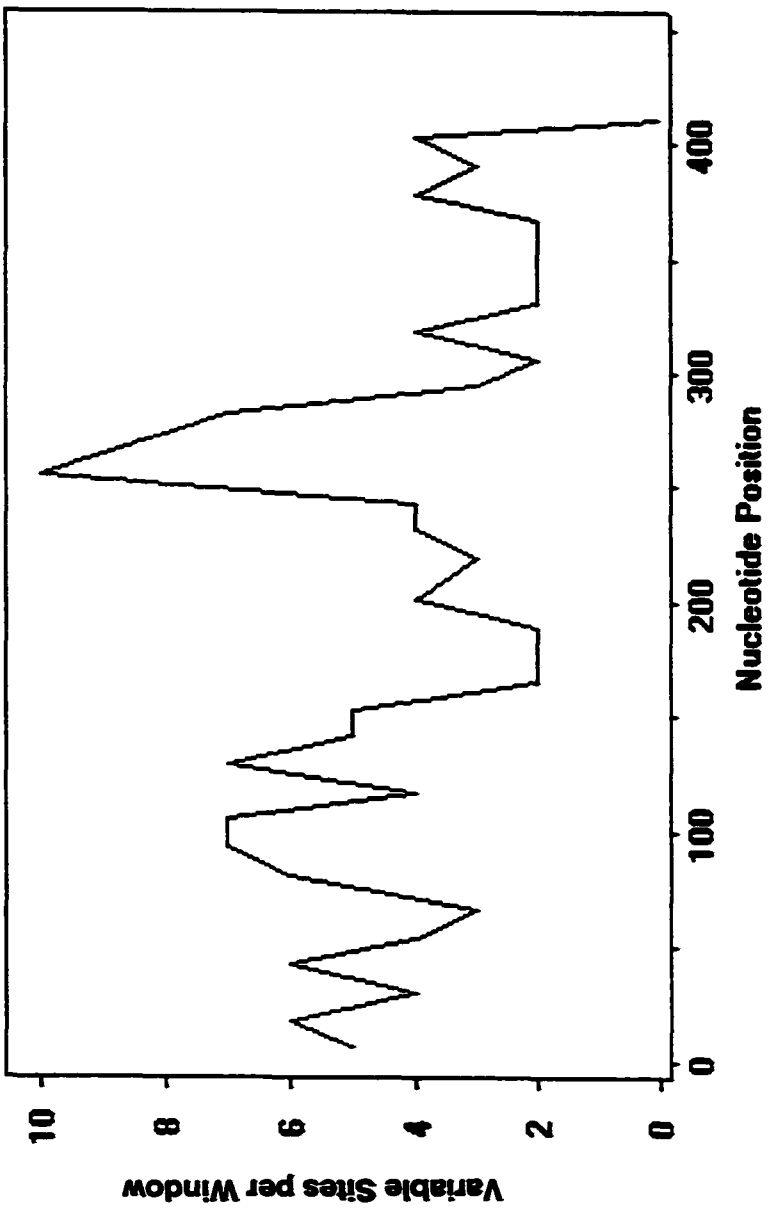


Figure 2-7

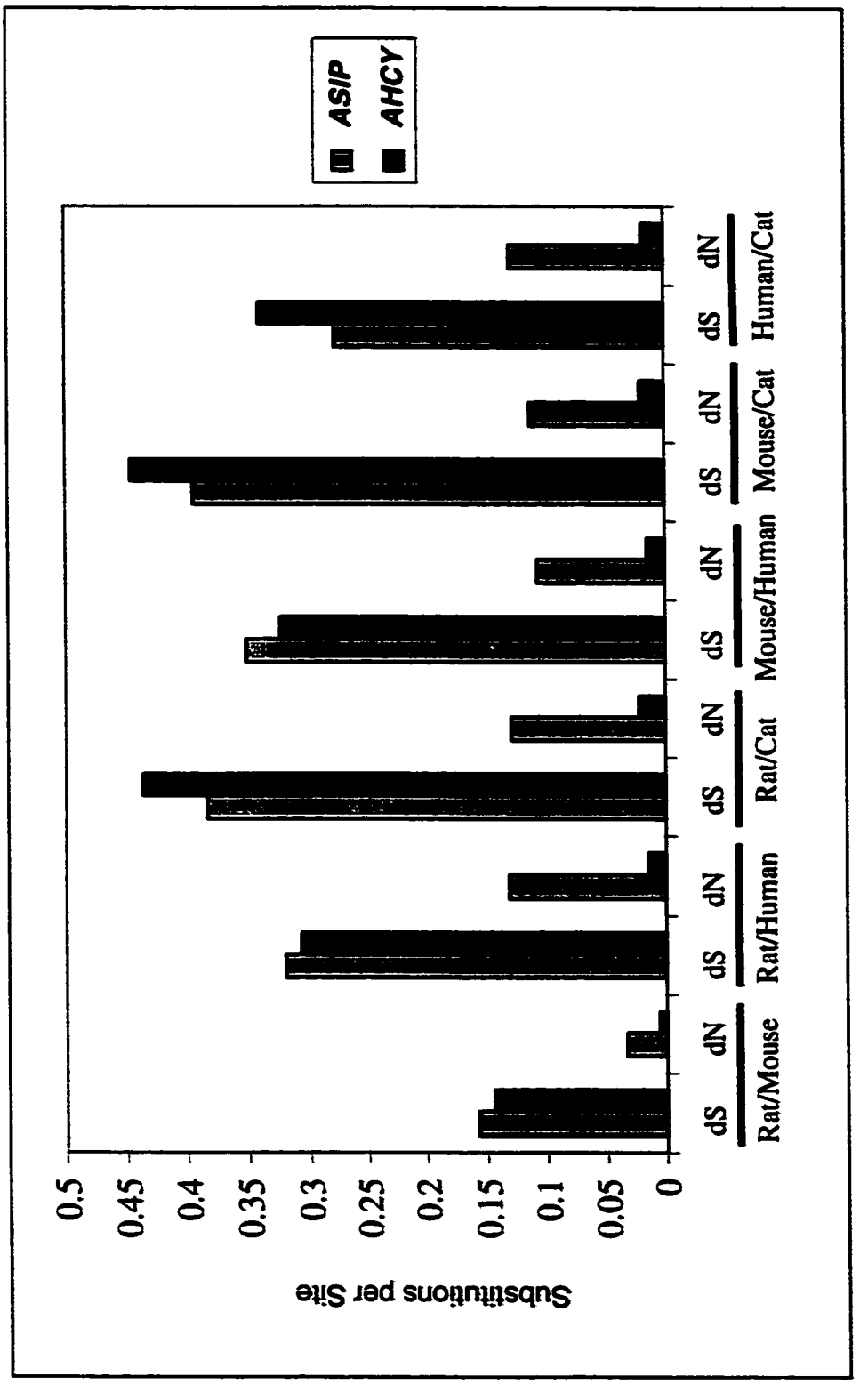


Figure 2-8

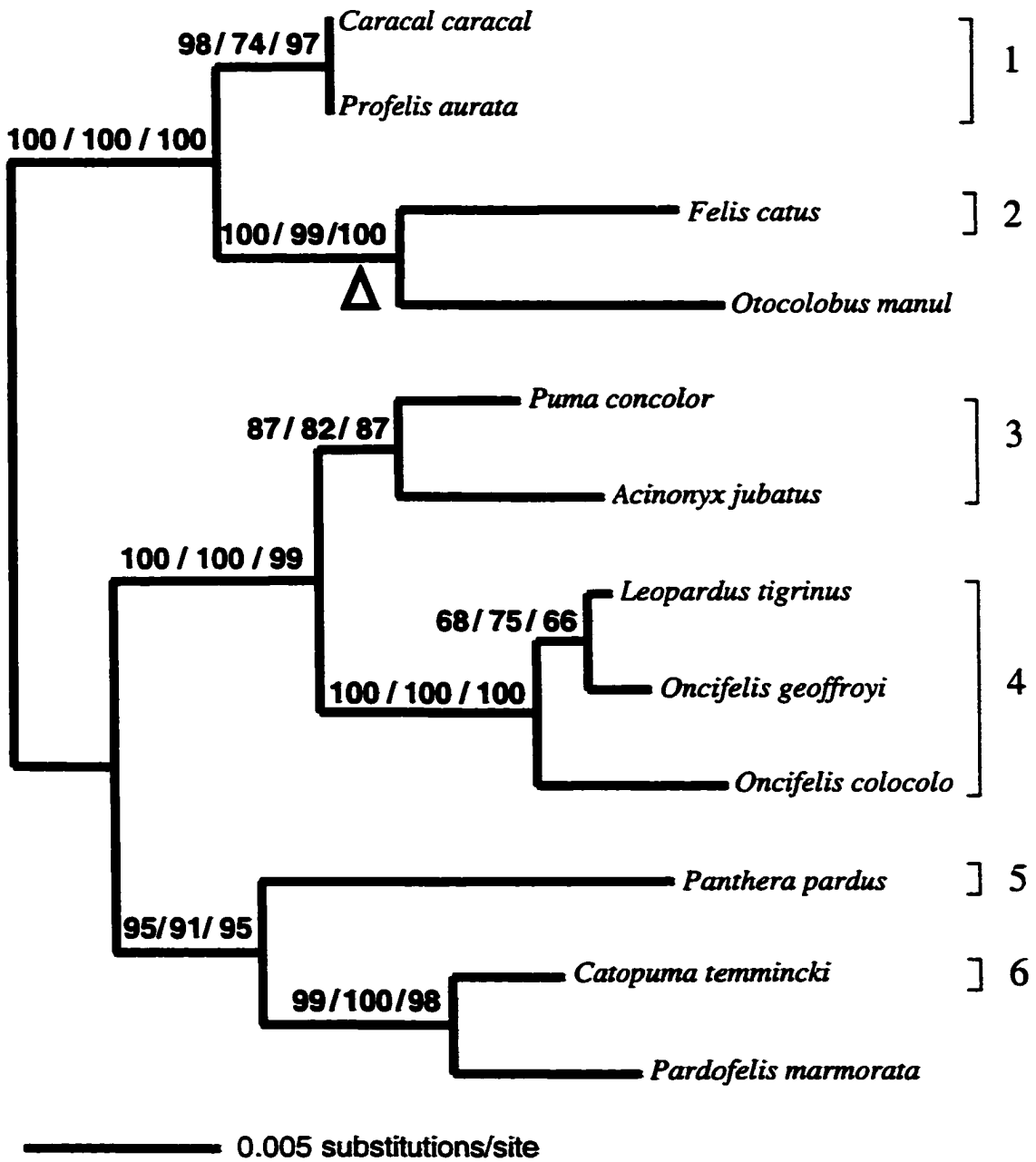


Figure 2-9

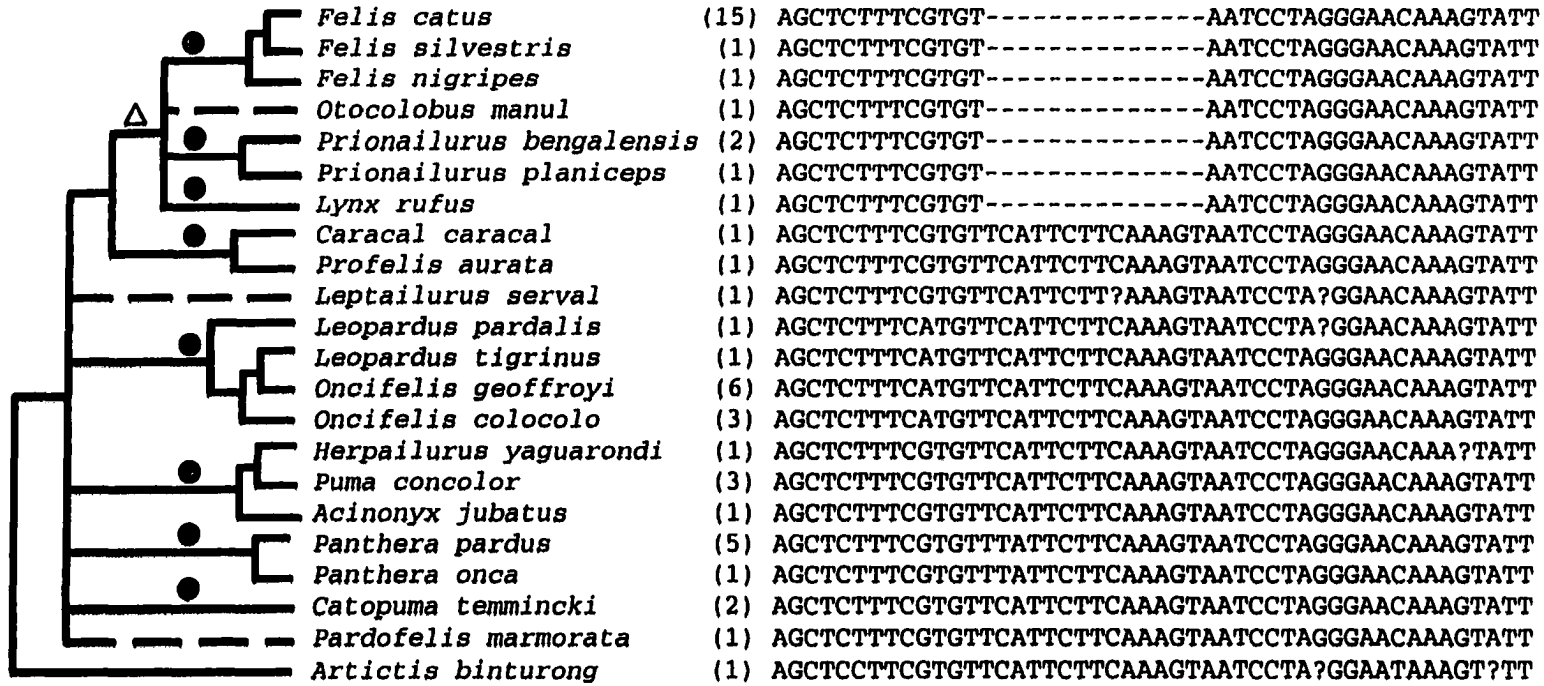


Figure 2-10

CHAPTER 3

Evolutionary Genomic Analysis of the Feline *MC1R* Gene

Abstract

Variants at the melanocortin-1 receptor (*MC1R*) gene have been implicated in pigmentation phenotypes in several mammalian and avian species. Although *MC1R* has been sequenced in several vertebrate species, so far no study has systematically investigated its patterns of variation among different evolutionary lineages, and among spatial components of the resulting protein and extended promoter region. This study describes the mapping of the *MC1R* gene in the domestic cat, and the cloning and sequencing of an 8.7 kb genomic segment containing *MC1R* in this species. Phylogenetic footprinting analyses including seven mammalian species (representing five eutherian orders) identified 25 evolutionarily conserved non-coding segments in the vicinity of this locus, only one of which had been described previously. A survey for transcription factor binding sites in these conserved segments is presented, along with a comparative analysis of the main *MC1R* promoter region in eight species. An evolutionary analysis of the *MC1R* coding region was also performed, examining patterns of nucleotide and amino acid variation among domains and different vertebrate lineages. In the concluding section of this chapter I survey the extent of sequence variation in the *MC1R* coding region of the domestic cat and 12 other felid species, and evaluate the utility of this gene for evolutionary studies of the Felidae.

Introduction

Of several genes found to influence coat color variation in mice (Searle 1968, Silvers 1979, Jackson 1994), the *melanocortin-1 receptor* (*MC1R*) has so far received the most attention with respect to molecular studies in other species (e.g. Klungland et al. 1995; Kijas et al. 1998, 2001; Våge et al. 1997, 1999; Rana et al. 1999; Harding et al. 2000, Makova et al. 2001; Rees 2000; Ritland et al. 2001; Smith et al. 2001; Theron et al. 2001). Interest has been drawn to this gene by the identification of variants involved in coloration phenotypes in several species (e.g. Robbins et al. 1993, Klungland et al. 1995; Valverde et al. 1995), and by the finding of high levels of intra-specific variability at this locus in humans (Rana et al. 1999, Harding et al. 2000, Makova et al. 2001, Smith et al. 2001). The investigation of this gene in diverse species has also been facilitated by the fact that it contains a single coding exon spanning *ca.* 1 kb, reducing the difficulty of obtaining complete sequences of its coding region from genomic DNA.

In mice, *MC1R* was found to correspond to the classical coat color locus *extension* (Robbins et al. 1993), which is associated with dominant melanism and recessive yellow (Silvers 1979; Jackson 1994; Barsh 1996). *Extension* variants have also been hypothesized to be involved in similar phenotypes in other mammals (Searle 1968; Robinson 1970a). The *MC1R* gene product (a seven-transmembrane G-protein coupled receptor) plays a critical role in the biochemical switch between the synthesis of eumelanin (dark pigment) *versus* pheomelanin (light pigment) in hair-follicle melanocytes (Robbins et al. 1993; Jackson 1994; Barsh 1996; Cone et al. 1996; He et al. 2001; see Chapter 1 for more details). *MC1R* is also expressed in immune system

cells such as T cells, macrophages and mast cells, where it appears to be involved in signaling pathways related to inflammatory responses (Adachi et al. 2000). *MC1R* mutations have been found to be implicated in variant pigmentation phenotypes in several species, including the mouse (Robbins et al. 1993), human (Valverde et al. 1995, Box et al. 1997), cow (Klungland et al. 1995), horse (Marklund et al. 1996), red fox (Våge et al. 1997), pig (Kijas et al. 1998, 2001), sheep (Våge et al. 1999), dog (Newton et al. 2000, Everts et al. 2000), black bear (Ritland et al. 2001), chicken (Takeuchi et al. 1996) and bananaquit birds (*Coereba flaveola*) (Theron et al. 2001). The phenotype-altering *MC1R* mutations observed in these species included both gain-of-function mutations associated with dominant melanism (e.g. in mice, pigs, cattle and sheep) and loss-of-function mutations associated with recessive light color (yellow, whitish or reddish) (e.g. in mice, pigs, dogs, black bears and humans). *MC1R* is so far the only locus found to be clearly associated with normal variation in hair and skin color in humans (Box et al. 1997; Smith et al. 2001; Sturm et al. 2001). It influences sensitivity to burning by ultra-violet radiation and is a genetic risk factor for skin cancer (Rees 2000). Additionally, its patterns of molecular diversity have been used to make inferences on the population-level evolutionary history of our species (Rana et al. 1999; Harding et al. 2000; Makova et al. 2001).

Even though *MC1R* diversity has been used to investigate human demographic history, and conservation of particular coding sites among species has been used to infer functional constraints (e.g. Kijas et al. 1998, Rana et al. 1999; Everts et al. 2000), the patterns of *MC1R* variation at different levels of divergence have not been studied in detail, and some important aspects of its molecular evolution remain unclear. While

considerable variability of *MC1R* sequences can be observed among mammals on the basis of previously published alignments (e.g. Kijas et al. 1998; Newton et al. 2000), and even within human populations (e.g. Rana et al. 1999, Harding et al. 2000), some studies have referred to this gene as being "highly conserved" (e.g. Klungland et al. 1999; Våge et al. 1999). Also, while some studies of human *MC1R* variability indicate that high diversity in non-African populations is mostly due to relaxation of selective constraints in these geographic regions (Harding et al. 2000, Rees 2000), others have suggested that diversifying selection has been operating in these areas (Rana et al. 1999; Sturm et al. 2001). A comprehensive analysis of *MC1R* sequence variation across multiple species at different levels of divergence can potentially shed light on spatial and temporal patterns of functional constraint enforced on the resulting protein, and thus aid in the interpretation of observed variants in specific taxa.

All functionally significant mutations identified so far in the *MC1R* locus have been in the coding region, but it is possible that transcriptional regulation of this gene may also affect pigmentation phenotypes (Moro et al. 1999; Makova et al. 2001; Smith et al. 2001). Recent studies have characterized the human and mouse *MC1R* promoter region, and identified one segment (*ca.* 500 bp upstream of the initiation codon) that is critical for activation of *MC1R* transcription, and for regulation of this gene by the microphtalmia transcription factor (MITF) in cultured mouse mast cells (Moro et al. 1999; Adachi et al. 2000; Smith et al. 2001). Two studies used phylogenetic footprinting (a search for non-coding sequence blocks that are conserved among different evolutionary lineages; Tagle et al. [1988]; Gumucio et al. [1996]) to demonstrate sequence conservation of this critical segment between human and mouse

(Makova et al. 2001; Smith et al. 2001). Results from these experimental and comparative analyses indicate that further regulatory elements (e.g. a tissue-specific silencer, and additional targets for MITF transactivation) should exist in addition to those identified in this immediate critical region, and suggested potential differences between human and mouse *MC1R* promoters (Moro et al 1999; Adachi et al. 2000; Makova et al. 2001; Smith et al. 2001). Extended genomic comparisons among multiple species may reveal additional conserved motifs potentially relevant for *MC1R* transcriptional regulation, and provide more details on the patterns of sequence conservation in this region.

The *MC1R* gene has not been identified or characterized in the domestic cat. Moreover, genetic variants fitting the *extension* series of coat color alleles have not been observed and may not exist in this species, since no mutants bearing dominant melanism or autosomal recessive yellow have been confirmed (Searle 1968, Robinson 1976). Yet domestic cats have a unique form of X-linked *Orange* coloration (Wright 1918, Robinson 1976), prompting speculation that the *MC1R* gene may have been translocated onto the X chromosome in this species, so that a loss-of-function "recessive yellow" mutant at this locus could be the basis of the X-linked *Orange* phenotype. Testing of this hypothesis would be important in the context of understanding the molecular basis of the X-linked *Orange* mutation (a classical coat color variant used to support the proposition of X-chromosome inactivation in mammals [Lyon 1961]).

This study describes the *MC1R* gene of the domestic cat, including its chromosomal location and sequence features of the coding region and adjacent

genomic segments. The domestic cat *MC1R* genomic sequence and those of six other mammals were used to perform a detailed characterization of evolutionarily conserved non-coding segments in the vicinity of this gene (both 5' and 3' of the coding region) that are likely involved in the regulation of *MC1R* or adjacent loci. An evolutionary analysis of patterns of *MC1R* coding region variation is also presented, based on comparisons with sequences from other mammals, birds and a fish. Lastly, a survey of sequence variability in the *MC1R* gene among different cat species is included, and the utility of this locus for evolutionary studies in the family Felidae is discussed.

Materials and Methods

Experimental procedures

Conserved PCR primers targeting the *MC1R* coding region were designed on the basis of available sequences for humans, mouse, horse and red fox (Mountjoy et al. 1992, Robbins et al. 1993, Marklund et al. 1996, Våge et al. 1997). Nine primers were designed and tested in all possible combinations, with varying annealing temperatures and $MgCl_2$ concentration. The primers extF4 and extR4 (see Table 3-1 for list of all primers used in this study) amplified a suitable product (770 bp long) under a touchdown PCR profile (decreasing annealing temperatures from 60°C to 50°C) with *AmpliTaq* Gold (Perkin Elmer) and 2.0 mM $MgCl_2$. The PCR products for three domestic cats were purified with Centricon-100 concentrators (Amicon) and sequenced using ABI Big Dye terminator chemistry; sequencing products were purified with CentriSep columns (Princeton Separations) and analyzed with an ABI

373 automated sequencer. Resulting sequences were verified with BLAST genomic searches (Altschul et al. 1990) and used for further primer design.

To map the genomic location of the domestic cat *MC1R* gene, a PCR-based radiation hybrid (RH) assay was developed: primers extRH-F1 and extRH-R1 were designed to amplify a 230-bp product within the coding region, and optimized to produce an intense, clean amplification in cat genomic DNA, and no product in the hamster background cell line. The PCR reactions (35 cycles with 59.5°C annealing temperature, *AmpliTaq Gold* DNA polymerase (PE) and 1.5 mM $MgCl_2$) were performed in duplicate in 96-well format using the domestic cat radiation hybrid panel (Murphy et al. 2000). Analysis was performed with RHMAP (Boehnke et al. 1991) relative to markers and genes previously mapped by Murphy et al. (2000).

To obtain the full sequence of the domestic cat *MC1R* gene and surrounding genomic regions, a BAC clone containing this locus was identified and characterized. A domestic cat BAC library (RPCI 86; Beck et al. 2001) was screened using a probe derived from a 230 bp domestic cat PCR product (amplified with primers extRH-F1 and extRH-R1) labeled with ^{32}P using a random priming kit (Boehringer-Manheim). The hybridization was performed using multiple probes for different genomic regions of interest (not shown). Positive clones for the *MC1R* gene were identified using direct PCR from bacterial colonies, and confirmed by sequencing. Four BAC clones positive for *MC1R* were identified (108d2, 130a19, 225b16, 248l10), and two were selected for characterization. BAC clones 225b16 and 248l10 were grown in large-scale (500 ml) culture, purified with a Qiagen Large-Construct Kit, and submitted to restriction enzyme digestion (using *Bam*H1 and *Hind*III, separately and combined).

Restriction fragments were visualized with ethidium bromide in a 1% agarose gel, and transferred to a Duralon membrane (Stratagene) overnight using standard conditions (Sambrook et al. 1989). Fragments were cross-linked to the membrane in a UV oven. The same PCR-derived domestic cat *MC1R* probe was used in an overnight hybridization to this membrane at 65°C, followed by washes with [2X SSC, 0.5% SDS] and [0.1X SSC, 0.5% SDS], and exposure with X-ray film for 20 minutes.

A single positive band was identified in each of the three enzyme treatments, and a *Bam*H1 fragment of *ca.* 9 kb in clone 248110 was selected for further characterization. This fragment was cut from a 0.8% agarose gel, purified by electro-elution in a dialysis tube (Gibco BRL), and confirmed with a second Southern blot using the same probe and procedure described above. The fragment was then ligated to a *Bam*H1-digested pBS KS$^-$ plasmid (treated with calf intestinal alkaline phosphatase [CIAP]), electroporated into *E.coli* DH10B competent cells (Gibco BRL), and plated for ampicillin selection. Positive colonies were streaked onto a new plate, grown overnight, and tested for the presence of the *MC1R* insert by PCR using primers extRH-F1 and extRH-R1. Only one clone produced a positive PCR result, and was then grown in liquid culture and additionally confirmed to contain the 9-kb fragment by *Bam*H1 digestion and agarose gel analysis. This recombinant plasmid was then fully sequenced using the AT-2 primer-island transposon kit (Perkin Elmer) to generate randomly-inserted priming sites. Forty-eight transposon-containing clones were isolated and sequenced in both directions in 96-well format with BigDye chemistry, followed by purification using Sephadex G-50 plates (Amersham Pharmacia) and analysis in an ABI 3700 automated sequencer. Sequences were

analyzed using a BioLIMS environment (PE informatics), including quality assessment, elimination of vector and *E. coli* sequences, and assembly into a single contig using PHRED, PHRAP and CONSED (Ewing et al 1998; Gordon et al. 1998).

From this genomic sequence a new primer set (extCOM-F1 and extCOM-R1) was designed to span the entire feline *MC1R* coding region (1079 bp product). This segment was amplified by PCR in three additional domestic cats, other felid species and a feloid carnivore outgroup (see Results) using a touchdown profile (total of 40 cycles, annealing temperature decreasing from 60°C to 50°C, with *AmpliTaq* Gold [PE] and 1.5 mM $MgCl_2$). Products were purified with Centricon-100 and directly sequenced using the PCR primers as well as internal primers extF4, extRH-F1, extR2 and extR4. Sequences were manually inspected and variants were identified using SEQUENCHER (Gene Codes).

Phylogenetic footprinting analyses

To identify and characterize conserved sequence blocks in non-coding regions adjacent to the cat *MC1R* gene, comparative analyses (phylogenetic footprinting: Tagle et al. [1988]; Gumucio et al. [1996]; see Chapter 2 for description of rationale behind comparative genomic analyses to identify conserved sequence regions) of the 5' and 3' regions were conducted relative to six available homologous mammalian sequences: dog (GenBank accession number AF064455), horse (AF288357), pig (AF326520), cow (AF445641), human (AF263461) and mouse (obtained from the CELERA Mouse Genome Data Base). Comparative analyses were performed using both local and global alignment applications. Initial comparisons were conducted with

BLAST, using both standard nucleotide searches against the non-redundant databases and pairwise gapped searches against specific sequences. Sequence comparisons were also conducted using the VISTA global alignment programs (Dubchak et al. 2000; Mayor et al. 2000; http://www-gsd.lbl.gov/vista/).

Identification of non-coding conserved sequence blocks using multiple-species VISTA (multiVISTA) was performed through extensive cross-species comparisons (focusing mostly on results from pairs including the domestic cat), varying parameters such as window size (10 or 20 bp) and threshold for minimum acceptable conservation (75% to 90% nucleotide identity per window). Final analyses used a 20 bp window size and minimum identity level of 88%. Conserved blocks that were not identified using these parameters for a given taxon comparison were further investigated using a 10 bp window (90% identity minimum threshold), and also with pairwise gapped BLAST searches. Refined comparative analyses of conserved non-coding segments 5' of the *MC1R* gene were performed for cat, human and mouse sequences using the Bayesian alignment approach implemented in the Bayes Block Aligner (Zhu et al. 1998). This method allows an evaluation of the probability of alignment between two genomic regions without the need to specify the penalties for gap creation and extension, two parameters that are often determined arbitrarily, and that may introduce relevant biases in the identification of conserved segments (Zhu et al. 1998). Putative binding sites for transcription factors in these conserved non-coding segments were identified using the regulatory VISTA (rVISTA) application (Loots et al. 2002), as well as direct visual comparisons with previously published experimental studies in humans and mice. In the rVISTA analysis, a database of

known transcription factor binding sites was screened for similarity with motifs present in the evaluated genomic region, and the resulting list of potential sites was filtered to exhibit only those contained in the segments found to be conserved among mammalian taxa.

Coding region analyses

Nucleotide and amino acid sequences of the *MC1R* coding region were aligned using ClustalX (Thompson et al. 1997) and visually checked. Sequence analyses were performed using MEGA 2.1 (Kumar et al. 2001) and PAUP 4.0 (Swofford 1998), and included (i) estimates of conservation and variability in different portions of the *MC1R* gene; (ii) characterization of rates and patterns of nucleotide substitution in different mammalian lineages, (iii) assessment of nucleotide composition and codon usage bias, (iv) phylogenetic inference; and (v) assessment of deviations from neutrality and rate-constancy in the evolutionary history of this gene in mammals. These analyses used a data set including the domestic cat *MC1R* sequence reported here, as well 15 additional mammals (representing five eutherian orders) whose sequences were available: domestic dog (*Canis familiaris*, GenBank accession AF064455), red fox (*Vulpes vulpes*, X90844), pig (*Sus scrofa*, AF326520), cow (*Bos taurus*, U39469), goat (*Capra hircus*, CHMC1R), sheep (*Ovis aries*, OAMC1R), muskox (*Ovibos moschatus*, OMMC1R), reindeer (*Rangifer tarandus*, RTMC1R), fallow deer (*Cervus dama*, CDAMC1R), red deer (*C. elaphus*, CEMC1R), moose (*Alces alces*, AAMC1R), horse (*Equus caballus*, AF288357), human (*Homo sapiens*, AF326275), chimpanzee (*Pan troglodytes*, PTR245705) and mouse (*Mus musculus*, X65635). The

data set for amino acid-level analyses also included three bird species (chicken [*Gallus gallus*, CHKM1R], bananaquit [*Coereba flaveola*, AF362575], and tanager [*Tangara cucullata*, AF362606]) and one fish (*Takifugu rubripes* [accession dbj|AB073674.1]). Nucleotide sequences from the three birds were used as outgroups for the mammalian data set in some DNA-based analyses.

Rates of synonymous and nonsynonymous substitutions in different *MC1R* domains were estimated for mammals, using the method of Li (1993) and Pamilo & Bianchi (1993), and also the modification of this approach proposed by Kumar (Nei & Kumar 2000), as implemented in MEGA. The existence of a constant rate of nucleotide substitution in the *MC1R* gene among different mammalian lineages was evaluated using a likelihood-ratio-based molecular clock test, as implemented in PAUP and described in Eizirik et al. (in press).

Results

The domestic cat *MC1R* gene

Initial PCR-based experiments produced amplification products for the domestic cat *MC1R* gene, confirmed by DNA sequence analysis. Radiation hybrid mapping of the domestic cat *MC1R* sequences showed it to be located on chromosome E2 (23.6 cR distal of marker FCA586; see Murphy et al. [2000] for reference markers), at a position homologous to its location on human chromosome 16q24.3. This autosomal location excludes the possibility that a loss-of-function allele at *MC1R* locus might be implicated in the domestic cat X-linked *Orange* mutant.

Characterization of a *Bam*H1 sub-fragment from BAC clone 248110 resulted in a 8,663 bp-long, finished sequence of the domestic cat *MC1R* gene and adjacent genomic segments, including 4.35 kb upstream of the coding region and 3.3 kb downstream of the termination codon (Figures 3-1 and 3-2). The domestic cat *MC1R* gene consists of an intron-less open reading frame comprising 951 bp (317 amino acids), similar in structure to previously described mammalian homologues (e.g. Mountjoy et al.1992; Våge et al. 1997), but exhibiting many nucleotide differences relative to other species (Figures 3-2 and 3-3; see comparative analyses below). Assessment of nucleotide composition in the domestic cat sequence revealed a high proportion of GC in this genomic region: 61% in the 5' non-coding segment, 62% in the 3' non-coding segment, and 65% in the coding region. When the three different codon positions were analyzed separately the GC content was 62% in the first, 42% in the second, and 91% in the third position.

Characterization of non-coding conserved sequence blocks

Genomic comparisons across the entire 8,663 bp domestic cat sub-clone revealed the existence of at least 25 non-coding segments that are highly conserved among mammals (Figures 3-1 and 3-2; Tables 3-2 and 3-3), both upstream and downstream of the *MC1R* coding region. These conserved sequence blocks (CSBs) were located up to 3.2 kb away from the *MC1R* gene, and displayed similar or sometimes higher levels of nucleotide conservation than portions of the coding region (Figure 3-2). Only segments that exhibited conserved motifs in the domestic cat and in at least two additional mammalian species are reported here.

Eleven CSBs were identified in the non-coding genomic region 5' of the cat *MC1R* gene, ranging in length from 8 bp to 56 bp, and showing 75% to 100 % nucleotide identity with other mammals (Figure 3-1, Table 3-2). Block 5h coincided with the experimentally-defined minimal *MC1R* promoter (Moro et al. 1999; Adachi et al. 2000), and was strongly conserved in all surveyed species (Figure 3-1, Table 3-2). It contained several putative Transcription Factor Binding Sites (TFBSs) including two SP1 motifs, one of which was contiguous to an E-box ('CANNTG'; Aksan & Goding 1998) in an identical disposition (CATGTGGCCGCCC) to the human segment functionally identified as critical for *MC1R* transcriptional activation (Moro et al. 1999). This E-Box was almost perfectly conserved across all surveyed mammals (Figure 3-4), whereas the adjacent E-box (CACATG) also identified to be functionally important in the mouse (Adachi et al. 2000) was not (see Figure 3-4 and Discussion).

To survey the occurrence of potential transcription factor binding sites (TFBSs) in the remaining CSBs 5' of *MC1R*, rVISTA analyses were performed based on cat/human and cat/cow genomic alignments. The resulting conserved sites were further filtered to include only genomic regions shared by other species as well. Clusters of TFBSs were identified in all of the 5' conserved blocks shown in Figure 3-1, except 5i. Identified clustered TFBSs included the following (blocks in which the TFBS was detected are given in parentheses, along with number of sites per block [in brackets] when more than one): **C/EBP** (5a [4], 5b, 5c, 5d, 5e, 5f, 5g, 5j, 5k); **AP-1** (5a, 5c, 5d, 5e); **AP-2** (5h); **AP-4** (5b); **Sp1** (5a, 5d [2], 5h [2], 5j [4]); **GKLF** (5g, 5h, 5j [3]); **ets1** (5a, 5c, 5e, 5g, 5h, 5j [2], 5k); **CREB** (5d); **YY1** (5a, 5b, 5e, 5g, 5h [3], 5k); **NFAT** (5a, 5g); **GATA-1** (5h); and **MZF1** (5e, 5g [2], 5h [2], 5j). A complete

list of TFBSs identified in each of the conserved sequence blocks 5' of *MC1R* is given in Appendix 2.

In the 3.3 kb genomic segment 3' of the *MC1R* coding region, 14 CSBs were identified, ranging in size from 8 bp to 174 bp, and displaying 78-100% nucleotide identity with other mammals (Table 3-3). The comparative location of conserved blocks relative to the human sequence was used to infer their position with respect to the *MC1R* 3' UTR, based on experimental evidence on the mapping of the poly-adenylation signal in this transcript (Smith et al. 2001). This comparison indicates that sequence blocks 3a, 3b and 3c would be contained in the 3' UTR, and the remaining 3' blocks seem to be located in the intergenic space downstream of *MC1R*. Block 3c corresponds to a portion of the second *MC1R* exon (195 bp long) reported by Tan et al. (1999) to be present in an alternatively spliced isoform (MC1R-B) identified in humans (see Discussion).

Comparative analysis of the *MC1R* coding region

The domestic cat *MC1R* gene was compared to available sequences from 15 other mammals, three birds and a fish (the latter used only for amino acid-level comparisons; see Materials and Methods for list of included taxa). Among mammals, the domestic cat *MC1R* coding region showed an overall proportion of nucleotide identity of 81% with respect to mouse, *ca.* 85% with human and cattle, *ca.* 87% with pig and horse, and 90% with dog and fox. Amino acid identity was 75% with mouse, 83% with human and cattle, *ca.* 85% with pig and horse, and 90% with dog and fox.

Nucleotide and amino acid variation across mammals was abundant throughout the *MC1R* gene (Figures 3-3 and 3-5). Overall, only a few short ($\leq$ 10 residues) continuous segments were completely conserved at amino acid level among all surveyed mammals (Figure 3-5). These included 106 residues that were conserved across vertebrates, which were mostly concentrated in the second intra-cellular domain (ID2) and the sixth transmembrane domain (TM6) (Table 3-4, Figure 3-5). Most of the variation was observed in the first and second extra-cellular domains (ED1 and ED2), and also in ID3 and portions of TM3, TM4 and TM7 (see Figures 3-3 and 3-5). ED1 was particularly variable, including multiple amino acid replacements and two insertion/deletion (indel) events among mammals: a two-amino acid deletion in the mouse and a three-residue insertion in the pig (the polarity of these indels was inferred based on the placental mammal phylogeny presented by Murphy et al. [2001]). Two additional deleted residues were observed in the ED1 of birds relative to the mammals and fish, and virtually no conservation at the amino acid level was observed among these three vertebrate groups for this portion of MC1R (Figure 3-5, Table 3-4).

When the fifteen *MC1R* domains were divided into three categories (extra-cellular [ED], intra-cellular [ID] and transmembrane [TM]), measures of variation among mammals, birds and fish were considerably higher in the ED than in the ID or TM, and particularly so in ED1 and ED3 (Table 3-4). Heterogeneity in divergence estimates was observed within the ID and TM categories, with the lowest values obtained for ID1, ID2 and TM6 for all comparisons. Divergence as estimated by the observed proportion of different residues (p-distance) seemed to be saturated in the bird-mammal comparison for ED1, ED3, ED4, ID2, ID3, TM2, TM4 and TM7, as

these values were similar to or higher than those observed in the much older comparison of these groups with fish.

In light of the observed saturation, nucleotide-level analyses of *MC1R* substitutions were limited to the mammalian data set, using the birds as outgroups when appropriate. This taxon set exhibited the same pattern of very high GC content at the third codon position observed in the domestic cat *MC1R*, with an average of 88% GC at these sites. The highest proportion of GC at the third codon position was observed in the birds (95%) and in the pig (97.5 %). The high GC content at third codon positions was correlated with the observation of strong bias in codon usage in this data set. Measures of Relative Synonymous Codon Usage (RSCU, Sharp et al. 1986) ranged from 0.01 for UUA (Leu) and 0.06 for CUA (Leu), the least used codons, to 3.11 for CCC (Pro) and 3.29 for CUG (Leu), the most used codons. All codons showing strong preferential usage (RSCU >2.0) contained G or C at the third position, including the ones mentioned above and also CUC (Leu), AUC (Ile), GCC (Ala) and CGG (Arg).

The number of synonymous (dS) and nonsynonymous (dN) substitutions among mammals, and the derived rates per 10^9 years (r_S and r_N, respectively), were estimated for different segments of the *MC1R* gene among various lineages (Figure 3-6, Table 3-5). While synonymous rates were quite homogeneous across the coding region, nonsynonymous rates varied considerably among domains, producing similar spatial results to those obtained in the mammal-bird-fish amino acid level comparisons (Table 3-4). Most nonsynonymous variability was observed in the extra-cellular domains, particularly ED1 and ED3, whereas the highest conservation was seen in ID1

(Figure 3-6). Considerable variation in rates and patterns of nucleotide substitution was observed among lineages. Results from likelihood-based molecular clock tests using either all coding sites or only the first and second codon positions (reflecting mostly nonsynonymous changes) showed significant departure from rate constancy ($P<0.01$), with the most salient feature being the acceleration of the mouse lineage (Figure 3-7). Removal of the mouse sequence in subsequent analyses revealed that rate heterogeneity among the remaining lineages was still significant ($P<0.05$), indicating that the assumption of a molecular clock does not hold for this *MC1R* data set even after exclusion of the murid representative.

MC1R Sequence Variation in the Cat Family

A survey of variation at the *MC1R* gene in the Felidae was performed by amplifying and sequencing its coding region in three additional domestic cats, and one individual from each of 12 other felid species: black-footed cat (*Felis nigripes*), leopard cat (*Prionailurus bengalensis*), jaguar (*Panthera onca*), jaguarundi (*Herpailurus yaguarondi*), puma (*Puma concolor*), cheetah (*Acinonyx jubatus*), marbled cat (*Pardofelis marmorata*), Asian golden cat (*Catopuma temmincki*), caracal (*Caracal caracal*), African golden cat (*Profelis aurata*), serval (*Leptailurus serval*) and Pallas cat (*Otocolobus manul*). These species represent six of the eight major evolutionary lineages in the cat family, along with three felids (marbled cat, serval, and Pallas cat) not consistently placed in the latest molecular studies of the cat phylogeny (see Chapters 1 and 5). A full *MC1R* coding sequence was also obtained for a feloid carnivore outgroup (the genet *Genetta genetta*: Carnivora, Viverridae),

which provides a closer outgroup for evolutionary studies than the available canids (Wozencraft 1989; Salles 1992; Nowak 1999).

High levels of variation were observed among the cat species included in this data set (Figure 3-8), including 51 variable nucleotides, 20 variable amino acids, and pairwise differences ranging from 1 to 17. In spite of the observed variability, phylogenetic analyses using this *MC1R* segment produced limited resolution of felid relationships. No portion of the tree was resolved with high bootstrap support, and some inconsistencies were observed among the phylogenies inferred with maximum likelihood (ML), maximum parsimony (MP) and minimum evolution (ME). The sister-taxon relationship between puma and jaguarundi identified in previous studies (e.g. Johnson & O'Brien 1997) was recovered with all methods, albeit with only weak bootstrap support (ML: 53%, MP: 62%, ME: 55%). The only other relationship consistently recovered with all methods was the placement of the serval in a clade with the two representatives of the Caracal Lineage (caracal and African golden cat), with low to moderate boostrap support (ML: 67%, MP: 53%, ME: 56%).

Two single-nucleotide-polymorphisms (SNPs) were identified in the domestic cat *MC1R* coding region: G/A at site 169, and T/C at site 886 (Figure 3-8). Both substitutions were non-synonymous, the former causing a Val/Met change at codon 57 (position 61 in Figure 3-5 [TM1]), and the latter causing a Phe/Leu change at codon 296 (300 in Figure 3-5; [TM7]). The Val57Met mutation was present in a single individual (Fca264), apparently homozygous for it. The Phe296Leu variant was homozygous in one individual (Fca 194) and heterozygous in another (Fca215) (Figure 3-8).

Discussion

Non-coding conserved sequence blocks: Evolutionary and functional inferences

Only one of the conserved sequence blocks characterized here (CSB 5h) had been identified previously based on human-mouse comparisons (Makova et al. 2001; Smith et al. 2001). This block contains the critical *MC1R* promoter (Moro et al. 1999; Adachi et al. 2000), and the observed conservation among multiple species (Figure 3-1, Table 3-2) supports a significant regulatory role for this segment across mammals. A detailed comparative analysis of CSB 5h in eight species (representing five mammalian orders) revealed interesting aspects of nucleotide conservation and turnover (Figure 3-4). Of the two E-Boxes experimentally identified as critical for *MC1R* transactivation by MITF in mouse mast cells (Adachi et al. 2000), one (CATGTG) was conserved across all surveyed mammals, suggesting a common regulatory role. However, experimental analysis of the human *MC1R* promoter has refuted the involvement of this site in MITF transactivation in our species (Smith et al. 2001). This result was interpreted as being due to the lack, in the human promoter, of a 'T' nucleotide on the 5' flank of this motif, which has been shown to be critical for MITF recognition of 'CATGTG' E-Boxes (Aksan & Goding 1998). Figure 3-4 shows that a 'T' is indeed conserved at that position in mammals (including the gorilla and chimpanzee), suggesting that loss of functionality at that motif occurred recently in the human lineage. The 'G' nucleotide located at this site in humans was not found to be polymorphic by Makova et al. (2001), suggesting that our species as a whole has an E-box with low MITF affinity at that location. On the other hand two different

conserved sites within block 5h (indicated in Figure 3-4) have been found to be polymorphic in humans (Makova et al. 2001), suggesting that they may underlie relevant variation in *MC1R* expression present in our species. In addition to the 'T' nucleotide 5' of the conserved E-Box, the horse promoter also contained an 'A' on the opposite flank of this motif ('T' in the complement strand, also implicated in high MITF affinity; Aksan & Goding 1998), which may imply increased activation by MITF in this species.

The second E-Box (CACATG; underlined in the mouse in Figure 3-4) was not conserved among mammals, suggesting divergent regulatory functions for this segment. Interestingly, the chimpanzee and gorilla also show a convergent E-Box at this same location. It is not known whether this represents an active regulatory element in these species, given that their E-Box is a CATGTG motif lacking a 'T' on either side. Even if this segment in gorillas and chimpanzees were not bound by MITF, these and other E-Boxes in this region may still be targets of other basic-Helix-Loop-Helix transcription factors, as suggested by Aksan & Goding (1998) and Smith et al. (2001).

Even though the conserved E-Box in CSB 5h was demonstrated not to mediate MITF activation of *MC1R* in humans (Smith et al. 2001), these authors reported that an active MITF regulatory element does exist in the human *MC1R* promoter region (within ca. 1,700 bp of the initiation codon); however, this motif has not been identified. Although several potential E-Box motifs are present in the genomic region flanking *MC1R*, none was located in any of the mammalian conserved blocks identified here, suggesting that sequence turnover may have generated a new

functional site for MITF activation in humans that is not shared by other species. One potential candidate motif (CACGTG) within this region was identified a few bases away from conserved block 5f, as part of a 10-bp segment that is 100% identical between cat and human (positions 3332-3341 in the cat sequence; *ca.* 1,240 bp upstream from *MC1R* in humans), albeit not conserved in the other surveyed mammals.

The identification of transcription factor binding sites (TFBSs) in most of the other CSBs 5' of *MC1R* suggests a potential role for these segments as enhancer elements. The activating role of Sp1 binding to the human *MC1R* minimal promoter has been experimentally demonstrated (Moro et al. 1999), and the results presented here suggest that this transcription factor may also bind to enhancers further upstream from the gene. Other factors identified here have been shown to interact with Sp1 in different cellular contexts to mediate transcriptional activation, including C/EBP, AP-1 and GKLF (Lopez-Rodriguez et al. 1997; Higaki et al. 2002). Interactions among AP-1, YY1 and NFAT have also been reported in other cell types (Sweetser et al. 1998). The identification of a potential cAMP response element (bound by CREB protein; Alberts et al. 1994) in block 5d adds to the hypothesis that *MC1R* expression may be upregulated by cellular levels of cAMP, which has also been inferred based on the role of MITF (itself induced by cAMP) in *MC1R* regulation (Adachi et al. 2000). Since MC1R signaling increases cellular levels of cAMP (Robbins et al. 1993) and expression of *MITF* (Bertolotto et al. 1998), these interactions seems to form a positive feedback network, as suggested by Adachi et al. (2000).

Interestingly, some of the families of transcription factors inferred here to bind to potential *MC1R* enhancers have been shown to be involved in epidermal-specific (keratinocyte) gene expression, including AP-1, AP-2, ets, Sp1, and possibly also C/EBP and GKLF (Sinha et al. 2000; Kaufman et al. 2002). Factors MZF1, GATA-1, and YY1 have been implicated in the formation of repressor complexes (Raich et al. 1995; Hromas et al. 1996), and their inferred binding to several of the conserved blocks identified here (including CSB 5h) suggests that they may be involved in the tissue-specific silencing of *MC1R* expression.

Comparison of the relative position of conserved segments in the *MC1R* promoter region indicates that blocks 5i, 5j and 5k would be located within the 5' UTR of the *MC1R* transcript, based on the mapping of the likely initiation point (cap site) in humans (Smith et al. 2001). This suggests that their sequence conservation may derive from post-transcriptional functional roles, but they may potentially be involved in transcription factor binding as well. These hypotheses, as well as the others presented above on the basis of evolutionary and computational analyses, suggest multiple *cis* elements and *trans* factors potentially involved in *MC1R* regulation, which can now be tested experimentally.

In the non-coding region 3' of *MC1R*, the comparative data set analyzed here allowed an assessment of evolutionary conservation in the segment reported to be an alternatively spliced second *MC1R* exon (Tan et al. 1999). These authors reported that the human MC1R-B form (containing 65 additional amino acids in the C-terminal intra-cellular domain) is expressed in melanoma cells, fetal heart and testis (Tan et al. 1999), however no further information exists on its functional significance or

occurrence in other species. Subsequent analyses failed to confirm the existence of MC1R-B in human melanocytic cells (Smith et al. 2001). The comparative genomic analyses performed here identified only a short (22 bp) block of conserved sequence between cat and human in this region. Little conservation in this segment (restricted to 9 bp) was detected in a comparison with the mouse sequence, and no observed conservation in this region was identified relative to the cow (Figure 3-1, Table 3-3). These results imply the existence of little functional constraint in the second exon reported to occur in the human MC1R-B isoform, suggesting that it either does not perform a critical role or that its functionality is not conserved among mammalian groups.

Evolution of the *MC1R* coding region

The *MC1R* coding region exhibited a very high GC content at third codon positions, which was correlated with the observation of considerable codon usage bias. Similar findings have been reported for other genomic regions of high GC content, and interpreted to derive either from biased mutation pressure (mutationalist view, in which case third codon positions should accumulate biased mutations at a similar rate as adjacent non-coding [neutral] nucleotides), or from selection favoring a high GC proportion or particular codons (selectionist view) (Li 1997). The pattern observed in *MC1R* does not fit an exclusive mutationalist scenario, since the GC content at third codon positions is *ca.* 50% higher than that observed in surrounding non-coding genomic areas.

An alignment of the *MC1R* coding region among 20 vertebrate species revealed the occurrence of short conserved motifs intercalated with highly variable segments. One-hundred and six amino acid residues were completely conserved among vertebrates, indicating that these sites are likely the ones under the strongest functional constraints. It can be thus expected that variants at these positions should have more meaningful effects on MC1R activity than those identified elsewhere in the gene.

Heterogeneity in evolutionary rates was apparent among lineages and among *MC1R* domains (Table 3-4; Figures 3-6 and 3-7). Evidence for taxon-specific acceleration of amino acid substitution rate was observed in ED4 and TM7, in both cases suggesting that mammals have evolved faster than birds at these particular *MC1R* segments (Table 3-4). At the nucleotide level, in most cases the mouse lineage showed a faster rate of nonsynonymous (and also synonymous [not shown]) change relative to the included primates (Figures 3-6 and 3-7), however in some domains (particularly ID4) this pattern was reversed (Figure 3-6). Also, comparisons within two mammalian orders (Carnivora: cat-canids; and Artiodactyla: Pig-Ruminants) revealed some contrasting patterns of *MC1R* variation. While in some domains (e.g. ED4, TM1) the carnivore comparisons indicated higher rates of nonsynonymous change in this group, in others (particularly TM4) the opposite was observed. In TM4 there were no replacement changes observed among carnivores, whereas estimated nonsynonymous divergence between the pig and ruminants reached 24%, above that observed in the much older comparison of Ferungulates (Carnivores, Perissodactyls, Artiodactyls) *versus* Primates. Such observations, interpreted in the context of overall

patterns of *MC1R* variation, may lead to interesting studies of lineage-specific changes in functional constraint or episodes of adaptive evolution in this gene.

To evaluate the extent of sequence conservation of *MC1R* relative to other genomic loci, the rates of synonymous (r_S) and nonsynonymous (r_N) nucleotide substitution at this gene were estimated using two different evolutionary calibrations (Table 3-5). The human-rodent (mouse) calibration at 80 million years ago (MYA) allowed a direct comparison with rates estimated for other loci, as this calibration has been used in previous studies (Li 1997; Makalowski & Boguski 1998). Rates estimated with the cat-dog calibration likely represent more realistic values for mammals in general, given the acceleration of the mouse lineage (see Figure 3-7) that will tend to inflate rate estimates (Li 1997; Bromham et al. 2000; Eizirik et al. 2001b).

The estimate of r_S for *MC1R* shown in Table 3-5 is similar to the average rates among loci presented by Li (1997) and Makalowski & Boguski (1998), whereas r_N is higher than the average reported by these authors (Li's [1997] mean r_N: 0.74; Makalowski & Boguski's [1998] mean r_N: 0.55 for mouse-human comparisons). Fifteen of the 47 genes surveyed by Li (1997) show a r_N higher than *MC1R*, placing this locus among the top 34% of loci according to rate of nonsynonymous evolution. If the extra-cellular domains of *MC1R* are considered separately, the estimated r_N places these segments among the 17% of loci with the fastest rates. Estimates of r_N are significantly lower ($P<0.05$) than r_S for the transmembrane and intra-cellular domain categories, and also for the total *MC1R* sequence (Table 3-5). This result indicates negative selection (functional constraint) in these regions, affecting estimates for the protein as a whole.

These observations indicate that *MC1R* should not be considered "highly conserved", as argued in previous reports (e.g. Klungland et al. 1999), as it exhibits faster rates of nonsynonymous substitutions than the average of loci surveyed so far. A similar conclusion was reached by Rana et al. (1999) based on an analysis of *MC1R* variation in hominoid primates and a human-rodent comparison relative to other melanocortin receptors. Overall, the observed patterns indicate that *MC1R* sequence variation is influenced by negative selection at an identified subset of its sites, and relaxation of such constraints at others, particularly in the extra-cellular domains. No conclusive evidence of diversification mediated by positive selection was identified, as nonsynonymous rates were lower than or similar to synonymous rates for all fifteen domains, when analyzed individually. One possible exception was the dog-fox comparison in the ED1 domain, where seven out of eight nucleotide changes were nonsynonymous, producing a suggestive though non-significant result (tested by assessing the overlap in confidence intervals, corrected for multiple comparisons; see Chapter 2 for details of testing rationale).

The identification of two variable sites (both causing amino acid changes) in a sample of only seven domestic cat chromosomes (including the BAC clone sequence [Fca273]) suggests that considerable variation exists in the *MC1R* coding region in this species, as observed in other organisms (e.g. humans: Rana et al. 1999; Harding et al. 2000). It can be hypothesized that the identified polymorphic replacements do not significantly affect *MC1R* function, as they involve the exchange of one hydrophobic residue for another, and in both cases the alternative forms at the site in question are seen in other mammals. At codon 57, valine is observed in most mammals, but

methionine occurs in the chimpanzee (Figure 3-5); at codon 296, phenylalanine is found in canids, whereas leucine occurs in all other surveyed mammals and birds (Figure 3-5). If these substitutions are experimentally confirmed to maintain an identical MC1R function, they illustrate the frequency at which homoplasious changes can occur at particular locations of this gene, and possibly also at other loci evolving under similarly heterogeneous functional constraints.

Table 3-1. PCR primers used in this study.

Primer[a]	Sequence (5' – 3')
ext-F4	TCAGCCTGGGGCTGGTG
ext-R2	AGAGGAGCGTAGCCACCCAGAT
ext-R4	GCTGCGGAAGGCGTAGAT
extRH-F1	TTCATCGCCTACTACGATCACA
extRH-R1	CCCCAGCAGAGAAAGAAAATG
extCOM-F1	ATGAAGCCTGCTGGAAGCAC
extCOM-R1	GATATCCCCACCTCCCTCTG

[a] For any primer the suffix containing 'F' refers to forward, and 'R' to reverse.

Table 3-2. Features of conserved non-coding sequence blocks identified in 4.35 kb genomic segment located 5' of the domestic cat *MC1R* coding region (see Figure 3-1 for availability of other mammalian sequences for comparison). Boundaries are defined based on multiVISTA results. Asterisks indicate blocks identified/refined with the Bayes Block Aligner.

Block	Position in domestic cat sequence [a]	Size of conserved block relative to each species (% nucleotide identity)
5a	1032 – 1087	**Cow:** 31 bp (97%) **Human:** 38 bp (90%) **Mouse:** 22 bp (91%)
5b	1124 – 1155	**Cow:** 32 bp (91%) **Human:** 14 bp (100%)* **Mouse:** 26 bp (92%)
5c	1169 – 1204	**Cow:** 36 bp (92%) **Human:** 20 bp (85%)* **Mouse:** 20 bp (95%)
5d	2175 – 2215	**Cow:** 34 bp (94%) **Human:** 31 bp (91%)
5e	2239 – 2274	**Cow:** 23 bp (87%) **Human:** 36 bp (92%); **Mouse:** 8 bp (100%)* + 9 bp (100%)*
5f	3352 – 3373	**Pig:** 22 bp (95%) **Cow:** 21 bp (91%) **Human:** 15 bp (93%) **Mouse:** 8 bp (100%)
5g	3375 – 3397	**Cow:** 23 bp (91%) **Human:** 18 bp (94%)
5h	3910 – 3956	**Horse:** 35 bp (94%) **Pig:** 11 bp (91%) + 29 bp (90%) **Cow:** 43 bp (95%) **Human:** 34 bp (88%) **Mouse:** 48 bp (92%)
5i	4003 – 4044	**Dog:** 9 bp (100%) **Horse:** 42 bp (93%) **Pig:** 12 bp (92%) **Cow:** 11 bp (91%)

Table 3-2. Continued.

Block	Position in domestic cat sequence [a]	Size of conserved block relative to each species (% nucleotide identity)
5j	4067 – 4138	**Dog:** 56 bp (93%) **Horse:** 15 bp (87%) + 15 bp (94%) + 11 bp (100%) **Pig:** 19 bp (95%) **Cow:** 32 bp (91%) **Mouse:** 11 bp (91%)
5k	4211 – 4231	**Dog:** 16 bp (94%) **Horse:** 8 bp (100%) **Cow:** 21 bp (91%) **Human:** 13 bp (92%)

[a] Outer boundaries of conserved blocks between cat and all other compared species.

Table 3-3. Features of conserved non-coding sequence blocks identified in a 3.3 kb genomic segment located 3' of the domestic cat *MCIR* coding region (see Figure 3-1 for availability of other mammalian sequences for comparison). Boundaries are defined based on multiVISTA results.

Block	Position in domestic cat sequence [a]	Size of conserved block relative to each species (% nucleotide identity)
3a	5319 – 5492 [b]	**Dog:** 5 sub-blocks: 11-25 bp (90% - 100%) **Horse:** 2 sub-blocks: 17 –24 bp (88% - 92%) **Pig:** 2 sub-blocks: 11-18 bp (91% - 100%) **Cow:** 5 sub-blocks: 8-16 bp (88% - 100%) **Human:** 2 sub-blocks: 17-21 bp (88% - 95%) **Mouse:** 1 sub-block: 9 bp (100%)
3b	5495 – 5524	**Dog:** 26 bp (89%) **Pig:** 22 bp (91%) **Cow:** 23 bp (96%) **Human:** 12 bp (92%)
3c	5656 – 5677	**Human:** 22 bp (91%) **Mouse:** 9bp (100%)
3d	6045 – 6073	**Cow:** 23 bp (91%) **Human:** 14 bp (93%)
3e	6160 – 6195	**Cow:** 11 bp (91%) + 9 bp (100%) **Human:** 36 bp (92%) **Mouse:** 12 bp (100%)
3f	6694 – 6724	**Cow:** 24 bp (92%) **Human:** 11 bp (91%) **Mouse:** 17 bp (88%)
3g	6928 – 6954	**Cow:** 14 bp (100%) **Human:** 27 bp (93%) **Mouse:** 14 bp (93%)
3h	6994 – 7011	**Cow:** 18 bp (94%) **Human:** 10 bp (90%)
3i	7084 – 7112	**Human:** 27 bp (93%) **Mouse:** 14 bp (93%)

Table 3-3. Continued.

Block	Position in domestic cat sequence [a]	Size of conserved block relative to each species (% nucleotide)
3j	7286 – 7306	**Cow:** 12 bp (100%) **Human:** 18 bp (94%) **Mouse:** 9 bp (100%)
3k	7383 – 7434	**Cow:** 38 bp (95%) **Human:** 42 bp (91%) **Mouse:** 40 bp (95%)
3l	7847 – 7946	**Human:** 56 bp (88%) **Mouse:** 76 bp (90%)
3m	8362 – 8387	**Human:** 25 bp (92%) **Mouse:** 11 bp (100%)
3n	8454 – 8477	**Human:** 8 bp (100%) **Mouse:** 21 bp (91%)

[a] Outer boundaries of conserved blocks between cat and all other compared species.

[b] Block 3a can be divided into five sub-blocks with varying conservation among different species, and varying connectivity depending on analysis method and parameters; it is depicted as a single block here for simplicity. Sub-block positions in cat sequence (defined using multiVISTA with a 10 bp window size) are as follows (species sharing detected conserved sub-block are given in parentheses): 5319-5329 (dog, cow); 5332-5352 (dog, horse, human, mouse); 5374-5397 (dog, horse, cow); 5431-5455 (dog, pig, cow, human); 5470-5492 (dog, pig, cow).

Table 3-4. Amino acid variation in the different domains of the MC1R protein. ED: Extra-cellular domain; ID: intra-cellular domain; TM: transmembrane domain.

Domain	No. Sites	Var. Sites	Cons. Sites	Indels	Mean p-distance between groups (Mammals, Birds, Fish)		
					M vs. B	M vs. F	B vs. F
ED1	41	36	2	3	0.78 ± 0.06	0.84 ± 0.05	0.82 ± 0.06
ED2	19	15	4	0	0.57 ± 0.09	0.69 ± 0.10	0.68 ± 0.10
ED3	4	4	0	0	0.96 ± 0.04	0.78 ± 0.19	0.67 ± 0.16
ED4	14	9	5	0	0.50 ± 0.13	0.64 ± 0.13	0.31 ± 0.12
ED total	78	64	11	3	0.69 ± 0.05	0.76 ± 0.04	0.68 ± 0.05
ID1	9	3	6	0	0.12 ± 0.10	0.23 ± 0.13	0.22 ± 0.14
ID2	22	9	13	0	0.22 ± 0.08	0.22 ± 0.08	0.18 ± 0.08
ID3	27	21	6	0	0.62 ± 0.08	0.60 ± 0.09	0.67 ± 0.09
ID4	22	9	8	5	0.22 ± 0.09	0.42 ± 0.11	0.47 ± 0.12
ID total	80	42	33	5	0.35 ± 0.05	0.40 ± 0.05	0.42 ± 0.05
TM1	26	14	11	1	0.22 ± 0.08	0.46 ± 0.10	0.39 ± 0.10
TM2	27	20	7	0	0.34 ± 0.08	0.38 ± 0.08	0.43 ± 0.09
TM3	22	16	6	0	0.27 ± 0.08	0.47 ± 0.10	0.50 ± 0.11
TM4	21	18	3	0	0.58 ± 0.10	0.64 ± 0.10	0.67 ± 0.10
TM5	25	15	10	0	0.32 ± 0.08	0.39 ± 0.09	0.39 ± 0.09
TM6	25	8	17	0	0.19 ± 0.07	0.25 ± 0.08	0.24 ± 0.08
TM7	22	14	8	0	0.23 ± 0.08	0.40 ± 0.10	0.24 ± 0.09
TM total	168	105	62	1	0.30 ± 0.03	0.42 ± 0.04	0.40 ± 0.04
Total	326	211	106	9	0.40 ± 0.02	0.50 ± 0.03	0.47 ± 0.03

Table 3-5. Rates (± standard error) of synonymous (r_S) and nonsynonymous (r_N) substitution at the *MC1R* gene, estimated with the method of Li (1993) and Pamilo & Bianchi (1993), using two different evolutionary calibrations. Estimates were generated with MEGA 2.1, and the standard error was computed using 1000 bootstrap replications. Rates are expressed in units of substitutions per site per 10^9 years.

MC1R Domains	Human-Rodent calibration [a]	Cat-Dog calibration [b]
Extra-Cellular [73 codons: Human/Mouse] [75 codons: Cat/Dog]	r_S: 3.44 ± 1.24	r_S: 1.97 ± 0.61
	r_N: 1.47 ± 0.39	r_N: 1.26 ± 0.32
Intra-Cellular [75 codons]	r_S: 3.66 ± 1.12	r_S: 3.32 ± 1.05
	r_N: 0.54 ± 0.18	r_N: 0.46 ± 0.17
Transmembrane [167 codons]	r_S: 3.28 ± 0.04	r_S: 2.10 ± 0.44
	r_N: 0.95 ± 0.16	r_N: 0.28 ± 0.10
Total [315-317 codons]	r_S: 3.38 ± 0.38	r_S: 2.31 ± 0.34
	r_N: 0.96± 0.11	r_N: 0.54 ± 0.09

[a] Calculated on the basis of an evolutionary calibration of 80 million years ago (MYA) for the split between humans and rodents, as applied by Li (1997) and followed by Makalowski & Boguski (1998) (see Discussion).

[b] Calculated on the basis of a fossil calibration of 50 MYA for the divergence between cat and dog (Benton 1993, Eizirik et al. 2001b, Eizirik et al. in press).

Figure Legends:

Figure 3-1. Schematic of the domestic cat *MC1R* genomic region characterized from a sub-fragment of BAC clone RPCI86-248I10. Length of the domestic cat sequence is 8,663 bp (a scale shown in the top right corner); horizontal lines indicate the available portion of comparable sequence from other mammalian species. The single *MC1R* coding exon is shown as a dark box (positions 4351-5304 in the cat sequence), and evolutionarily conserved sequence blocks (CSBs) in non-coding regions are depicted as vertical rectangles / lines. The exact location of each conserved block in the cat sequence is shown in Tables 3-2 and 3-3, along with its length and similarity to other species. CSBs on the 5' non-coding region of the *MC1R* gene are numbered 5a-5k; those on the 3' region are numbered 3a-3n. Conserved blocks were defined using multiVISTA comparisons (in most cases supported by similar BLAST results): open rectangles are blocks identified using a 20-bp window size and a minimum of 88% identity; blocks detected only using a 10-bp window size (90% minimum identity) are shown as black vertical lines. Block 5f in the pig sequence (shaded in gray) was identified only with the BLAST analysis (21/22 identical nucleotides). Bayes Block alignments performed for the cat *MC1R* 5' region relative to human and mouse corroborated the mVISTA and BLAST results, with four exceptions: the presence of blocks 5b and 5c in the human sequence (hatched rectangles) was detected only with this approach (with 0.8 to 1.0 peak probability of alignment), and the presence of blocks 5f and 5j was not confirmed in the mouse (dotted lines) using this method.

Figure 3-2. Multiple-species VISTA plot indicating areas of high sequence conservation in the *MC1R* genomic region between cat and other mammals. Areas in blue denote the *MC1R* coding region, areas in red are conserved non-coding segments. The X axis represents base-pair positions in the domestic cat genomic sequence; the Y axis indicates sequence identity for each pairwise comparison (from the top: cat/cow, cat/human, cat/mouse) using a window size of 20 bp. Only segments exhibiting at least 70% sequence identity in each pairwise comparison are shown in the graph and a minimum level of conservation of 88% (empirically determined) was used to highlight conserved non-coding segments.

Figure 3-3. Nucleotide sequence of the domestic cat *MC1R* coding region (from clone RPCI86-248I10, derived from individual Fca273) aligned with that of dog (GenBank accession AF064455) and human (accession AF326275). Dots indicate identity to top sequence. Transmembrane (TM) domains 1-7 are shown in boldface, and delimited by a line above the sequence.

Figure 3-4. Alignment of conserved sequence block 5h (see Figure 3-1) for eight species representing five mammalian orders. Nucleotide positions in the domestic cat genomic sequence are indicated on top; dots indicate identity to cat sequence. Sites 3910-3952 in the cat sequence (upper case) comprise the core conserved segment of block 5h (arrow indicates the 5' end of block 5h). Adjacent sites 3905-3909 (lower case) are included to show the second E-box (underlined) present in this region in the mouse, chimpanzee (Chimp in the figure) and gorilla. The main E-Box, conserved

across all taxa, is delimited by a box. The nucleotide site flanking the 5' end of this E-Box (a 'T' in all species but humans) is shaded, as is the 'A' on the opposite flank in the horse (see Results and Discussion). The Sp1 motif adjacent to the E-box is indicated by the double underline in the cat sequence. Asterisks indicate polymorphic sites in humans (Makova et al. 2001) that are conserved across all other taxa. Accession numbers are listed in the Materials and Methods, except for the chimpanzee (AF387969) and gorilla (AF387968) promoter sequences.

Figure 3-5. Alignment of the inferred amino acid sequence of the domestic cat (*Felis catus*, Fca273) MC1R with those from 15 other mammals, three birds and one fish. Represented mammalian species are as follows (see Materials and Methods for full species name and Accession Numbers): domestic dog , red fox, pig, cow, goat, sheep, muskox, reindeer, fallow deer (*C. dama*), red deer (*C. elaphus*), moose, horse, human, chimpanzee (Chimp) and mouse. Birds are represented by chicken, bananaquit (*Coereba*) and tanager (*Tangara*); the fish is *Takifugu rubripes*. Dots indicate identity to top sequence. The beginning position of each of the 15 different MC1R domains is indicated above the sequences: extra-cellular domains (ED) are colored in blue, trans-membrane domains (TM) in red, and intra-cellular domains (ID) in black. Domain boundaries are based on Robbins et al. (1993). Asterisks indicate the two variable positions identified in domestic cats. Amino acid residues completely conserved across the surveyed vertebrates are shaded.

Figure 3-6. Graph showing estimated number of nonsynonymous substitutions per nonsynonymous site (calculated using the Kumar modification of the Pamilo-Bianchi-Li approach, see Methods) for each *MC1R* domain, at varying levels of phylogenetic depth among mammalian groups. Substitutions are estimated as the mean of pairwise comparisons between the two groups in question, using the nucleotide sequences for all the mammalian taxa shown in Figure 3-4. Groups are as follows: Canids: dog, fox; Ruminants: cow, goat, sheep, muskox, reindeer, fallow deer, red deer, moose; Primates: human, chimpanzee. Ferungulates are represented in this data set by Carnivores (Cat + Canids) + Artiodactyls (Pig + Ruminants). Pairwise group comparisons span different phylogenetic depths in the mammalian evolutionary tree: Cat-Canids: 50 million years ago (MYA); Pig-Ruminants: 60 MYA, Ferungulates-Primates (= Ferungulates-Mouse): *ca.* 94 MYA. Arrows indicate unusual patterns of molecular evolution observed in the phylogenetic comparisons involving ID4 and TM4 (see Results and Discussion).

Figure 3-7. Phylogenetic tree of 16 mammalian *MC1R* nucleotide sequences, generated using only the first and second codon positions (composing a data set with 642 bp), and rooted using three bird outgroups. The topology and branch lengths were estimated using a maximum likelihood (ML) approach, with a GTR+G+I model of nucleotide evolution, and parameters estimated from the data set. Similar results (and increased acceleration of the mouse lineage) were observed when third codon positions were also included. The same pattern was inferred when the topology was constrained to conform to recent estimates of mammalian relationships (Eizirik et al.

2001b; Murphy et al. 2001), and also using different phylogenetic methods (not shown).

Figure 3-8. Nucleotide variation in the felid *MC1R* gene. The figure shows an alignment of *MC1R* nucleotide sequences from four domestic cat individuals (Fca273 is the sequence from BAC clone 248l10), and one individual each from 12 additional cat species: *Felis nigripes* (Fni), *Prionailurus bengalensis* (Pbe), *Panthera onca* (Pon), *Herpailurus yaguarondi* (Hya), *Puma concolor* (Pco), *Acinonyx jubatus* (Aju), *Pardofelis marmorata* (Pma), *Catopuma temmincki* (Pte), *Caracal caracal* (Cca), *Profelis aurata* (Pau), *Leptailurus serval* (Lse) and *Otocolobus manul* (Oma); the last line is an *MC1R* sequence from a genet (*Genetta genetta*: Carnivora, Viverridae), a feloid carnivore outgroup. Only variable sites are shown; numbers on top (vertical notation) refer to nucleotide positions in Figure 3-3. Nucleotide positions involved in an amino acid replacement are shaded

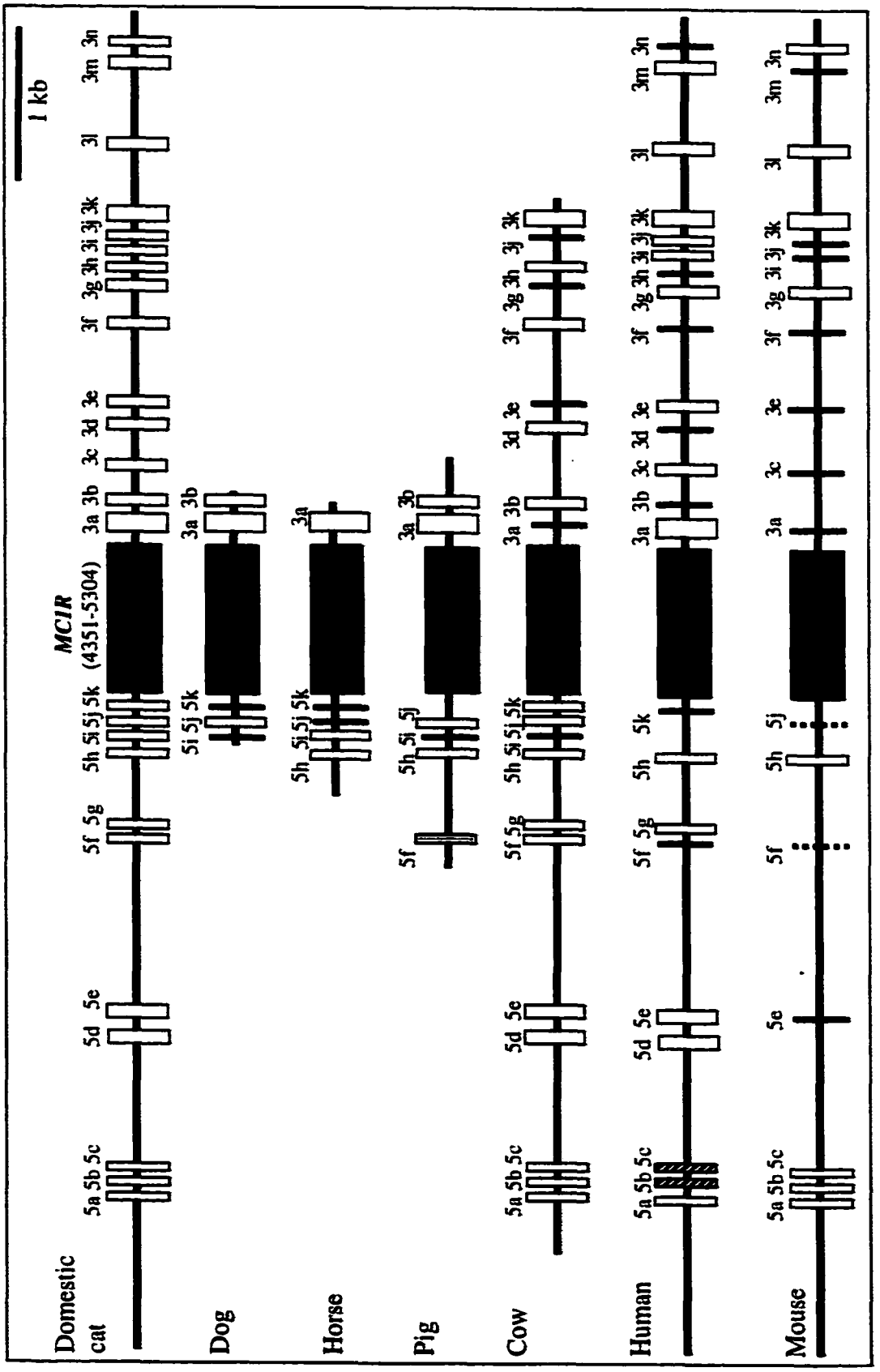


Figure 3-1

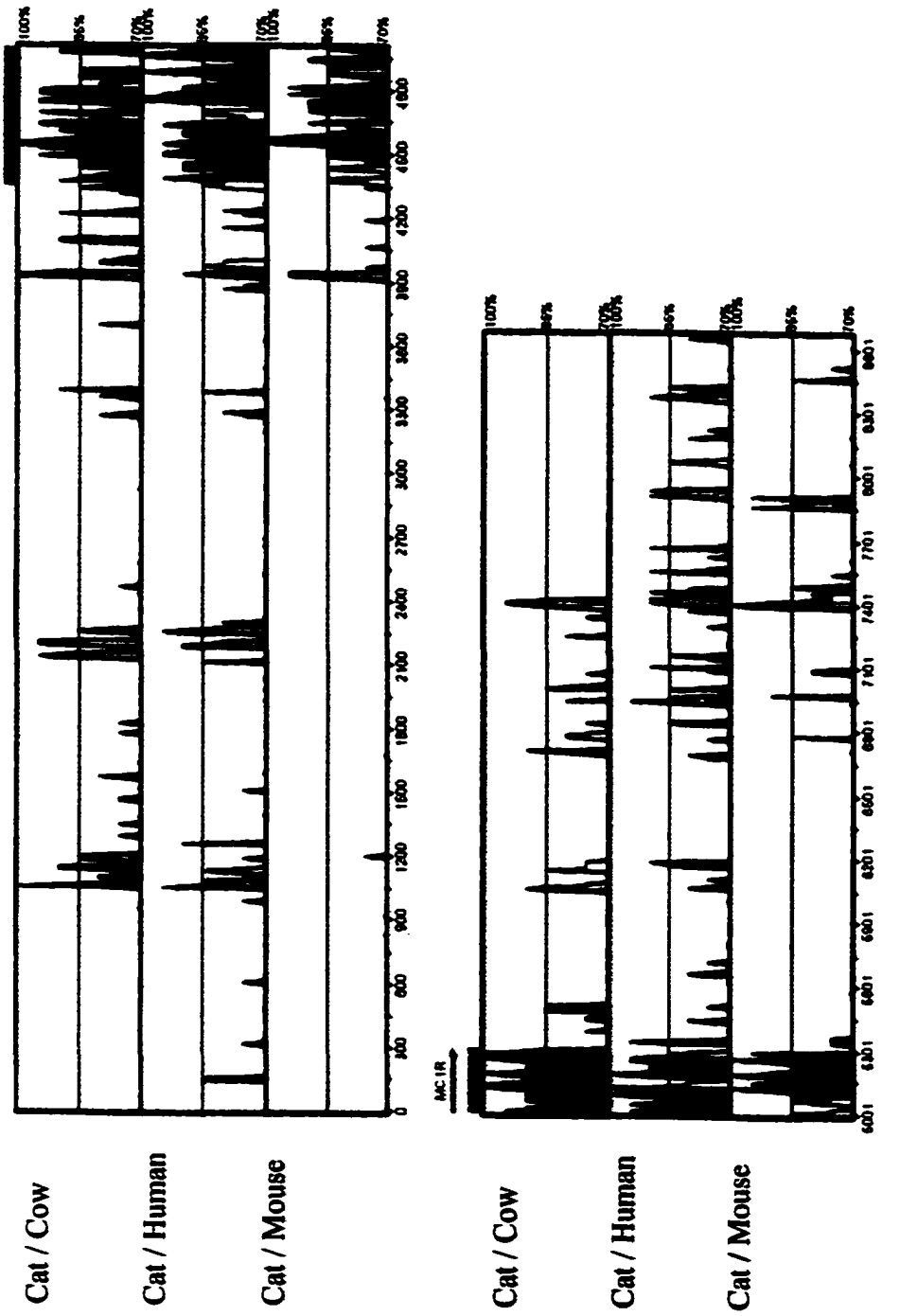


Figure 3-2

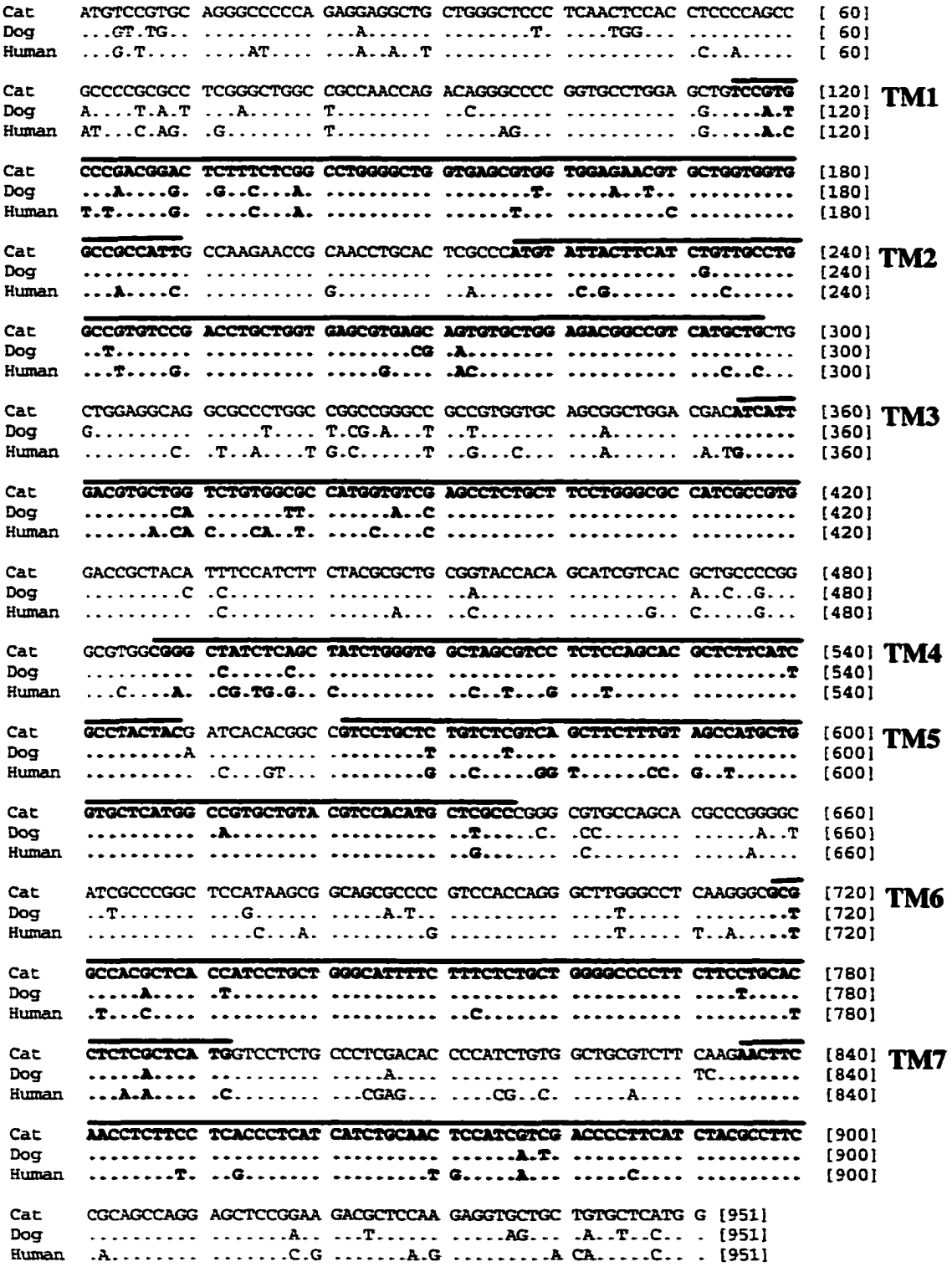


Figure 3-3

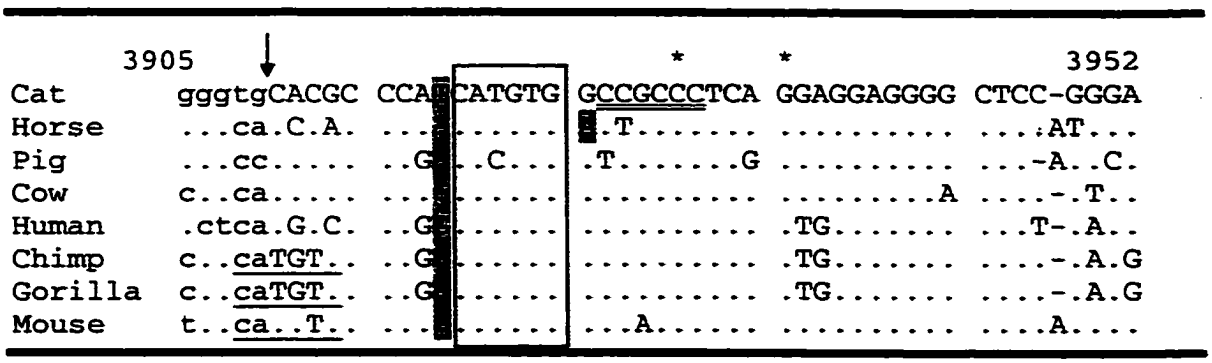


Figure 3-4

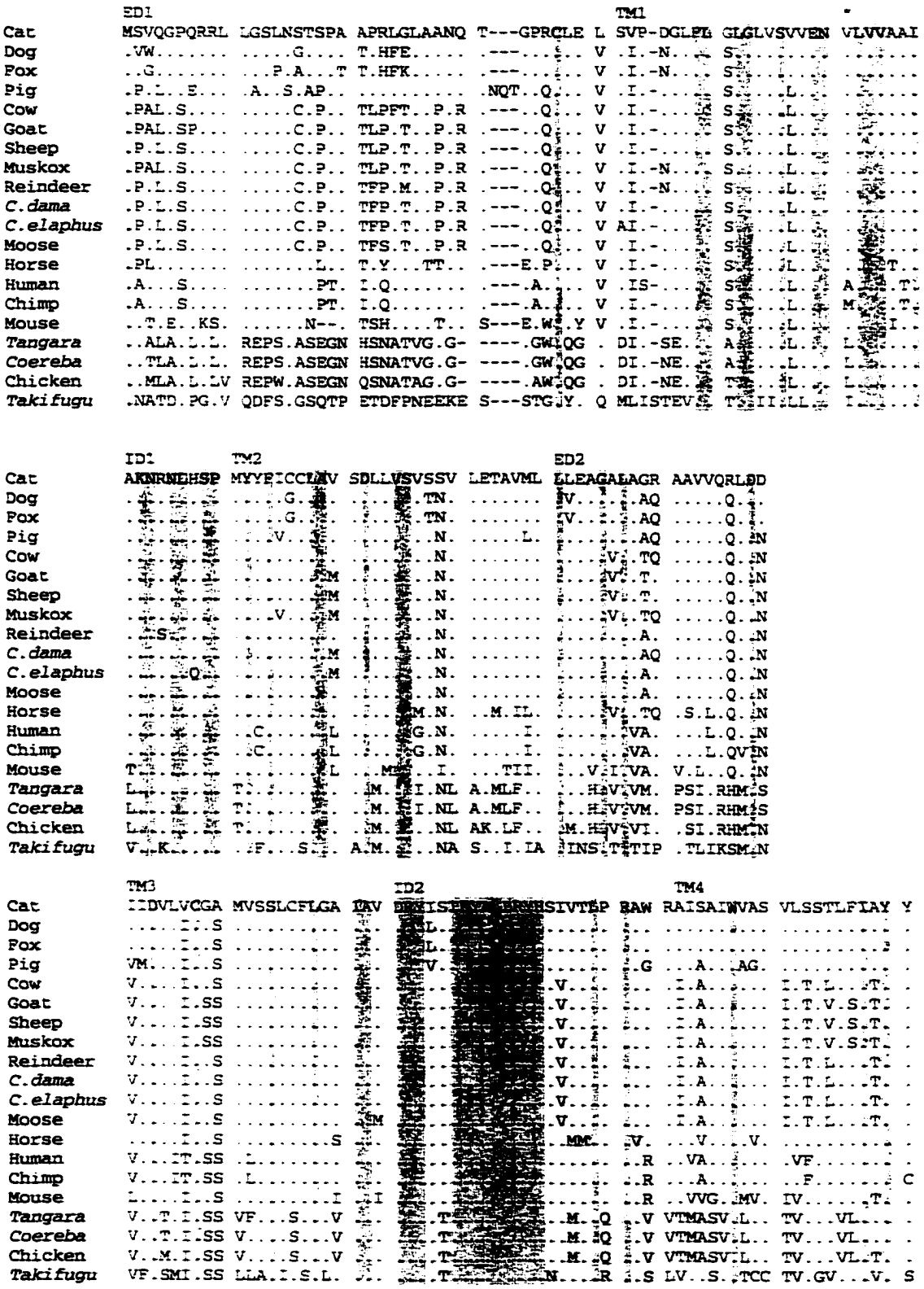


Figure 3-5

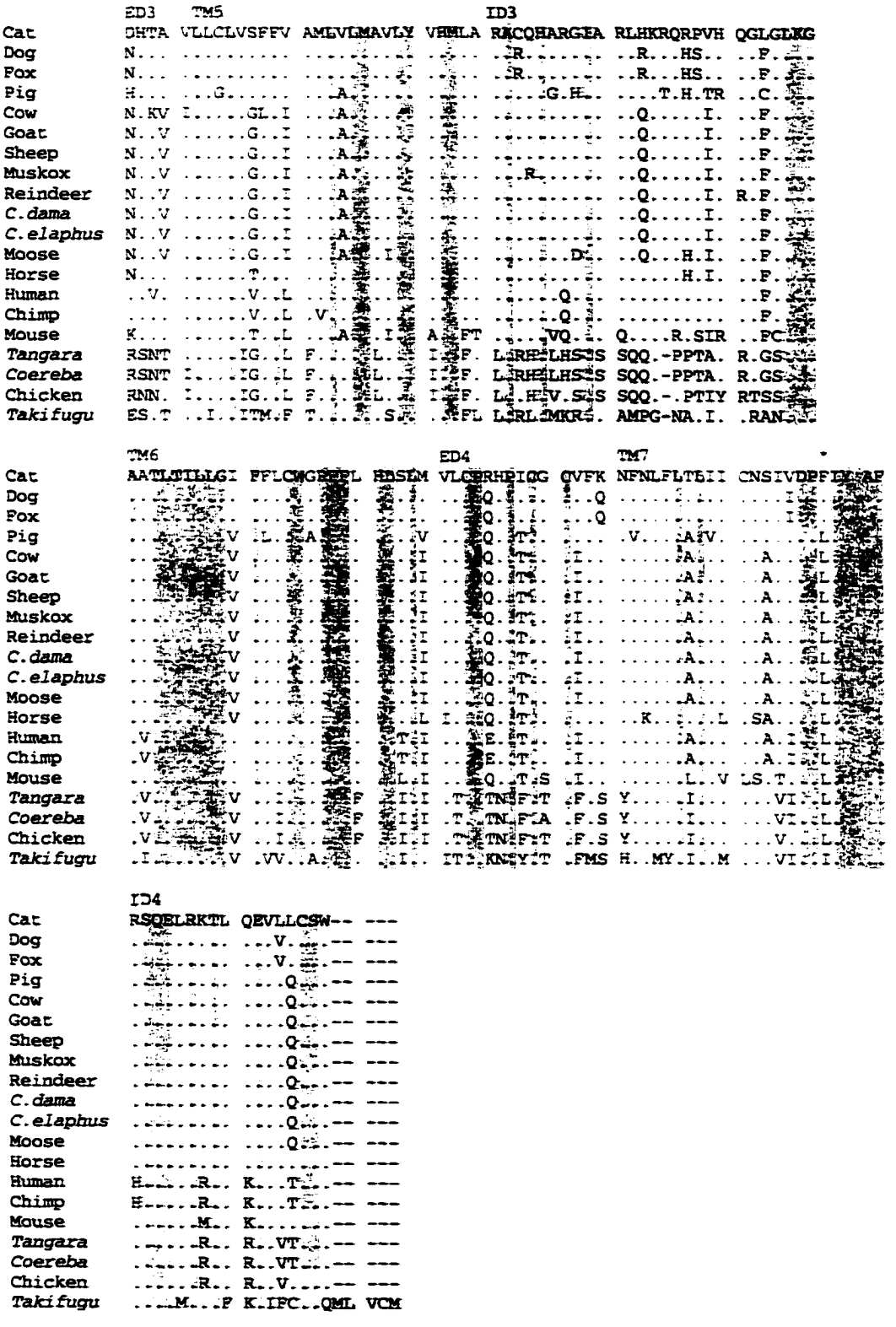


Figure 3-5 (continued)

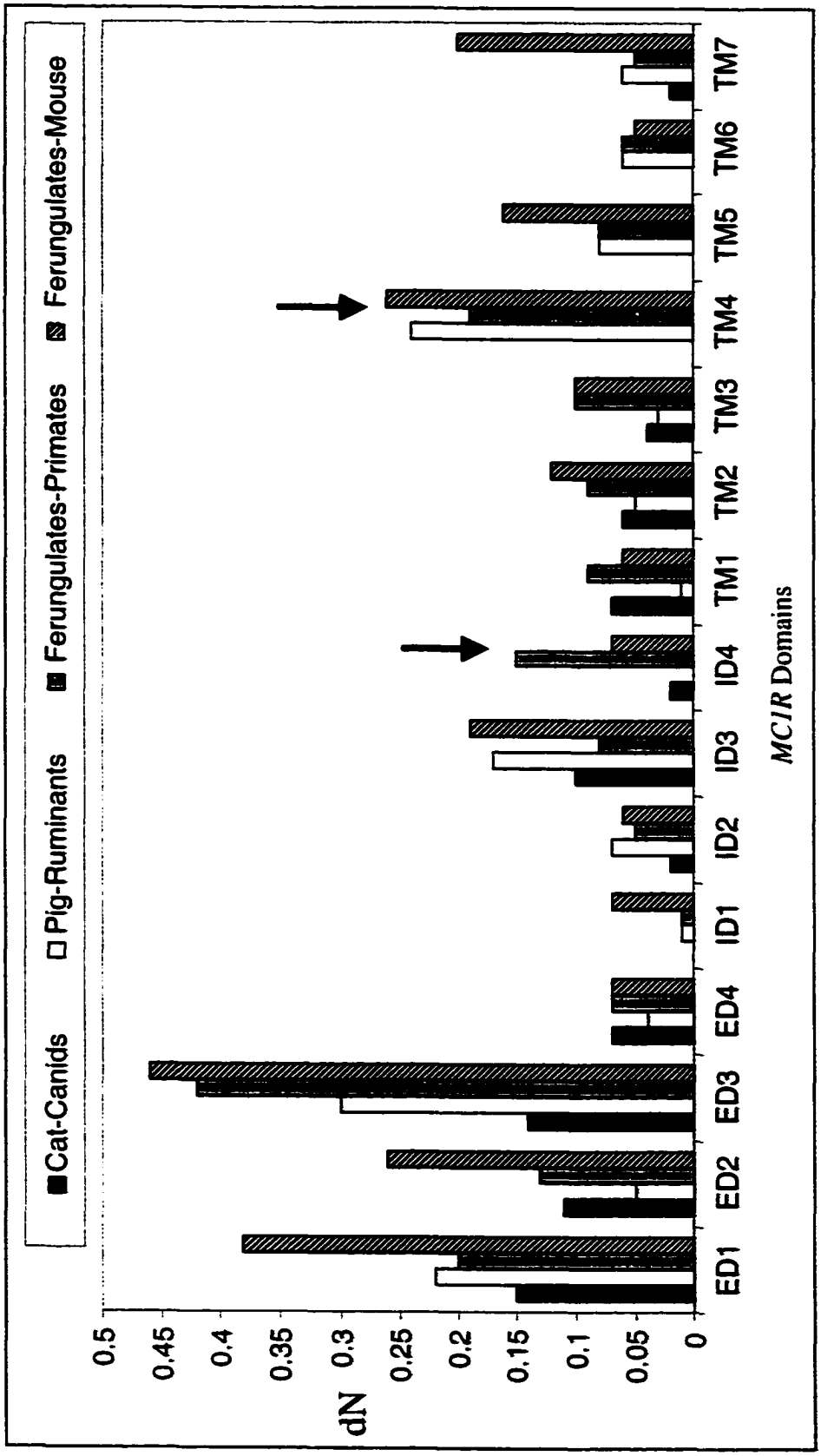


Figure 3-6

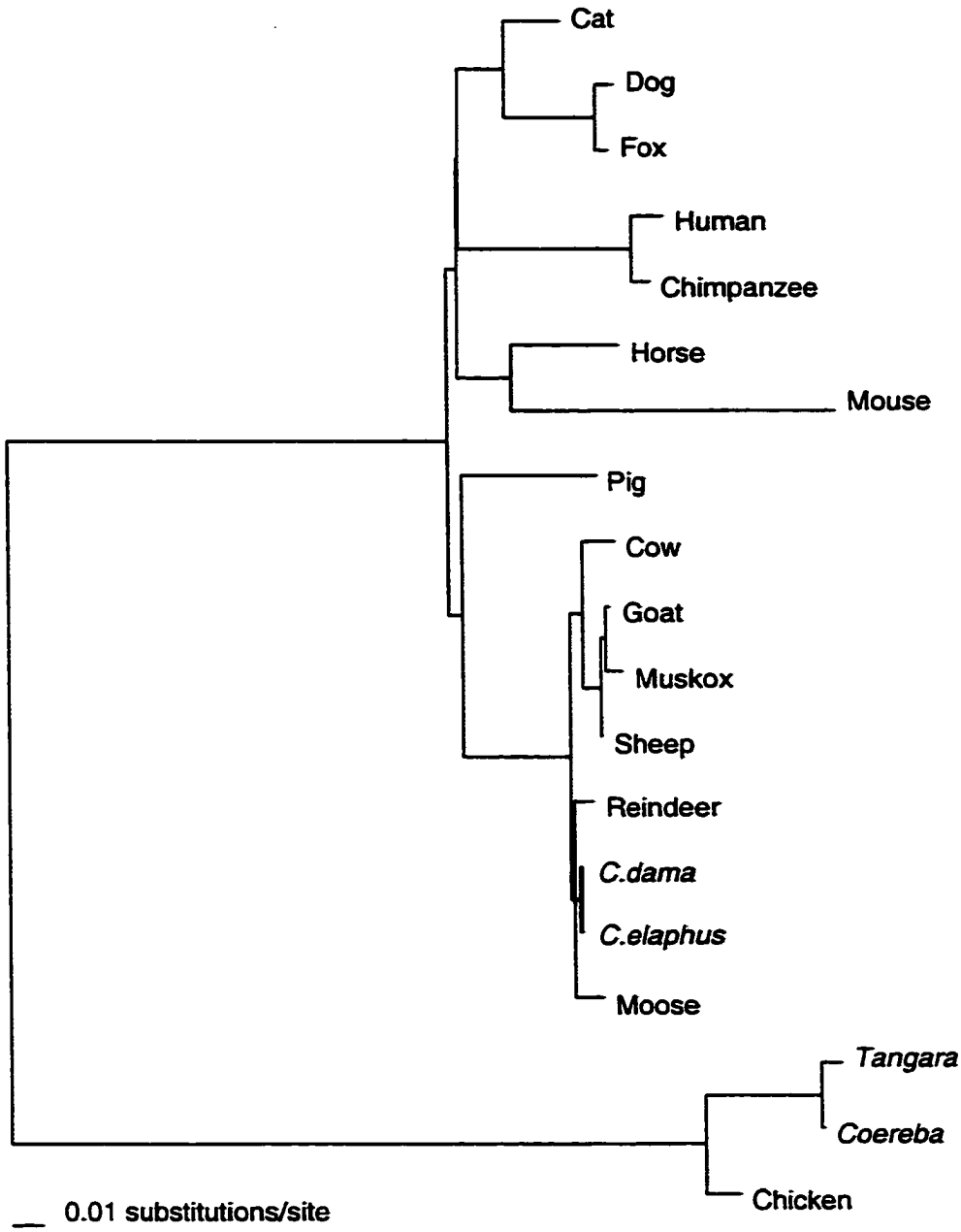


Figure 3-7

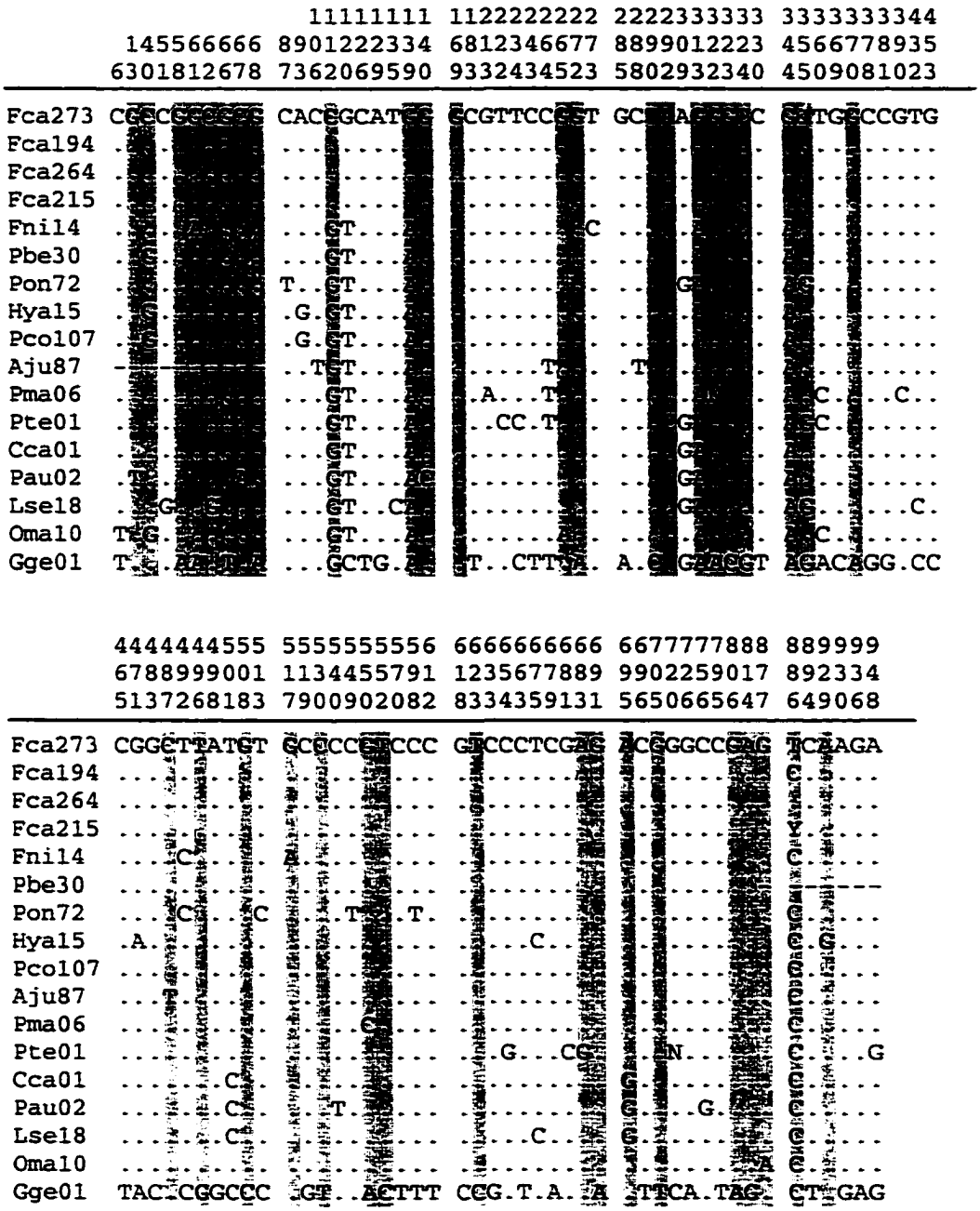


Figure 3-8

CHAPTER 4

Molecular Genetics and Evolution of Melanism in the Felidae

Abstract

Melanism occurs as a polymorphism in at least eleven cat species, reaching high frequencies in some populations but never achieving fixation. To understand the origin and evolution of melanism in the cat family, we have characterized the domestic cat homologues of two candidate genes (*ASIP* and *MC1R*), and identified three independent deletions associated with dark coloration in three different species. A 2 bp deletion in the *ASIP* gene specifies melanism (black color) in the domestic cat, whereas two distinct exonic deletions in the *MC1R* gene are associated with gain-of-function and dark coloration in the jaguar and jaguarundi. Evidence for natural selection favoring melanism was apparent in the jaguarundi, where the inferred ancestral coloration (reddish) has been largely replaced by dark individuals with a derived *MC1R* mutation. Melanistic individuals from five other felid species did not carry any of these three mutations, implying at least four independent origins for melanism in the cat family, and multiple episodes of frequency increase possibly mediated by natural selection.

Introduction

The genetic basis of pigmentation in mammals has fascinated biologists since the late 19th century (e.g. Wright 1918; Cott 1940; Haldane 1977; see Chapter 1). Molecular studies have recently identified variants associated with coloration mutants in several mammalian species, principally laboratory, domestic or farm animals (e.g. Bultman et al. 1992; Mountjoy et al. 1992; Robbins et al. 1993; Jackson 1994; Hustad et al. 1995; Barsh 1996; Våge et al. 1997, 1999; Kijas et al. 1998; Newton et al. 2000; Rieder et al. 2001; Ritland et al. 2001). Nonetheless, little is known about the molecular basis and adaptive significance of natural coat color variation in free-ranging mammals, and so far no study has jointly addressed these issues in multiple polymorphic species from the same family. The cat family (Felidae) exhibits a wide diversity of coat colors and patterns (Weigel 1961; Nowell & Jackson 1996; Nowak 1999) including melanism in at least eleven species (see Chapter 1) and different color 'phases' in others, such as the jaguarundi (*Herpailurus yaguarondi*). In this small Neotropical cat the coloration varies from dark brown/gray to a lighter reddish hue, with the dark form being most common throughout the species' geographic range, and thus widely regarded as the wild type (Robinson 1970a, 1976; Kitchener 1991).

Molecular genetic studies in mice have identified several genes involved in pigmentation phenotypes (Silvers 1979; Jackson 1994; Barsh 1996), and three loci in particular have been shown to be involved in melanism: *agouti/ASIP* (*Agouti Signaling Protein*) (Bultman et al. 1992; Miller et al. 1994; Hustad et al. 1995), *extension/MC1R* (melanocortin-1 receptor) (Mountjoy et al. 1992; Robbins et al. 1993) and *mahogany/ATRN* (Gunn et al. 1999; Nagle et al. 1999). These genes

participate in the biochemical switch between eumelanin (dark pigment: black or brown) and pheomelanin (light pigment: red or yellow) production in hair-follicle melanocytes (Jackson 1994; Barsh 1996; He et al. 2001; see Chapter 1 for more details). Loss of function mutations in the mouse *agouti* gene cause recessively inherited melanism, while gain-of-function mutations in *MC1R* confer a dominant form of melanism (Jackson 1994; Siracusa 1994; Barsh 1996; Lu et al. 1998). Similar phenotypic effects have been observed in other mammalian species (e.g. Våge et al. 1997, 1999; Kijas et al. 1998; Rieder et al. 2001; reviewed in Chapters 1, 2 and 3).

Melanism (black coloration) is inherited as a monogenic autosomal recessive trait in the domestic cat (*Felis catus*) (Whiting 1918; Robinson 1959, 1991), suggesting *agouti/ASIP* as a potential candidate gene. In contrast, melanism has been reported to be inherited as a dominant autosomal trait in jaguars (*Panthera onca*) (Dittrich 1979), suggesting that *MC1R* might be implicated in this phenotype in this species. To allow the investigation of the involvement *ASIP* and *MC1R* in melanism in these and other felids, I have previously mapped, cloned and sequenced the domestic cat homologues of these genes, and developed genomic tools to study them in other species (Chapters 2 and 3). The domestic cat *ASIP* gene maps to chromosome A3 (Chapter 2), and *MC1R* to chromosome E2 (Chapter 3), in both cases corresponding to the homologous genomic position of their human counterparts (Murphy et al. 2000). The feline *ASIP* gene consists of three coding exons (Figure 4-1A) as in other mammals, but comprises 405 bp (135 codons) as opposed to 393-396 in other species (e.g. Bultman et al. 1992; Miller et al. 1993; Rieder et al. 2001), due to a three-residue insertion after codon 84 (see Chapter 2 for more details). The cat

MC1R gene consists of an intron-less 951 bp (317 codons) open reading frame, similar in structure to other mammalian homologues (e.g. Mountjoy et al. 1992; Robbins et al. 1993; Kijas et al. 1998) but exhibiting abundant sequence variation relative to other species (see Chapter 3 for details).

In this study I describe the occurrence of genetic linkage between melanism and genomic markers adjacent to the *ASIP* gene, and characterize at the sequence level a molecular variant at this locus that is implicated in this phenotype. I also extend the molecular analysis of the *MC1R* gene to additional felids exhibiting melanism, and describe two independent 'in-frame' deletions in this gene associated with melanism in two different species. I investigate the occurrence of these molecular variants in multiple cat species, and discuss these findings in the context of the evolutionary history and adaptive significance of melanism in the Felidae, also drawing inferences on the functional implications of the identified mutants in the ASIP and MC1R proteins.

Materials and Methods

Analyses of the *ASIP* gene

To investigate whether *ASIP* is implicated in melanism in the domestic cat, linkage analysis was performed using STR markers tightly linked to this gene, and an 89-individual pedigree comprising 109 melanism meioses (listed in Appendix 3). This pedigree is a subset of the Purina Pedigree, a 259-individual kindred of domestic cats which was characterized and expanded as part of this study, to be used as a tool for the investigation of coat color genes in felids, and also for construction of reference

linkage maps. Specific crosses were performed in the years 1998 and 1999 to maximize the information content of this pedigree for melanism.

To identify appropriate markers for linkage analysis, PCR primers were designed for all 20 STR loci identified in the *ASIP* BAC clone RPCI86-188e3 (Chapter 2), and tested for polymorphism in the pedigree founders as well as a random sample of domestic cats, using dUTP incorporation and an ABI 377 automated sequencer. Eight STR loci were found to be polymorphic (Table 4-1), three of which were employed in the linkage analysis using fluorescently labeled primers to genotype all animals in the pedigree. Genotypes were checked and categorized with Genotyper (ABI); Mendelian inheritance and pedigree accuracy were verified using the program PedCheck (O'Connell & Weeks 1998); and linkage analyses were performed with the software LINKAGE (Lathrop et al. 1984).

Exon-specific primer pairs (agoEx2-F1/R1; agoEx3-F1/R1; agoEx4-F1/R1; see Table 4-2 for list of all primers) were used to obtain the full *ASIP* coding sequence of unambiguously melanistic or non-melanistic domestic cats. Amplifications with primer pairs agoEx2-F1/R1 and agoEx3-F1/R1 used a touchdown profile (40 cycles with annealing temperature decreasing from 60°C to 50°C); PCR with the agoEx4-F1/R1 pair used 35 cycles with 59.5°C annealing temperature. All reactions used *Taq Gold* (Perkin Elmer) and 1.5 mM $MgCl_2$. After identification of a 2 bp deletion in exon 2 of the melanistic animals, a new primer set (agoFcaDel-F1/R1; 153 bp PCR product in the non-deleted allele) was designed around its location, and modified to contain a fluorescent label to allow size-based genotyping of this variant. This segment was amplified using 30 cycles and 58°C annealing temperature. Genotyping

was checked for accuracy by sequencing in several animals, and then carried out as above in the Purina pedigree (listed in Appendix 3) and a panel of unrelated cats (listed in Appendix 4), including redundant typing of multiple individuals.

Analyses of the *MC1R* gene

Using the PCR primers extCOM-F1 and extCOM-R1 (Table 4-2) developed in Chapter 3, the full coding region of the *MC1R* gene was amplified and sequenced in four non-melanistic (wild-type color) and four melanistic jaguars, as well as two jaguarundis of different colorations. Cloning of the PCR products [using a Topo TA kit (Invitrogen)] from the melanistic jaguars revealed that they were heterozygotes carrying the wild type allele and a 15 bp deletion allele. Direct sequencing of the jaguarundi *MC1R* gene revealed that the dark individual was a homozygote for a 24 bp deletion adjacent to that found in melanistic jaguars. A single fluorescent genotyping assay for both *MC1R* deletions was developed, using primer pair extDel-F1/R1, *Taq Gold*, 1.5 mM $MgCl_2$ and 33 cycles with 58°C annealing temperature (173 bp PCR product for non-deleted alleles). This assay was applied to a broad geographic sample of each species, as well as other felids (Appendix 4), using the same methodology described above.

Biological Samples

DNA samples from wild and captive jaguars spanning the whole geographic range of the species (Mexico to southern Brazil) were previously obtained from field projects and captive breeding institutions for use in population genetic studies (Eizirik

et al. 2001a). Additional samples for this study, including free-ranging melanistic jaguars, were obtained from field ecology projects (P.E. Morro do Diabo, SP, Brazil [collected by L. Cullen, Instituto de Pesquisas Ecológicas, Brazil]; and UHE Porto Primavera, MS/SP, Brazil [collected by D. Sana, Associação Pró-Carnívoros and CENAP/IBAMA, Brazil]) as well as captive breeding facilities (EFBC's Feline Conservation Center, CA, USA; Fundação Zoobotânica, RS, Brazil). The new samples included a pedigree of eight captive-bred jaguars. For the new samples (whole blood, blood clots or primary fibroblast cultures), DNA was extracted using a standard phenol/chloroform protocol (Sambrook et al. 1989) or a DNeasy kit (Qiagen). DNA samples from wild and captive jaguarundis spanning the whole range of the species (Mexico to Argentina) and representing different colors (from very dark to dark brown/gray to reddish) were obtained for use in population genetic analyses (W. E. Johnson et al., in preparation). A full list of all samples analyzed in this study (including domestic cats, jaguars, jaguarundis and 22 other wild cat species) is given in Appendix 4.

Results and Discussion

Implication of *ASIP* in domestic cat melanism

Linkage analysis performed with three STR loci isolated from the immediate vicinity of the *ASIP* gene revealed significant linkage with no recombination between these markers and the melanism phenotype in domestic cats. Highest LOD scores were at 0.0 cM: 16.39, 11.35, and 11.03 for FCA708, FCA718, and FCA719,

respectively. A complete list of genotypes for these three loci in the analyzed pedigree is given in Appendix 3.

Sequence characterization of the *ASIP* gene in multiple domestic cats revealed that black individuals were homozygous for an allele (named *ASIP*-Δ2) where a 2 bp deletion at nucleotide positions 123-124 induces a frame-shift in the inferred protein, causing complete loss of the C-terminal active domain (Figure 4-1). Family transmission analysis in the same pedigree used for the linkage study demonstrated perfect co-segregation between the *ASIP*-Δ2 allele and black coloration, conforming to the recessive mode of inheritance [Robinson 1959, 1976; Searle 1968] (Figure 4-2). Genotyping of the presence of this deletion in unrelated individuals showed that 57 black domestic cats collected throughout the world were homozygous for the *ASIP*-Δ2 allele, whereas 26 non-black individuals carried at least one wild-type allele (Figure 4-3), demonstrating perfect association between coloration phenotype and molecular genotype ($P < 0.001$ in a Fisher's exact test using a recessive mode of inheritance). A genotyping survey of 20 other cat species (including melanistic individuals for seven of them) representing all major lineages in the Felidae revealed no other individual bearing this deletion allele (Figure 4-3).

Implication of *MC1R* in jaguar and jaguarundi melanism

I confirmed the dominant mode of inheritance of melanism in jaguars by performing transmission analysis in a 116-individual captive pedigree (shown in Appendix 5). Given the dominant inheritance and the absence of the *ASIP*-Δ2 mutation in melanistic jaguars (Figure 4-3), I tested whether the *MC1R* gene is

implicated in jaguar melanism by obtaining its full coding sequence in several wild-type (yellowish background coloration) and black individuals. Melanistic animals were found to carry at least one copy of a mutant allele bearing a 15 bp (five-codon) in-frame deletion at positions 301-315 (Figure 4-4). This allele (designated *MC1R*-Δ15) also displayed two non-synonymous nucleotide substitutions immediately adjacent to the deletion (changing a CTG codon to ACG), which result in a Leu/Thr replacement relative to the ancestral jaguar sequence at a codon that is otherwise conserved across mammals (Figure 4-4). An additional variable site was identified at position 825, where a synonymous T/C polymorphism was observed. The non-deleted alleles contained either C or T at this position, whereas the surveyed *MC1R*-Δ15 alleles all had a T at that site.

Forty-six jaguars differing in coloration phenotype were examined for their *MC1R* genotypes. Ten unrelated melanistic jaguars were either homozygous or heterozygous for the *MC1R*-Δ15 allele, whereas all 36 wild-type coloration jaguars (sampled from the whole range of the species [Mexico to southern Brazil]) were homozygous for the wild-type allele ($P<0.001$ for color-genotype association using a dominant model, Figure 4-5). Transmission analysis in a captive jaguar pedigree demonstrated exact co-segregation between the deletion genotype and the melanistic phenotype, supporting a dominant mode of inheritance mediated by the *MC1R*-Δ15 allele (Figure 4-6).

Neither *ASIP*-Δ2 nor *MC1R*-Δ15 were seen in a screen of 40 jaguarundis including animals which varied in coat color from very dark brown/gray to red

(Figures 4-3 and 4-5). However, a different in-frame deletion (designated *MC1R*-Δ24) was discovered in jaguarundis, which removed 24 bp (8 codons) at a position adjacent to, but distinct from the deletion seen in the jaguar *MC1R*-Δ15 allele (Figure 4-4). The jaguar *MC1R*-Δ15 and the jaguarundi *MC1R*-Δ24 deletions seem to be independent, given the sequence homology of adjacent nucleotides, the fact that each species belongs to a separate lineage in the Felidae (Johnson & O'Brien 1997; Pecon-Slattery & O'Brien 1998), and the persistence of non-deleted (ancestral) alleles in both of them (Figures 4-4 and 4-5). In addition to the presence of the deletion, the jaguarundi *MC1R*-Δ24 allele also differed from the ancestral-type sequence at three amino acid positions (P22L, I63V and Q310R; see Figure 4-4). These substitutions may influence MC1R activity in the jaguarundi, but their conservative nature (P22L, I63V) and/or the occurrence of identical or similar residues at these positions in other mammals (22L is found in cattle; 310K occurs in humans and mice) suggest that their impact on the protein structure and function is not as significant as that of the deletion itself.

Genotyping of the *MC1R* gene among 40 unrelated jaguarundis sampled across the entire geographic range of the species (Mexico to Argentina) revealed widespread occurrence of the *MC1R*-Δ24 variant, and a significant association with coat color ($P<0.001$ based on a dominant inheritance mode; n=29 with known coloration: Figure 4-5). Jaguarundi individuals bearing the *MC1R*-Δ24 allele, and particularly those homozygous for it, were consistently darker than the ancestral-type homozygotes, which were in turn exclusively red/reddish in coloration (Figure 4-5). These results implicate *MC1R*-Δ24 as a derived melanistic variant responsible for jaguarundi coat

color polymorphism. Interestingly the reddish color, which heretofore was considered as mutant, is actually the ancestral-type presentation, based on its *MC1R* genotype.

Molecular evolution of melanism in the Felidae

The results from this study strongly suggest that the *ASIP* and *MC1R* deletions identified here have causative effects on the occurrence of melanism in three different Felidae species. The *ASIP*-Δ2 deletion identified in the domestic cat likely leads to complete loss of function, as the protein sequence is totally modified after that position (Figure 4-1) and an early stop codon removes most of the biologically critical C-terminal domain (Perry et al. 1996; Miltenberger et al. 2002; see Chapter 2). Complete loss of function at *ASIP* has also been associated with recessively inherited extreme melanism in the mouse, rat and horse (Hustad et al. 1995; Miltenberger 2002; Kuramoto et al. 2001a; Rieder et al. 2001).

The two distinct *MC1R* deletions identified in melanistic jaguars and jaguarundis are associated with a dominant (or semi-dominant) effect, as observed in 'gain-of-function' melanistic *MC1R* mutations in other mammals (e.g. Robbins et al. 1993; Våge et al. 1997, 1999; Kijas et al. 1998). All such previously described mutations have been amino acid substitutions located in the same region of the gene (Figure 4-7), and were shown or inferred to cause constitutive activation or increased basal signaling for eumelanin, most likely due to conformational changes (Robbins et al. 1993; Våge et al. 1997, 1999). In-frame deletions in this region of *MC1R* have not been reported for any species, and our results indicate that they can have a similar effect on MC1R function. Under the alignment scheme presented in Figure 4-4, the

observation that the jaguar *MC1R*-Δ15 and the jaguarundi *MC1R*-Δ24 deletions overlap by two amino acids (101L, 102E) suggests a critical role for these residues in mediating MC1R inactivation. Additionally, the next two amino acids deleted in the jaguarundi *MC1R* (99L, 100L) have been previously implicated in melanistic phenotypes in mice, cattle and pigs, suggesting critical roles for these residues as well (Figure 4-7). Under an alternative, equally parsimonious alignment scheme, conserved residues 93L and 94E would have been deleted in this jaguarundi allele (instead of 101L, 102E; see Figure 4-4). In this case the deletion would include three residues where melanism-implicated substitutions have been previously identified in other species (Figure 4-7). Given the functional studies performed in other mammals (Robbins et al. 1993; Våge et al. 1999), it is likely that these deleted residues are important to maintain an inactive conformation of the MC1R protein, and/or they are critical for binding of the antagonist peptide ASIP.

The three distinctive melanistic deletions identified here appear to be species-specific, as no other surveyed felid was found to carry any of them (Figures 4-3 and 4-5). The absence of these variants in melanistic individuals from five other felid species (leopard [*Panthera pardus*], Geoffroy's cat [*Oncifelis geoffroyi*], oncilla [*Leopardus tigrinus*], pampas cat [*Oncifelis colocolo*] and Asian golden cat [*Catopuma temmincki*]) would suggest that melanism arose independently at least four times in the family Felidae, in many cases rising to high population-level frequencies (Cott 1940; Searle 1968; Kitchener 1991; Robinson 1976; Nowell & Jackson 1996). The significant elevation of independent gene variants in parallel Felidae lineages raises the prospect of an adaptive advantage of melanistic mutants under certain

ecological circumstances (Cott 1940; Kitchener 1991). A striking example is the jaguarundi, whose 'wild type' dark coloration is here shown to be a derived condition, having replaced the ancestral reddish form throughout its continental range, in an expansion process that was likely mediated by natural selection.

Conclusions

The results presented here indicate that loss of function of the *ASIP* gene causes black coloration in the domestic cat, while two independent deletions in the *MC1R* gene are implicated in gain-of-function melanistic phenotypes in the jaguar and jaguarundi. These findings implicate a minimum of four separate origins of melanism in the Felidae, followed by independent increases in frequency in each lineage, as observed in the jaguarundi. The observation that melanism has not achieved fixation in any cat species (Robinson 1976; Kitchener 1991; Nowell & Jackson 1996) suggests the possibility that advantageous effects in some areas may be balanced by negative selection in others, or that processes such as frequency-dependent selection may be involved in the maintenance of polymorphism in these traits. In a broader context, the prospect of directly inspecting gene variants that specify phenotypic variation potentially subject to natural selection will allow the direct study of such traits in free-ranging populations of endangered and/or elusive species, including the use of non-invasive sampling. These and other applications of such integrated genetic approaches will hopefully enhance our understanding of species survival, diversification and adaptive evolution over space and time.

Table 4-1. STR loci from domestic cat BAC clone188e3 containing the *ASIP* gene.

Locus	Repeat [a]	Observed alleles [b]	Forward primer (5' – 3')	Reverse primer (5' – 3')
FCA701	$(AC)_{18}$	146, 150, 152, 154	GGCAACCTACTGAATGAAAGAA	CCAGCCAGGCATTCAATC
FCA702	$(GA)_{15}$	276, 278	CTCCTCAGCTTGGAGGACAG	CTTGATTTCAGCTCAGGTCATG
FCA703	$(GA)_{18}$	229**	AGCTCTCTGGGCTTTAAAGTTC	CATCTGGCTCTGTCCTGACA
FCA704 *	$(GA)_{12}$	132, 136, 138	ATGGAATTTAGAGGTGGTGAGG	CTGTCCCCATTTCTCTCTGC
FCA705	$(CTTT)_{16}$	0**	CGGATGCTTAACTGACTGAGC	ACTGAGCCCCCACCTAGG
FCA706	$(AG)_{12}$	150	TGTAGCTTCTTTGGTGAACTGG	GTTTGTGAGTTTGAGCCTTGC
FCA707 *	$(CT)_{12}$	132, 138, 140	CAGCTGGGAGACTGGAGC	TTGGCCATACACTCTTTAACCA
FCA708 *	$(AC)_{22}$	117, 125, 127, 129, 139, 141, 145, 147, 149	CAGGAAACTTTTATCCCAATGG	CACACTTCAGGGGTTAGGGA
FCA709	$(AAC)_{11}$	0**	TACAGTCTGCGCTGCAGC	ACTTAAAATTTAAGGAGGCGGG
FCA710	$(AG)_{15}$	110, 120	AGTGTGACTGGGAGAGGGG	TGGTTTGTGAGCTTGAGCC
FCA711	$(GA)_{16}$	132, 158, 160	CCCTGTTGGACAGCAAAATT	AGCATGGAACCTGCTTCG
FCA712	$(TC)_{17}$	188, 190	AGGTCATGATCTCACCGTTTG	CACTGTGATTCTCTGAAGCAGG
FCA713	$(AG)_{21}$	222	TGCAATCCCAATTAAAAATTCC	CAGGCTCTGTCCTGAACACA
FCA714	$(AT)_{22}$	250	CTCTACCAACTGAGCCAGCC	CCTCAGATTTAGAGCATTTGGC
FCA715 *	$(GA)_{10}$	128, 132, 138, 140	TGAATGCTGAGGAGATCTTGG	TTCTCTCTTTCTGCCTCTCCC
FCA716 *	$(TG)_{20}$	208, 212, 218, 220	TCCCACAGGCTAAGAACATACA	AATCAACTGAGCCACCAAGG
FCA717 *	$(AC)_{10}$	208, 210, 212, 218, 224	GCATCCAAATTGGAAAGGAA	GCCACTCAGGCACTCCTTAG

Table 4-1. Continued.

Locus	Repeat [a]	Observed alleles [b]	Forward primer (5' – 3')	Reverse primer (5' – 3')
FCA718 *	$(AAAG)_{19}$	227, 235, 239, 243, 247, 251, 255	TGACAGCTCAGAGCCTAAAGC	GAGTGCACCCCTCCCATAC
FCA719 *	$(AAAG)_{15}$	186, 190, 194, 198, 202, 206, 210	CAACATGCCTTTACTAAGCGC	GTTTGGGGATTGGTTTCCTT
FCA720	$(TC)_{14}$	316	ATAAGGGTGCCTAGGTGGCT	TAAATCTGGCAACCCTACTTGG

* Loci confirmed to be polymorphic and selected for additional analysis in the Purina Pedigree.

** Loci presenting inconsistent amplification or no PCR product in the expected size range (0**).

[a] Number of repeats observed in BAC clone RPCI86-188e3.

[b] Shown as allele size in bp. Observed alleles were obtained from two test assays (pooled DNA from ten random cats, and separate reactions for six founders of the Purina Pedigree) for all loci, as well as large-scale results from the Purina Pedigree (for loci FCA708, FCA718 and FCA719, used in the Linkage analysis).

Table 4-2. Primers used in this study.

Primer [a]	Sequence (5' – 3')
***ASIP* gene**	
AgoEx2-F1	CAGAGCACCAGCCCAAAG
AgoEx2-R1	CCTTCTCATGGACCGGAGT
AgoEx3-F1 *	TCCACTCCTCCCACTTTACTG
AgoEx3-R1 *	CCCTTAGCTCTCTGGGCTTC
AgoEx4-F1	CCAAGGCTGAGGTGAAGAGG
AgoEx4-R1	AAGCCCCGCCTCCTGAAA
AgoFcaDel-F1	CCTACAGTCACCTGGCACCT
AgoFcaDel-R1	CCTTCTCATGGACCGGAGT
***MC1R* gene**	
ExtCOM-F1 **	ATGAAGCCTGCTGGAAGCAC
ExtCOM-R1 **	GATATCCCCACCTCCCTCTG
ExtDel-F1	CTGCACTCGCCCATGTATTA
ExtDel-R1	CCACAGACCAGCACGTCAAT

[a] For all primers, a suffix 'F' refers to forward, and 'R' to reverse.

* Primers also utilized in Chapter 2

** Primers also utilized in Chapter 3

Figure Legends

Figure 4-1. Nucleotide sequence of the three coding exons of the domestic cat *ASIP* gene, shown for a wild type (Fca-WT) and a melanistic allele (Fca-Black); amino acid sequences are given above (Fca-WT) and below (Fca-Black) the third position of each codon. Dots indicate identity to the top sequence. In exon 2 the initiation codon is underlined, and the two (bold underlined) nucleotides deleted in the melanism allele *ASIP*-Δ2 (indicated by asterisks) are boxed. In exon 4 the ten boxed cystein residues are conserved across mammals and critical for agouti function. Termination codons are underlined and marked with an X. See Chapter 2 for more details on *ASIP* structure and sequence.

Figure 4-2. Domestic cat nuclear family (part of the pedigree analyzed here) showing the co-segregation of the *ASIP*-Δ2 deletion allele and melanism, with a recessive mode of inheritance. Melanistic animals are indicated by dark symbols. + indicates an intact (wild type) allele; Δ2 (Δ in offspring) is the *ASIP*-Δ2 mutant allele. Perfect co-segregation was also observed in the remainder of the pedigree (total of 89 individuals comprising 109 melanism meioses; see Appendix 3 for full list of deletion genotypes).

Figure 4-3. Genotyping results for the *ASIP*-Δ2 deletion allele identified in domestic cats. Tree on the left indicates phylogenetic relationships of species with melanistic forms included in this study (Johnson & O'Brien 1997; Johnson et al. 1999); thick branches on the tree indicate major lineages in the Felidae (see Chapter 1). Only

unrelated individuals from each species were included. *Number of melanistic individuals assayed is given in parentheses. **Includes black domestic cats from Botswana (n=1), Brazil (n=10), Israel (n=5), Mongolia (n=4) and USA (n=37); USA black cats include random-bred individuals as well as representatives of the following breeds: Bombay, Maine Coon, Norwegian Forest Cat, Cornish Rex, Turkish Van, and Sphynx. *** Includes one or two individuals from each of the following cat species: *Felis nigripes, Felis margarita, Panthera leo, Panthera tigris, Panthera uncia, Neofelis nebulosa, Leopardus pardalis, Puma concolor, Acinonyx jubatus, Prionailurus bengalensis, Caracal caracal, Profelis aurata* and *Lynx rufus*. A full list of all individuals and their deletion genotypes is given in Appendix 4. The column f [Δ2] represents the frequency of the *ASIP*-Δ2 allele in each species; the domestic cat frequency was calculated exclusively from the non-melanistic genotype frequencies, assuming Hardy-Weinberg equilibrium.

Figure 4-4. Amino acid sequence alignment of the *MC1R* gene from a domestic cat (D.cat), wild-type (non-melanistic) jaguar allele (Pon-WT), melanistic jaguar allele *MC1R*-Δ15 (Pon-Mel), dark-brown (melanistic) jaguarundi allele *MC1R*-Δ24 (Hya-Dark), reddish jaguarundi (Hya-Red), human (GenBank accession number AF326275) and mouse (accession X65635). Dots indicate identity to top sequence; codon positions are shown above sequence, at the end of each block. The jaguar and jaguarundi melanistic deletions are shaded (dashes indicate deleted sites), and the region containing these deletions is boxed. The conserved leucine codon 3' of the jaguar deletion is marked with an asterisk, with the amino acid change in the *MC1R*-

Δ15 allele in bold. See Chapter 3 for additional comparisons with other *MC1R* homologues.

Figure 4-5. Genotyping results for the gain-of-function deletions identified in the *MC1R* gene of jaguars and jaguarundis. Only unrelated animals from each species were included. Tree on the left indicates phylogenetic relationships of species with melanistic forms included in this study (Johnson & O'Brien 1997; Johnson et al. 1999); thick branches on the tree indicate major lineages in the Felidae (see Chapter 1). Alleles are coded as follows: Δ15 is the jaguar deletion allele *MC1R*-Δ15; Δ24 is the jaguarundi deletion allele *MC1R*-Δ24; + indicates an ancestral-type (non-deleted) allele. *Number of melanistic individuals assayed is given in parentheses. **Includes one or two individuals from each of the following cat species: *Felis silvestris, Felis nigripes, Felis margarita, Panthera leo, Panthera tigris, Panthera uncia, Neofelis nebulosa, Leopardus pardalis, Puma concolor, Acinonyx jubatus, Prionailurus bengalensis, Prionailurus planiceps, Caracal caracal, Profelis aurata, Lynx rufus, Catopuma badia* and *Leptailurus serval.* A full list of all individuals and their deletion genotypes is given in Appendix 3. The columns f [Δ15] and f [Δ24] are the frequencies of the *MC1R* -Δ15 and *MC1R* -Δ24 alleles, respectively, calculated from the available sample for each species.

Figure 4-6. Pedigree of captive jaguars showing co-segregation of melanism and the *MC1R*-Δ15 deletion allele, following a dominant mode of inheritance. Melanistic animals are indicated by dark symbols. Alleles are coded as in Figure 4-5.

Figure 4-7. Partial diagram of the MC1R protein in (A) jaguar (*Panthera onca*) *MC1R* -Δ15 allele and (B) jaguarundi (*Herpailurus yaguarondi*) *MC1R*-Δ24 allele, focusing on the region where deletions were identified in these variants. The inset in the top left corner shows a schematic view of the whole protein in the melanocyte membrane, with the enlarged segment defined by an ellipse. The detailed views in A and B start in the first intra-cellular loop and end at the beginning of the third transmembrane domain (residues 63 to 125 in the cat); spatial arrangement adapted from Robbins et al. [1993], with site numbers based on the cat sequence). Circles bordered by a thick black line represent amino acid residues conserved between the mouse (Robbins et al. 1993) and each of the deletion alleles shown; those with a thin blue border are different (see Figure 4-4 for sequence). Amino acids shaded in brown (residue and site number in cats are indicated in A) represent all those at which substitutions have been previously reported to cause gain-of-function melanistic phenotypes in other species: S71L (E^{tob}), E94K ($E^{so\text{-}3J}$), L100P (E^{so}) in the mouse (Robbins et al. 1993) (E94K is also found in melanistic chickens and bananaquits [Takeuchi et al. 1996; Theron et al. 2001]); L99P in cattle (Klungland et al. 1995); C125R in the red fox (Våge et al. 1997); L99P and D121N in the pig (Kijas et al. 1998); and M73K and D121N in sheep (Våge et al. 1999). The replacement S90G (shaded in orange) has been proposed to be potentially associated with melanism in the domestic dog (Newton et al. 2000). The hatched residues (encompassed by a dashed line) are those deleted in melanistic jaguars (A) and jaguarundis (B). In (A) the two leucine residues bordering the deletion (one of which has been deleted) are indicated in red letters, with the one at position 106 being changed to a threonine, as

observed in the *MC1R* -Δ15 allele (Figure 4-4). Arrows in (B) indicate two amino acids that are included in the jaguarundi *MC1R* -Δ24 deletion under an alternative alignment scheme (see Figure 4-4 and text).

Exon2: 170 bp

```
                          M  N  I  L  R  L  L  L  A  T  L  L  V  C  L  C [16]
Cat-WT      GCCTTCAAGG ATG AAT ATC CTC CGC CTA CTC CTG GCC ACC CTG CTG GTC TGC CTG TGC
Cat-Black   .......... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
                       .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  . [16]

              L  L  T  A  Y  S  H  L  A  P  E  E  K  P  R  D  D  R  N [35]
Cat-WT      CTC CTC ACT GCC TAC AGT CAC CTG GCACCT GAG GAA AAA CCC AGA GAT GAC AGG AAC
Cat-Black   ... ... ... ... ... ... ... ... ..... ... ... ... ... ... ... ... ... ...
              .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  . [35]

              L  R  S  N  S  S  M  N  M  L  D  L  S  S  V  S  I  V [53]
Cat-WT      CTG AGG AGC AAC TCC TCC ATG AAC ATG TTG GAT CTC TCT TCT GTC TCT ATT GTA G
Cat-Black   ... ... ... ... ... .. .. ... ... ... ... ... ... ... ... ... ... .
              .  .  .  .  .  .  S  E  H  V  G  S  L  F  C  L  Y  C [52]
```

Exon3: 62 bp

```
              A  L  N  K  K  S  K  K  I  S  R  K  E  A  E  K  K  R  S  S  K [74]
Cat-WT      CG CTG AAC AAG AAA TCC AAA AAG ATC AGC AGA AAA GAG GCG GAA AAG AAG AGA TCT TCC AAG
Cat-Black   .. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
              S  A  E  Q  E  I  Q  K  D  Q  Q  K  R  G  G  K  E  E  I  F  Q [73]
```

Exon4: 186 bp

```
              K  K  A  S  M  K  N  V  A  Q  P  R  R  P  R  P  P  P  P  A  P  C  V  A  T  R  D  S [102]
Cat-WT      AAA AAG GCT TCG ATG AAG AAT GTT GCT CAG CCT CGG CGG CCC CGG CCT CCG CCG CCC GCC CCC TGC GTG GCC ACT CGT GAC AGC
Cat-Black   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
              E  K  G  F  D  E  E  C  C  S  A  S  A  A  P  A  S  A  A  R  P  L  R  G  H  S  X [99]

              C  K  P  P  A  P  A  C  C  D  P  C  A  S  C  Q  C  R  F  F  R  S  S  C  S  C  R  V [130]
Cat-WT      TGC AAG CCG CCG GCG CCC GCC TGC TGC GAC CCG TGC GCC TCC TGC CAG TGC CGC TTC TTC CGC AGC TCC TGC TCC TGC CGA GTG
Cat-Black   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

              L  N  P  T  C  X [135]
Cat-WT      CTC AAC CCC ACC TGC TGA
Cat-Black   ... ... ... ... ... ...
```

Figure 4-1

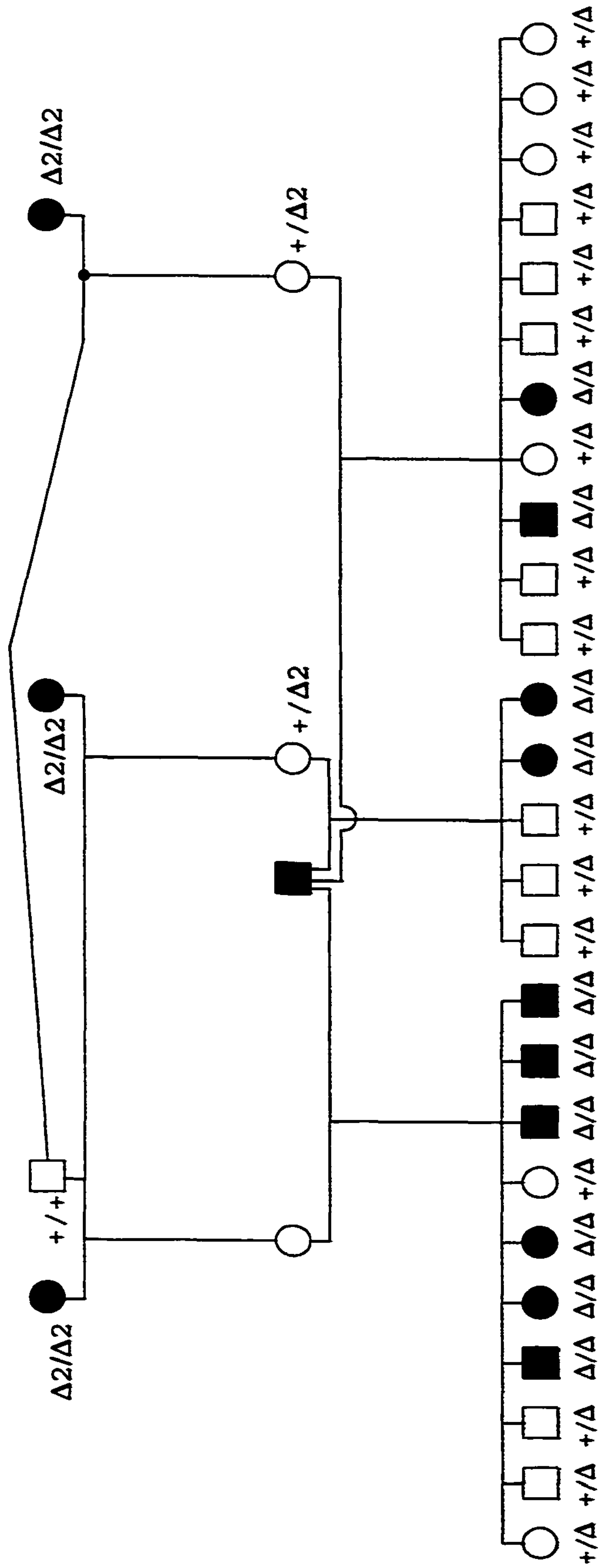
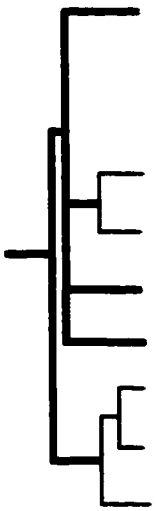


Figure 4-2



Felidae Species	n (mel)*	*ASIP* Genotypes			
		Δ2/Δ2	Δ2/+	+/+	f [Δ2]
Felis catus					**0.41**
Melanistic**	57 (57)	57	-	-	
Non-melanistic	26 (0)	-	15	11	
Panthera onca	10 (4)	-	-	10	0.00
Panthera pardus	4 (4)	-	-	4	0.00
Herpailurus yaguarondi	15 (8)	-	-	15	0.00
Catopuma temmincki	2 (1)	-	-	2	0.00
Leopardus tigrinus	3 (1)	-	-	3	0.00
Oncifelis geoffroyi	4 (2)	-	-	4	0.00
Oncifelis colocolo	3 (1)	-	-	3	0.00
Other Felidae***	15 (0)	-	-	15	0.00
Σ	139 (78)	57	15	67	

Figure 4-3

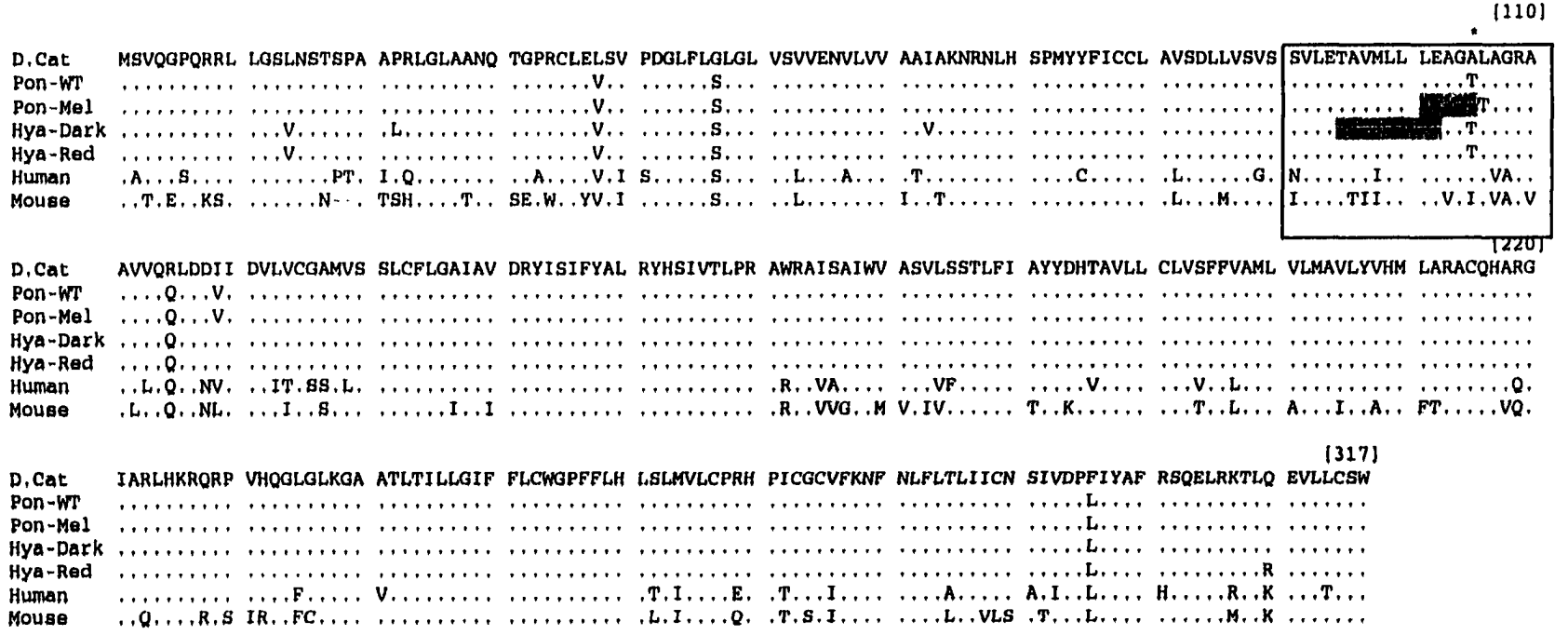


Figure 4-4

Felidae Species	n (mel)*	*MC1R* Genotypes						
		Δ15/Δ15	Δ15/+	Δ24/Δ24	Δ24/+	+/+	f [Δ15]	f [Δ24]
Felis catus	43 (28)	-	-	-	-	43	0.00	0.00
Panthera onca							**0.12**	0.00
Non-melanistic	36 (0)	-	-	-	-	36		
Melanistic	10 (10)	1	9	-	-	-		
Panthera pardus	8 (4)	-	-	-	-	8		
Herpailurus yaguarondi							0.00	**0.50**
Very dark	7 (7)	-	-	6	1	-		
Dark brown / gray	12 (12)	-	-	3	9	-		
Red / reddish	10 (0)	-	-	-	4	6		
Unknown color	11	-	-	3	2	6		
Catopuma temmincki	2 (1)	-	-	-	-	2	0.00	0.00
Leopardus tigrinus	3 (1)	-	-	-	-	3	0.00	0.00
Oncifelis geoffroyi	4 (2)	-	-	-	-	4	0.00	0.00
Oncifelis colocolo	4 (1)	-	-	-	-	4	0.00	0.00
Other Felidae**	22 (0)	-	-	-	-	22	0.00	0.00
Σ	172 (66)	1	9	12	16	134		

Figure 4-5

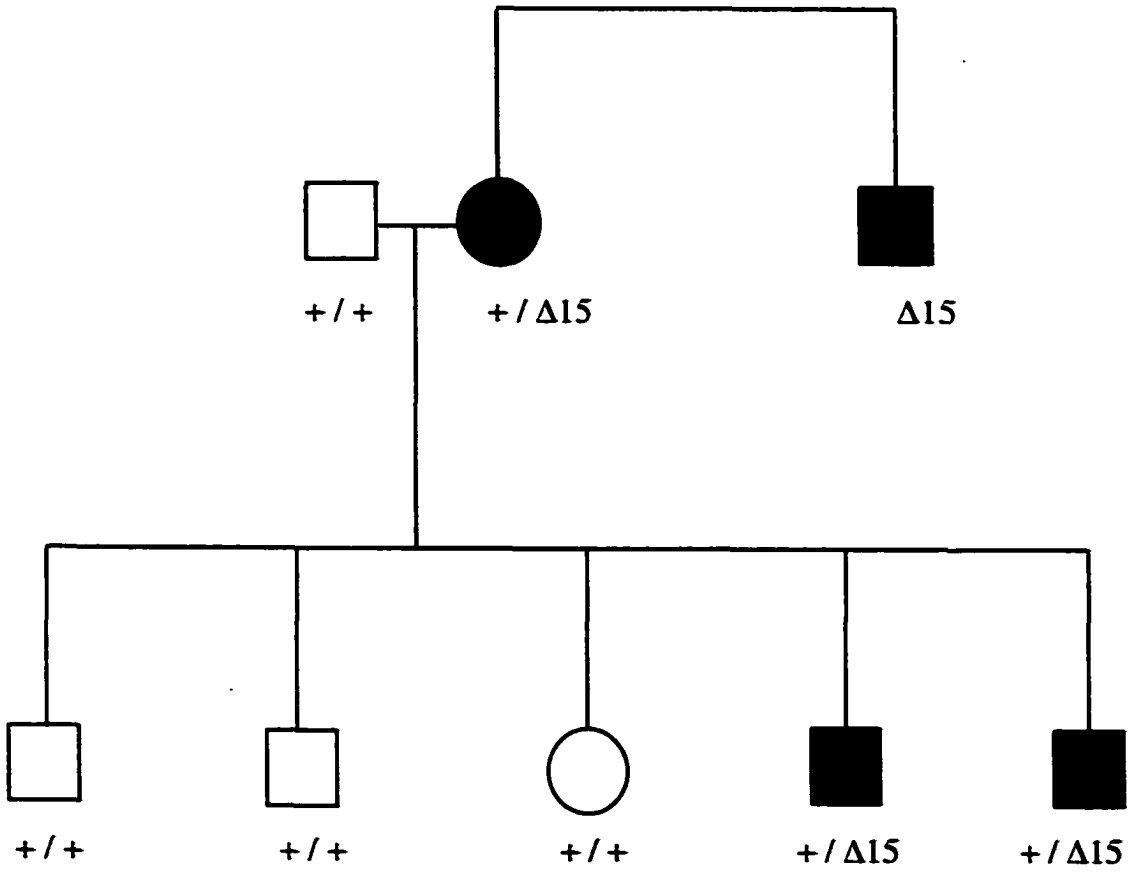


Figure 4-6

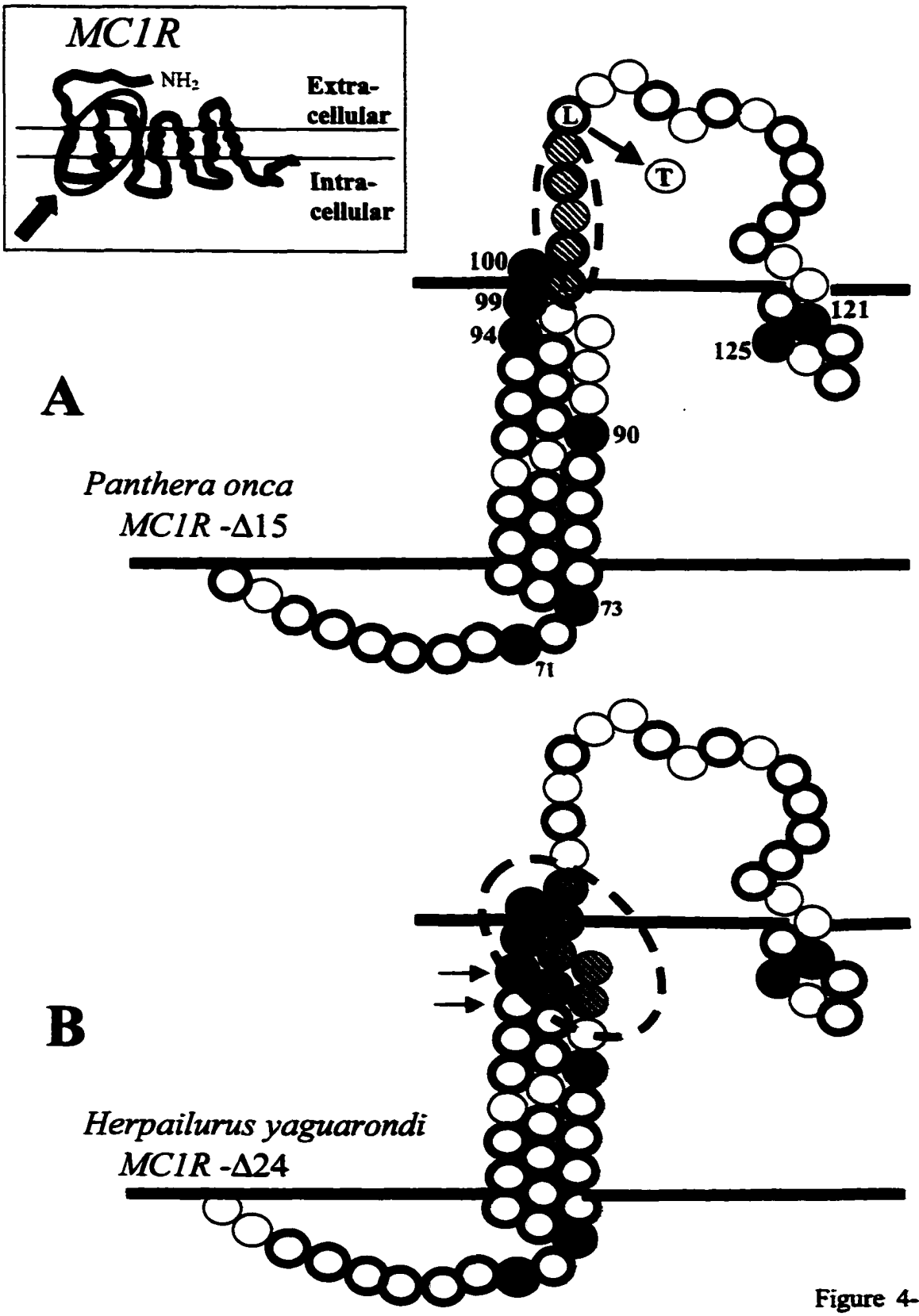


Figure 4-7

CHAPTER 5

General Discussion

Evolutionary and functional inferences on the *ASIP* and *MC1R* genes

The present study included comparative analyses of rates and patterns of molecular evolution of the *ASIP* and *MC1R* genes among sequences representing different mammalian groups (and other vertebrates, in the case of *MC1R*). Overall, *MC1R* exhibits higher substitution rates than *ASIP*, with the difference appearing less dramatic for nonsynonymous than synonymous changes (Tables 2-4 and 3-5). The higher synonymous rate observed at *MC1R versus* the *ASIP / AHCY* region suggests that the former may be located in a region subjected to higher mutation frequency (or lower levels of background selection), although no strict comparison was attempted in view of the considerable variance associated with sampling few loci. Nevertheless, this observation may be helpful to interpret the comparison of nonsynonymous rates between *ASIP* and *MC1R*, implying that part of the detected difference may be due to higher mutation rates in the latter. The overall pattern suggests that both genes evolve under a similar level of functional constraint, when their coding region is considered as a whole. Both *ASIP* and *MC1R* seem to accumulate nonsynonymous changes faster than the average observed in broader genomic surveys (Li 1997; Makalowski & Boguski 1998), indicating that the level of evolutionary constraints on these genes is lower than that observed at many other loci, and/or that positive selection may be responsible for accelerated evolution in particular segments of these two genes.

In this context, it does not seem justified to consider these genes to be 'highly conserved', as argued in the literature (e.g. Lu et al. 1998; Klungland et al. 1999; Miltenberger et al. 2002). The magnitude and specific details of molecular differences observed among species may be relevant when attempting to make inferences on functional properties of identified variants, as experimental approaches are often limited to heterologous model systems (e.g. transgenic mice or human and mouse cell lines; see Lu et al. 1998). When analyzing the function of molecular variants identified in other species by introducing them into these experimental systems, it is important to bear in mind the potential role of the remaining differences from the original protein in modifying the impact of that particular mutation. Even though technical difficulties may continue to restrict the ability to test identified variants in their original molecular and cellular environment, potential controls for this possibility (e.g. introducing multiple changes separately and combined) may be incorporated into such studies.

A complementary approach is to combine functional assays with the evolutionary analysis of large comparative data sets including multiple taxa. These analyses may indicate which residues are more or less constrained over evolutionary time, which pairs or groups of sites may show correlated patterns of substitution, and which types of change seem to be 'allowed' or recurrent at particular locations. Such an analysis was attempted in this study for both *ASIP* and *MC1R* (Chapters 2 and 3), indicating that both genes exhibit similarly heterogeneous constraints, in which highly variable segments are interspersed with conserved motifs. Evidence of positive selection generating diversity at the variable segments was not detected in this study,

but remains a possibility. Suggestive evidence of functional divergence among different evolutionary lineages was observed for *MC1R*, in which patterns of both regulatory and coding region (nonsynonymous) variation were altered in a taxon-specific manner (see Figures 3-4 and 3-6). In contrast, evidence for conservation of function among taxa was obtained in comparisons of evolutionary and functional data for *ASIP*, as all sites found to be mutated in mouse null alleles (and most of those mutated in hypomorphic alleles) are completely conserved among taxa (Figure 2-6). This observation should be re-evaluated with expanded taxonomic sampling for *ASIP* sequences, and may help in selecting additional sites for future experimental assays.

Phylogenetic relationships in the Felidae

The evolutionary relationships among the 37 species contained in the Felidae have been a matter of extensive study and controversy for over a century (e.g. Pocock 1917; Collier & O'Brien 1985; Wozencraft 1989, 1993; Salles 1992; Bininda-Emonds et al. 2001). In spite of the amount of effort invested on this topic, conflicting results from diverse lines of evidence have continually precluded a consensual understanding of the phylogeny and appropriate taxonomy for this group. Recent studies using DNA sequence data from mitochondrial genes and nuclear introns located on the X and Y chromosomes indicated the existence of eight major lineages in the Felidae (Johnson & O'Brien 1997; Pecon-Slattery & O'Brien 1998; see Figure 1-1), which were supported by a subsequent study incorporating molecular, morphological and karyological data (Mattern & McLennan 2000). The basal relationships among these lineages have still not been resolved conclusively, and the placement of four additional

species (Pallas cat [*Otocolobus manul*], marbled cat [*Pardofelis marmorata*], serval [*Leptailurus serval*] and rusty-spotted cat [*Prionailurus rubiginosus*]) has been unstable in these recent studies. Analyses of additional phylogenetic characters are needed to conclusively resolve the pattern of this radiation, which would be required not only to establish a consistent taxonomy for this group, but also to serve as a basis for other evolutionary studies, including investigations on the origins and adaptive significance of coat color variants (e.g. Ortolani & Caro 1996).

One of the goals of the present study was to evaluate the extent of molecular diversity in portions of the *ASIP* and *MC1R* genes in the Felidae, and to assess their utility in reconstructing the phylogenetic relationships among cat species. The results shown in Chapters 2 and 3 reveal considerable levels of variation in both of the analyzed segments (*ASIP* intron 2 and exons 2 and 3 [Chapter 2], and the whole *MC1R* coding region [Chapter 3]), but contrasting levels of phylogenetic resolution between them. Considerable support for resolved phylogenetic relationships was observed with the *ASIP* segment, whereas only two nodes of the tree were consistently recovered with *MC1R*. The difference is not surprising, since the *ASIP* segment is longer and mostly intronic, showing two-fold more variation among felids than is observed in the *MC1R* coding region.

Overall, the phylogenetic analyses performed with both genes provided novel information on the relationships among felid lineages, and also on the placement of three of the four unaligned species. These results indicate that (i) the Pallas cat is closely related to the Domestic Cat, Leopard Cat and Lynx Lineages (Chapter 2), suggesting an Eurasian radiation of this group; (ii) the marbled cat is related to the Bay

Cat Lineage (Chapter 2), constituting a southeast Asian clade; and (iii) the serval is related to the Caracal Lineage (Chapter 3), forming an African group of felids. If confirmed by further analyses, these results will add to the understanding of the biogeographic history of extant felid radiations. Some of the observed relationships among the main clades constitute novel hypotheses on felid phylogeny (e.g. the proximity of the Puma and Ocelot Lineages, and between the Domestic Cat and Caracal Lineages), which can be tested using additional information for these and other character sets. Further phylogenetic studies using these genes should focus on incorporating additional taxonomic representation for the *ASIP* segment, especially regarding the generation of complete sequences from suitable carnivore outgroups (e.g. viverrids), allowing the reliable placement of the felid root and confident interpretation of the observed clades.

Evolution of melanism in the Felidae

Three different deletions implicated in felid melanism have been described in this study: one in the domestic cat *ASIP* and two in the *MC1R* of jaguars and jaguarundis (Chapter 4). The functional interpretation of the *ASIP* mutation seems straightforward, as different deletions leading to frameshifts have been implicated in complete loss of *ASIP* function in mice (Miltenberger et al. 2002), and inferred in other species (Kuramoto et al. 2001a; Rieder et al. 2001). This result has been observed in frameshift mutations occurring near the C-terminus of ASIP (Miltenberger et al. 2002), implying even stronger impacts of a deletion occurring on the N-terminal half of the protein, such as observed in the domestic cat.

The location of the *MC1R* deletions observed in jaguars and jaguarundis suggests that the included residues are critical for maintaining an inactive conformation of the receptor, and/or interact with the ASIP antagonist to mediate inhibition (see Chapter 4). Adjacent to the five amino acids deleted in the jaguar *MC1R*-Δ15 allele, two nonsynonymous nucleotide substitutions changed a leucine codon (conserved across all surveyed vertebrates [Figure 3-5]) to a threonine (see Figure 4-4). Given the assumption that a single deletion event occurred, two equally parsimonious (identical) scenarios can be considered for the source of this mutated codon, namely the leucine at position 101, or that at position 106 (Figure 4-4). Here the second scenario (codon 106) is favored, as its location fits a plausible mechanism for the origin of the deletion through slippage or unequal cross-over mediated by the CTG leucine codons (see Figure 4-4; note that the jaguarundi *MC1R*-Δ24 deletion may also have been generated by a similar process, mediated by the 'CTG,GAG' codons that border its deleted segment). In either case a conserved leucine (a non-polar residue) is changed to a threonine (uncharged polar), requiring two different non-synonymous steps. This may be interpreted as suggestive evidence for positive selection acting after the deletion event to change the remaining leucine residue (which might pull the shortened extra-cellular loop towards the membrane, possibly leading to some functional impairment; see Figure 4-7) to a polar amino acid , perhaps conferring additional flexibility to this domain. This hypothesis can be tested directly with functional assays addressing the biological properties of these sequence variants, comparing the activity of alleles bearing the deletion in combination with leucine or threonine at that position.

From a broader perspective, the results presented here indicate that melanism arose independently at least four times in the Felidae (Chapter 4). Given the current knowledge on the felid phylogeny (Chapter 1 and section above) and the information available on the inheritance of melanism in additional cat species (Chapter 1), it is likely that this phenotype has arisen at least nine times in the family, and it is conceivable that it may be species-specific in all cases. A potential exception may be the Ocelot Lineage, in which four species forming a sub-clade (*Oncifelis geoffroyi*, *O guigna*, *O. colocolo* and *Leopardus tigrinus*) exhibit melanism at appreciable frequencies (Nowell & Jackson 1996; Dunstone et al 1998; L. Silveira, personal communication; E.E., personal observation), while the three remaining species have never been confirmed to harbor this polymorphism (see Figure 1-1). It is intriguing to speculate that melanism may have arisen in the ancestor of these four species, and been maintained as a trans-species polymorphism through three speciation events. In both the trans-species and recurrent-origins scenarios, an important role for natural selection is implied, favoring melanism in some contexts but precluding its complete fixation. Additional studies including molecular, population-genetic, ecological and behavioral components should provide a more detailed understanding of the evolutionary history and dynamics of this phenotype within and among different felid species (see section on "Prospects for Future Work") .

Molecular genetics of polymorphic phenotypes

The results obtained in this study, along with other recent reports (e.g. Ritland et al. 2001; Theron et al. 2001), illustrate the potential of direct molecular

investigations of intra-specific phenotypic diversity to provide information on the functional basis and evolutionary significance of such phenomena. As more genomic and functional data accumulate from traditional model systems and also from other species, direct studies of variation in candidate genes in wild populations will become progressively more feasible. In addition to shedding light on the molecular mechanisms underlying observed phenotypes, these investigations can potentially uncover the historical origins and evolutionary dynamics of these traits, and provide genetic tools for the analysis of their adaptive significance in natural populations. These include not only inferences on historical adaptive processes (e.g. using assessment of nonsynonymous *versus* synonymous changes, and diversity at closely linked markers), but also ecological investigations on the frequencies and spatial dynamics of observed variants using non-invasive sampling and molecular markers derived from genetic studies.

The incorporation of direct molecular evolutionary analyses in the study of polymorphic phenotypes can add new dimensions to the way in which these traits are described and interpreted. One example is the ability to infer the evolutionary polarity of specific changes in phenotypes (i.e. identifying which form is ancestral, independent of current frequency) by analyzing the phylogenetic pattern of molecular changes associated with the trait of interest. A consequence of this approach is that the most common allele (classically referred to as the 'wild type' [Fisher 1930; Silvers 1979; Hartl & Clark 1997; Li 1997]) may be found not to be ancestral, thus allowing a direct investigation of ongoing processes of allele replacement and/or maintenance of polymorphism in natural populations. Such a phenomenon was observed in the

jaguarundi in this study (Chapter 4), where the most frequent color morph was inferred to be a derived condition.

This observation illustrates the temporally restricted nature of any definition of 'wild type', which is usually understood in theory (e.g. Li 1997) but so far rarely documented in natural populations. Most cases where allelic replacement processes have been documented in nature consist of very recent and rapid episodes witnessed directly by human observers (e.g. melanism in insects [Majerus 1998] and cases of pesticide resistance [May 1985; Raymond et al. 2001]), whose molecular basis is either unknown or was identified *a posteriori*. The direct molecular study of polymorphic phenotypes allows the inference of such processes over a much broader window of evolutionary time, potentially revealing patterns of allelic replacement historically associated with adaptation to varying environments.

Prospects for future work

The inferences and conclusions derived from this study suggest several avenues for further investigation of the evolution of melanism in the Felidae, as well as similar endeavors addressing other phenotypes. An important development expected for the future is the use of functional assays to investigate the biological consequences of the identified *ASIP* and *MC1R* mutations. This would contribute to verify experimentally the implication of these variants in melanistic phenotypes, and also shed light on the functional mechanisms modified by these mutations.

Another intended goal is the detailed characterization of *ASIP* and *MC1R* alleles present in melanistic individuals of additional felid species, aiming to verify the

involvement of these genes in other melanism-inducing mutations in this group. The prospect of identifying additional cases of *ASIP* / *MC1R* involvement in polymorphic pigmentation phenotypes is promising, given the results obtained so far in felids and other groups. Considering the existence of several additional loci causing melanistic phenotypes in mice (Silvers 1979), it may seem surprising that mutations in *ASIP* and *MC1R* have so often been implicated in melanism and other pigmentation variants. A possible explanation is that disruption/alteration of function at the remaining loci might produce more significant pleiotropic effects than changes *ASIP* and *MC1R*, thus leading to more frequent selective removal of variants from natural (or domesticated) populations. The potential for such pleiotropic effects has been documented for the attractin/mahogany locus (e.g. Gunn et al. 2001; He et al. 1991; Kuramoto et al. 2001b), and remains to be investigated as additional melanism loci are identified and characterized.

Future projects on felid melanism may also focus on population-level investigations on the evolutionary history and dynamics of molecular variants associated with this phenotype. The intra-specific molecular variants identified here (SNPs and STR loci in or near the *ASIP* gene, and coding region polymorphisms in *MC1R*) may be used to perform DNA-based analyses of population history, attempting to differentiate demographic *versus* selective processes affecting these genes over evolutionary time. Several analytical methods have been designed for this purpose (e.g. Hudson et al. 1987; Tajima 1989; Fu & Li 1993; Fu 1996; McDonald & Kreitman 1991; reviewed by Li [1997] and Kreitman [2000]), and some approaches

that seem to be most likely applicable to the variants identified here will be described briefly below.

The 'signature' of natural selection at the DNA level can be investigated using several population-genetic approaches that test for departures from expectations under a neutral model of molecular evolution (Kimura 1983). Under neutral evolution it is expected that synonymous and nonsynonymous changes will occur at the same frequency (per synonymous and nonsynonymous site, respectively), that intra-specific polymorphism and inter-specific divergence (fixed differences) will be both linearly correlated with the neutral mutation rate (Kimura 1983; Kreitman 2000), and that different genomic regions will exhibit similar patterns of nucleotide polymorphism (conditioned on the stochastic nature of the mutational process) (Kreitman 2000). These expectations of the neutral-equilibrium model can be statistically compared to molecular data obtained from natural populations using tests designed to identify different forms of departure from neutrality (reviewed by Kreitman [2000]). One of the most widely used approaches, likely applicable to the problem described in this study, is Tajima's D test (Tajima 1989). This test compares two estimates of the neutral parameter θ ($4N_e\mu$ for nuclear autosomal genes, where N_e is the effective population size and μ is the mutation rate per sequence per generation), $\theta(K)$ and $\theta(\pi)$. The former is based on the number of segregating (variable) sites in a sample of DNA sequences, and is not influenced by the frequency of different variants. In contrast, $\theta(\pi)$ represents the mean number of pairwise nucleotide differences in the sample (Tajima 1983), and is strongly influenced by the frequency of variants in the sample. Under neutrality, these two estimates should be identical (or not differ significantly

from each other). The statistical comparison of these two estimates (Tajima 1989) can indicate departures from neutral expectations, in which significantly positive values of D (larger $\theta[\pi]$) suggest the occurrence of balancing selection, and significantly negative values (larger $\theta[K]$) suggest the occurrence of directional selection (a recent selective sweep) (Kreitman 2000; Verrelli et al. 2002). It is important to note, for Tajima's and many other tests of neutrality, that natural selection is not the only force that can lead to departure from neutral expectations, i.e. similar skews in the polymorphism spectrum can be caused by changes in patterns of genetic drift caused by demographic processes (Fu 1996). In the case of Tajima's test, significantly positive values may also be due to population subdivision, and significantly negative values may be due to a recent expansion in population size (Kreitman 2000). A logical way to differentiate these two types of inference is to analyze multiple genomic loci, since demographic processes should influence the whole genome, and the 'signature' of natural selection is expected to be locus-specific.

Other tests for departure from neutral expectations based on polymorphism data were developed by Fu & Li (1993) and Fu (1996, 1997). Fu and Li's (1993) tests compare the frequency of young *versus* old mutations in a gene genealogy, and may be effective in detecting the occurrence of purifying (negative) or balancing selection, and also of recent selective sweeps. Fu's (1996, 1997) studies evaluated patterns of polymorphism in DNA samples under different demographic and/or selective scenarios that depart from neutral expectations, and led to the development of new statistical tests to detect the occurrence of background selection, genetic hitchhiking, balancing selection, and also population subdivision, shrinkage or growth.

A different class of tests was developed to evaluate the neutral expectation of correlated levels of intra-specific polymorphism *versus* inter-specific divergence. The Hudson-Kreitman-Aguadé (HKA) test (Hudson et al. 1987) compares levels of polymorphism and divergence between different genomic regions, and may be effective in detecting the occurrence of recent selective sweeps or balancing selection. The McDonald-Kreitman test (McDonald & Kreitman 1991) compares synonymous *versus* nonsynonymous substitutions within a species (polymorphisms) and between species (fixed differences), and is a widely used approach to detect positive selection leading to accelerated protein evolution. Finally, recent studies have used the analysis of multi-locus (or multi-marker) haplotypes to infer departures from neutral expectations in the history of alleles of interest (e.g. Stephens et al. 1998; Tishkoff et al. 2001). Statistical tests addressing the recombination- and/or mutation-based evolution of haplotypes have been designed to test for departures from neutrality (Slatkin & Rannala 2000; Sabeti et al. 2002), and may be powerful tools to identify the historical effects of natural selection on specific genomic regions.

Given the data sets required for these tests, it is likely that future studies applying these approaches to population-level investigations of *ASIP* and *MC1R* in the context of felid melanism will likely involve the characterization of the identified variants in larger samples of individuals from natural populations, and the sequencing of adjacent genomic segments to survey levels of silent polymorphism (e.g. to apply the HKA test). In case haplotype-based analyses are desired, the STR loci identified in the *ASIP* genomic region may useful (although more distant loci may also be required for recombination-based approaches), and similar markers should be

obtained for the *MC1R* region of felids. It will also be important to pursue similar population-genetic studies addressing other genomic regions of these species, to derive estimates of historical demographic processes that may confound inferences on the 'signature' of natural selection at a particular locus. As more loci involved in coloration phenotypes are described and characterized, a broader picture of the impact of natural selection on their interactive evolutionary history may emerge. In combination with ecological and behavioral studies investigating the adaptive role of these phenotypes in the wild, such population genetic approaches may help uncover the evolutionary significance of coloration diversity in the Felidae and in other groups of organisms.

APPENDIX 1

Computer programs used in this study

Computer Program	Application in this study	Reference / Website
BLAST	DNA or protein comparison to genomic databases	Altschul et al. 1990
mVISTA	Genomic comparison	Dubchak et al. 2000; Mayor et al. 2000; http://www-gsd.lbl.gov/vista
rVISTA	Search for Transcription Factor Biding Sites	Loots et al. 2002; http://www-gsd.lbl.gov/vista
Bayes Block Aligner	Nucleotide Alignment	Zhu et al. 1998
ClustalX	Nucleotide Alignment	Thompson et al. 1997
Primer3	Primer design	http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi
RHMAP	Radiation hybrid mapping	Boehnke et al. 1991
Genotyper	Allele sizing and categorization	Commercial – ABI
PedCheck	Mendelian inheritance check	O'Connell & Weeks 1998
LINKAGE	Linkage analysis	Lathrop et al. 1984
FASTLINK	Linkage analysis	Cottingham et al. 1993
PHRED	DNA Base calling	Ewing et al 1998
PHRAP / CONSED	Preparation and assembly of genomic contigs	Gordon et al. 1998
SEQUENCHER	DNA Base calling	Commercial - Gene Codes
ReapeatMasker	Identification and exclusion of genomic repeats	A.F.A. Smit & P. Green, unpublished; http://ftp.genome.washington.edu/cgi-bin/RepeatMasker

Computer Program	Application in this study	Reference / Website
Microsatellite Target Identification Program	Search for long, uninterrupted STR loci in genomic sequences	R. Stephens and V.A. David [NCI-Frederick], unpublished
MEGA 2.1	Sequence analysis	Kumar et al. 2001
PAUP 4.0	Phylogeny estimation, molecular clock test	Swofford 1998
DnaSP	Sequence analysis	Rozas & Rozas 1999

APPENDIX 2

Transcription Factor Binding Sites (TFBSs) identified in conserved non-coding sequence blocks 5' of the *MC1R* gene (described in Chapter 3). Numbers refer to the beginning position of each TFBS in the domestic cat *MC1R* genomic sequence (Figure 3-1).

Block 5a
1043 CEBPA
1040 CEBP
1044 CEBPA
1043 OCT1_Q6
1037 STAT
1044 OCT1
1042 NFAT_Q6
1047 YY1
1043 CDXA
1050 GKLF
1044 CDXA
1048 OCT1
1047 EN1
1048 GR_Q6
1052 LDSPOLYA_B
1048 MEIS1
1048 TGIF
1047 AP1FJ_Q2
1048 AP1
1054 CEBPA
1058 CEBPA
1051 AHR
1050 P53
1054 P53
1053 SP1
1056 MSX1
1056 SRY
1061 CETS1P54
1060 IK2

Block 5b
1135 NF1_Q6
1130 AP4
1133 AP4_Q5
1135 YY1
1143 RFX1
1138 CEBPA
1143 CEBP
1140 CEBPA
1143 HLF
1145 MYB_Q6
1145 SRY

Block 5c
1168 CEBP
1176 NF1_Q6
1175 LDSPOLYA_B
1170 HSF1
1177 HAND1E47
1182 CETS1P54
1183 IK2
1182 EN1
1191 EN1
1195 AP1
1197 SRY

Block 5d
2182 CAAT
2183 MYB_Q6
2185 EN1
2191 SP1
2195 BARBIE
2197 CEBP
2196 CDXA
2200 P300
2197 CDXA
2199 ATF_B
2199 CREB
2199 HSF1
2204 HSF2
2201 NKX25
2204 AP1
2209 SP1

Block 5e
2247 CEBP
2248 CDXA
2254 AP1
2255 YY1
2258 OCT1
2255 TCF11
2258 CDXA
2263 CETS1P54
2262 MZF1

Block 5f
3353 DELTAEF1
3362 CEBP_Q2
3358 RORA1
3362 CDXA
3364 CDXA
3366 NKX25
3368 AML1
3369 SRY

Block 5g
3380 CEBP
3386 BRN2
3384 ELK1
3389 GKLF
3384 CETS1P54
3383 IK2
3383 MZF1
3385 NFAT_Q6
3389 YY1
3384 ETS1_B
3386 CDXA
3391 MZF1
3392 MZF1

Block 5h
3914 YY1
3917 YY1
3920 YY1
3915 GATA1
3919 TAL1ALPHAE47
3917 EN1
3919 MAX
3919 USF
3920 USF
3921 USF_Q6
3917 CDXA

3919 LMO2COM
3920 LMO2COM
3919 MYCMAX
3920 MYCMAX
3919 MYOD
3920 NMYC
3919 NMYC
3919 SREBP1
3918 MYCMAX_B
3920 MYOD_Q6
3917 TCF11
3920 NKX25
3920 PAX4
3922 NF1_Q6
3922 AML1
3926 GC
3927 ZF5_B
3924 ZF5_B
3926 SP1
3927 SP1
3929 AP2_Q6
3938 GKLF
3939 GKLF
3933 IK2
3934 DELTAEF1
3933 CETS1P54
3934 ETS1_B
3934 USF_C
3937 ETS1_B
3936 MZF1

3940 MZF1
3941 MZF1
3938 GC
3938 SP1
3945 CETS1P54
3946 CETS1P54
3948 E2F1_Q3

Block 5j
4076 AML1
4081 AML1
4083 CETS1P54
4088 ELK1
4092 GKLF
4088 CETS1P54
4087 IK2
4088 STAT
4087 CETS1P54
4089 GABP_B
4089 NRF2
4094 AREB6
4094 DELTAEF1
4095 GC
4100 GKLF
4100 GC
4104 GKLF
4095 SP1
4098 SP1
4100 SP1_Q6
4099 SP1

4093 AML1
4095 CEBP
4094 PAX4
4099 CETS1P54
4098 IK2
4098 MZF1
4100 GC
4098 LYF1
4098 SP1
4099 SP1
4103 SP1
4103 GC

Block 5k
4227 GR_Q6
4224 HNF4
4221 YY1
4223 USF
4224 USF
4223 LMO2COM
4224 LMO2COM
4223 PAX4
4223 MYOD_Q6
4225 SRY
4225 CEBP_Q2
4224 CDXA
4228 CETS1P54
4229 CETS1P54

APPENDIX 3

Purina Pedigree segregating for melanism: STR data for linkage analysis and genotypes for the *ASIP*-Δ 2 allele.

Purina Pedigree was divided into 4 families (Mel1, Mel2, Mel4, Mel5) to break inbreeding loops; individuals with the same ID (or suffixes 'a','b', 'c') represent the same animal, duplicated among or within families for the linkage analysis.

Allele codes for STR loci are shown at the bottom; Δ is the *ASIP*- Δ 2 allele; + is wild type allele
Zeroes for Sire and Dam indicate a founder individual (no parents included); zeroes for STR alleles indicate unavailable genotype.

Family	Purina ID	Sire	Dam	Sex	Melanism	FCA1008a		FCA1018a		FCA1019a		*ASIP*-Δ 2
Mel1	922033	0	0	M	Melanistic	1	1	4	6	6	6	Δ/Δ
Mel1	41253a	0	0	M	WT	4	4	0	0	0	0	+/+
Mel1	41253b	0	0	M	WT	4	4	0	0	0	0	+/+
Mel1	41253c	0	0	M	WT	4	4	0	0	0	0	+/+
Mel1	CB-460	0	0	F	Melanistic	1	1	4	7	3	3	Δ/Δ
Mel1	CB-461	0	0	F	Melanistic	1	1	0	0	3	3	Δ/Δ
Mel1	CE-147	0	0	F	Melanistic	1	2	5	5	0	0	Δ/Δ
Mel1	CH-27	41253a	CB-460	F	WT	1	4	0	0	3	6	+/Δ
Mel1	CF-29	41253b	CB-461	F	WT	1	4	4	7	3	6	+/Δ
Mel1	CG-159	41253c	CE-147	F	WT	0	0	4	5	4	6	+/Δ
Mel1	CK-144	922033	CF-29	M	WT	1	4	4	6	6	6	+/Δ
Mel1	CK-145	922033	CF-29	M	WT	1	4	4	6	6	6	+/Δ
Mel1	CK-146	922033	CF-29	M	Melanistic	1	1	6	7	3	6	Δ/Δ
Mel1	CK-147	922033	CF-29	F	WT	1	4	0	0	6	6	+/Δ
Mel1	CK-148	922033	CF-29	F	Melanistic	1	1	6	7	3	6	Δ/Δ
Mel1	CK-55	922033	CF-29	M	WT	1	4	4	6	0	0	+/Δ
Mel1	CK-56	922033	CF-29	M	WT	1	4	4	6	0	0	+/Δ
Mel1	CK-57	922033	CF-29	M	WT	1	4	4	6	6	6	+/Δ
Mel1	CK-58	922033	CF-29	F	WT	1	4	4	6	0	0	+/Δ
Mel1	CK-59	922033	CF-29	F	WT	1	4	4	6	0	0	+/Δ
Mel1	CK-60	922033	CF-29	F	WT	1	4	4	4	0	0	+/Δ
Mel1	CK-82	922033	CG-159	M	WT	1	4	4	6	0	0	+/Δ
Mel1	CK-83	922033	CG-159	M	WT	1	4	4	4	0	0	+/Δ
Mel1	CK-84	922033	CG-159	M	WT	1	4	4	6	0	0	+/Δ

Family	Purina ID	Sire	Dam	Sex	Melanism	FCA1008a		FCA1018a		FCA1019a		*ASIP*-Δ2
Mel1	CK-85	922033	CG-159	F	Melanistic	1	1	5	6	0	0	Δ/Δ
Mel1	CK-86	922033	CG-159	F	Melanistic	1	1	4	5	0	0	Δ/Δ
Mel1	CK-130	922033	CH-27	M	Melanistic	1	1	4	4	3	6	Δ/Δ
Mel1	CK-132	922033	CH-27	M	Melanistic	1	1	4	6	0	0	Δ/Δ
Mel1	CK-133	922033	CH-27	M	Melanistic	1	1	4	4	3	6	Δ/Δ
Mel1	CK-134	922033	CH-27	F	WT	1	4	4	6	6	6	+/Δ
Mel1	CK-135	922033	CH-27	F	Melanistic	1	1	4	4	3	6	Δ/Δ
Mel1	CK-136	922033	CH-27	F	Melanistic	1	1	4	4	3	6	Δ/Δ
Mel1	CK-46	922033	CH-27	M	Melanistic	1	1	4	4	3	6	Δ/Δ
Mel1	CK-47	922033	CH-27	M	WT	1	4	4	4	6	6	+/Δ
Mel1	CK-48	922033	CH-27	M	WT	1	4	4	4	6	6	+/Δ
Mel1	CK-49	922033	CH-27	F	WT	1	4	4	4	0	0	+/Δ
Mel2	41253a	0	0	M	WT	4	4	0	0	0	0	+/+
Mel2	41253b	0	0	M	WT	4	4	0	0	0	0	+/+
Mel2	932244	0	0	F	WT	1	4	4	7	6	6	+/Δ
Mel2	CB-460a	0	0	F	Melanistic	1	1	4	7	3	3	Δ/Δ
Mel2	CB-460b	0	0	F	Melanistic	1	1	4	7	3	3	Δ/Δ
Mel2	CB-461	0	0	F	Melanistic	1	1	0	0	3	3	Δ/Δ
Mel2	CE-115a	0	0	M	WT	1	3	3	5	3	4	+/Δ
Mel2	CE-115b	0	0	M	WT	1	3	3	5	3	4	+/Δ
Mel2	CE-147	0	0	F	Melanistic	1	2	5	5	0	0	Δ/Δ
Mel2	CF-101	0	0	F	WT	1	4	2	4	3	6	+/Δ
Mel2	CH-27	41253a	CB-460a	F	WT	1	4	0	0	3	6	+/Δ
Mel2	CF-29	41253b	CB-461	F	WT	1	4	4	7	3	6	+/Δ
Mel2	CG-159	41253b	CE-147	F	WT	0	0	4	5	4	6	+/Δ
Mel2	CH-4	0	0	F	WT	1	4	0	0	5	6	+/Δ
Mel2	CI-13	0	0	F	WT	1	4	2	4	3	6	+/Δ
Mel2	CJ-26	CE-115a	932244	M	Melanistic	1	1	3	7	4	6	Δ/Δ
Mel2	CG-15	CE-115a	CB-460b	F	Melanistic	1	1	3	7	0	0	Δ/Δ
Mel2	CG-115	CE-115a	CF-101	M	WT	3	4	4	5	3	6	+/+
Mel2	CJ-10	CE-115a	CF-29	F	WT	1	4	0	0	3	3	+/Δ
Mel2	CJ-153	CE-115a	CF-29	M	WT	3	4	4	5	3	6	+/+
Mel2	CJ-155	CE-115a	CF-29	M	WT	1	4	3	4	4	6	+/Δ

Family	Purina ID	Sire	Dam	Sex	Melanism	FCA1008a		FCA1018a		FCA1019a		*ASIP*-Δ2
Mel2	CJ-157	CE-115a	CF-29	F	WT	3	4	4	5	3	6	+/+
Mel2	CJ-158	CE-115a	CF-29	F	Melanistic	1	1	3	7	3	4	Δ/Δ
Mel2	CJ-7	CE-115a	CF-29	M	Melanistic	1	1	3	7	3	4	Δ/Δ
Mel2	CJ-8	CE-115a	CF-29	M	WT	3	4	4	5	3	6	+/+
Mel2	CJ-175	CE-115b	CG-159	F	WT	1	3	5	5	3	4	+/Δ
Mel2	CJ-143	CE-115b	CH-27	M	Melanistic	1	1	3	4	3	4	Δ/Δ
Mel2	CJ-144	CE-115b	CH-27	M	Melanistic	1	1	3	4	3	4	Δ/Δ
Mel2	CJ-145	CE-115b	CH-27	M	WT	1	4	3	4	4	6	+/Δ
Mel2	CJ-146	CE-115b	CH-27	F	WT	3	4	4	5	4	6	+/+
Mel2	CJ-147	CE-115b	CH-27	F	WT	1	4	3	4	0	0	+/Δ
Mel2	CJ-51	CE-115b	CH-27	F	WT	1	3	4	5	3	3	+/Δ
Mel2	CJ-52	CE-115b	CH-27	M	WT	1	3	4	5	3	3	+/Δ
Mel2	CJ-32	CE-115b	CH-4	M	WT	1	4	3	4	4	6	+/Δ
Mel2	CJ-36	CE-115b	CH-4	M	Melanistic	1	1	3	3	4	5	Δ/Δ
Mel2	CK-06	CE-115b	CI-13	M	WT	1	4	3	4	4	6	+/Δ
Mel2	CK-07	CE-115b	CI-13	F	WT	3	4	4	5	3	6	+/+
Mel2	CK-08	CE-115b	CI-13	F	Melanistic	1	1	2	3	3	4	Δ/Δ
Mel4	41253	0	0	M	WT	4	4	0	0	0	0	+/+
Mel4	932244	0	0	F	WT	1	4	4	7	6	6	+/Δ
Mel4	CE-147	0	0	F	Melanistic	1	2	5	5	0	0	Δ/Δ
Mel4	CH-4	0	0	F	WT	1	4	0	0	5	6	+/Δ
Mel4	CG-158	41253	CE-147	M	WT	2	4	4	5	4	6	+/Δ
Mel4	CH-5	0	0	F	WT	1	2	6	7	3	6	+/Δ
Mel4	CK-50	CG-158	932244	M	Melanistic	1	2	5	7	4	6	Δ/Δ
Mel4	CK-52	CG-158	932244	F	WT	4	4	4	4	6	6	+/+
Mel4	CK-69	CG-158	CH-4	M	WT	4	4	3	4	0	0	+/+
Mel4	CK-70	CG-158	CH-4	F	Melanistic	1	2	3	5	0	0	Δ/Δ
Mel4	CJ-86	CG-158	CH-5	M	WT	1	4	4	7	6	6	+/Δ
Mel4	CJ-87	CG-158	CH-5	M	WT	2	4	4	6	3	6	+/+
Mel4	CJ-88	CG-158	CH-5	M	WT	0	0	5	6	3	4	+/Δ
Mel4	CJ-89	CG-158	CH-5	F	WT	2	2	0	0	3	4	+/Δ
Mel4	CJ-90	CG-158	CH-5	F	WT	2	4	4	6	0	0	+/+
Mel4	CK-34	CG-158	CH-5	F	Melanistic	1	2	5	7	4	6	Δ/Δ

Family	Purina ID	Sire	Dam	Sex	Melanism	FCA1008a		FCA1018a		FCA1019a		*ASIP*-Δ2
Mel4	CK-35	CG-158	CH-5	F	Melanistic	1	2	5	7	4	6	Δ/Δ
Mel4	CK-36	CG-158	CH-5	F	WT	2	4	4	6	0	0	+/+
Mel5	41253	0	0	M	WT	4	4	0	0	0	0	+/+
Mel5	C7-142	0	0	F	WT	1	2	3	6	3	5	+/Δ
Mel5	CE-147	0	0	F	Melanistic	1	2	5	5	0	0	Δ/Δ
Mel5	CH-78	0	0	M	WT	1	7	3	4	3	3	+/Δ
Mel5	CF-64	41253	C7-142	F	WT	1	4	0	0	5	6	+/Δ
Mel5	CI-39	0	0	F	WT	1	5	3	4	3	6	+/Δ
Mel5	CK-79	CH-78	CE-147	M	WT	1	7	3	5	0	0	+/Δ
Mel5	CK-80	CH-78	CE-147	F	Melanistic	0	0	4	5	0	0	Δ/Δ
Mel5	CJ-93	CH-78	CF-64	M	Melanistic	1	1	3	4	3	5	Δ/Δ
Mel5	CJ-94	CH-78	CF-64	M	WT	1	4	4	4	3	6	+/Δ
Mel5	CJ-95	CH-78	CF-64	M	WT	1	4	4	4	3	6	+/Δ
Mel5	CJ-97	CH-78	CF-64	F	WT	1	4	4	4	3	6	+/Δ
Mel5	CJ-98	CH-78	CF-64	F	WT	1	7	3	3	0	0	+/Δ
Mel5	CJ-68	CH-78	CI-39	F	WT	1	7	3	3	3	6	+/Δ

FCA1008a alleles (in bp): 127=1, 129=2, 139=3, 141=4, 145=5, 147=6, 149=7.
FCA1018a alleles (in bp): 227=1, 235=2, 239=3, 243=4, 247=5, 251=6, 255=7.
FCA1019a alleles (in bp): 186=1, 194=2, 198=3, 202=4, 206=5, 210=6.

APPENDIX 4

Samples included in this study, with genotypes for *ASIP* and *MC1R* deletions (except where stated, all listed individuals are unrelated).

Alleles are coded as follows: + = non-deleted (ancestral) allele; Δ2 = *ASIP*-Δ2 deletion allele; Δ15 is the *MC1R*-Δ15 deletion allele; Δ24 is the *MC1R*-Δ24 deletion allele.

NT= not typed; LGD=Sample collection of the Laboratory of Genomic Diversity.

LGD code	Color	Geographic Origin / Breed	Contact	Genotype *ASIP*	Genotype *MC1R*
Domestic cat (*Felis catus*)					
FCA-123*	?	domestic shorthair, USA	LGD	NT	+/+
FCA-159	spotted	Egyptian Mau	LGD	+/Δ2	NT
FCA-1670	black/white	outbred, MD, USA	M. Roelke	Δ2/Δ2	+/+
FCA-1902	abyssinian	Abyssinian	LGD	+/+	+/+
FCA-1904	brown tabby/white	domestic shorthair, USA	LGD	+/Δ2	+/+
FCA-1910	ruddy abyssinian	Somali	LGD	+/+	+/+
FCA-1911	black smoke	Maine Coon	LGD	Δ2/Δ2	+/+
FCA-1912	silver tabby	Maine Coon	LGD	+/Δ2	+/+
FCA-1913	brown tabby	Maine Coon	LGD	+/Δ2	NT
FCA-1914	brown blotched tabby	Norwegian Forest Cat	LGD	+/+	NT
FCA-1915	black	Norwegian Forest Cat	LGD	Δ2/Δ2	+/+
FCA-1928	tabby/white	British Shorthair	LGD	+/Δ2	+/+
FCA-194*	?	domestic shorthair, USA	LGD	NT	+/+
FCA-2096	black	domestic shorthair, Frederick, MD, USA	K. Nalewaik	Δ2/Δ2	+/+
FCA-2119	brown mackerel tabby	Norwegian Forest Cat	LGD	+/Δ2	NT
FCA-2148	agouti/white	Norwegian Forest Cat	LGD	+/Δ2	NT
FCA-215*	?	domestic shorthair, USA	LGD	NT	+/+
FCA-2152	black/white	Norwegian Forest Cat	LGD	Δ2/Δ2	+/+
FCA-2162	black	Norwegian Forest Cat	LGD	Δ2/Δ2	+/+
FCA-2297	abyssinian	Abyssinian	LGD	+/+	+/+
FCA-2305	black	Bombay	LGD	Δ2/Δ2	NT
FCA-2308	silver tabby	Maine Coon	LGD	+/Δ2	NT
FCA-2348	silver mackerel tabby	Scottish Fold	LGD	+/+	+/+
FCA-2352	brown blotched tabby	domestic shorthair, USA	LGD	+/Δ2	NT
FCA-2355	brown tabby	domestic shorthair, USA	LGD	+/Δ2	NT
FCA-2384	black smoke	Cornish Rex	LGD	Δ2/Δ2	NT
FCA-2410	brown blotched tabby	domestic shorthair, USA	LGD	+/Δ2	NT
FCA-2417	black/white	Turkish Van	LGD	Δ2/Δ2	+/+
FCA-2461	black/white	domestic shorthair, USA	LGD	Δ2/Δ2	+/+
FCA-2468	black	domestic shorthair, USA	LGD	Δ2/Δ2	NT
FCA-2545	blue abyssinian	Abyssinian	LGD	+/+	+/+

LGD code	Color	Geographic Origin / Breed	Contact	Genotype *ASIP*	Genotype *MC1R*
FCA-2556	black	Bombay	LGD	Δ2/Δ2	+/+
FCA-2580	black	Bombay	LGD	Δ2/Δ2	+/+
FCA-2612	black	domestic shorthair, Clinton, NY, USA	P. Menotti	Δ2/Δ2	+/+
FCA-2618	brown tabby	domestic shorthair, USA	LGD	+/Δ2	NT
FCA-264	black/white	domestic shorthair, USA	L. Lyons	Δ2/Δ2	+/+
FCA-2641	black	outbred, Israel	C. Driscoll	Δ2/Δ2	+/+
FCA-2682	black	domestic shorthair, Sylvania, OH	N. Huntzinger	Δ2/Δ2	+/+
FCA-2683	black/white	domestic shorthair, USA	V. Hughes / L. Strathern	Δ2/Δ2	+/+
FCA-2684	black	outbred, Frederick, MD, USA	L. Lyons/LGD	Δ2/Δ2	+/+
FCA-2685	black	outbred, Porto Alegre, RS, Brazil	T. C. Trigo	Δ2/Δ2	+/+
FCA-2686	black	outbred, Porto Alegre, RS, Brazil	E. Rodrigues, M. Gomes	Δ2/Δ2	NT
FCA-2687	black/white, S=7	outbred, Porto Alegre, RS, Brazil	E. Rodrigues, M. Gomes	Δ2/Δ2	+/+
FCA-2688	black/white, S=4	outbred, Porto Alegre, RS, Brazil	E. Rodrigues, M. Gomes	Δ2/Δ2	+/+
FCA-2689	black-calico	outbred, Porto Alegre, RS, Brazil	E. Rodrigues, M. Gomes	Δ2/Δ2	NT
FCA-2690	black/white, S=6	outbred, Porto Alegre, RS, Brazil	E. Rodrigues, M. Gomes	Δ2/Δ2	NT
FCA-2691	black w/ white, S=1	outbred, Porto Alegre, RS, Brazil	E. Rodrigues, M. Gomes	Δ2/Δ2	NT
FCA-2692	black-calico	outbred, Porto Alegre, RS, Brazil	E. Rodrigues, M. Gomes	Δ2/Δ2	NT
FCA-2693	black	outbred, Porto Alegre, RS, Brazil	E. Rodrigues, M. Gomes	Δ2/Δ2	NT
FCA-2694	black	outbred, Porto Alegre, RS, Brazil	E. Rodrigues, M. Gomes	Δ2/Δ2	NT
FCA-2699	black	Bombay	LGD	Δ2/Δ2	+/+
FCA-273	blotched tabby	domestic shorthair, USA	LGD	+/Δ2	+/+
FCA-2782	black	Sphynx	LGD	Δ2/Δ2	+/+
FCA-2815	brown blotched tabby	British shortahir	LGD	+/+	+/+
FCA-2817	silver mackerel tabby	Manx	LGD	+/Δ2	NT
FCA-2877	black	outbred, Pennsylvania, USA	N. Sullivan	Δ2/Δ2	NT
FCA-2878	black	outbred, Pennsylvania, USA	N. Sullivan	Δ2/Δ2	NT
FCA-2879	black	outbred, Pennsylvania, USA	N. Sullivan	Δ2/Δ2	NT
FCA-2880	black	outbred, Pennsylvania, USA	N. Sullivan	Δ2/Δ2	NT
FCA-2881	black	outbred, Pennsylvania, USA	N. Sullivan	Δ2/Δ2	NT
FCA-2882	black	outbred, Pennsylvania, USA	N. Sullivan	Δ2/Δ2	NT
FCA-2883	black	outbred, Pennsylvania, USA	N. Sullivan	Δ2/Δ2	NT
FCA-2884	black	outbred, Pennsylvania, USA	N. Sullivan	Δ2/Δ2	NT
FCA-2885	black	outbred, Pennsylvania, USA	N. Sullivan	Δ2/Δ2	NT
FCA-2886	black	domestic shorthair, Oldwick, NJ, USA	N. Sullivan	Δ2/Δ2	NT
FCA-296	black	domestic shorthair, USA	LGD	Δ2/Δ2	+/+
FCA-2976	abyssinian-lined tabby	Abyssinian heterozygote	LGD	+/+	NT
FCA-3004	Silver spotted	Egyptian Mau	LGD	+/Δ2	+/+
FCA-4060	black	outbred, Israel	A. Boldo, G. Bar-Gal	Δ2/Δ2	NT

LGD code	Color	Geographic Origin / Breed	Contact	Genotype *ASIP*	Genotype *MC1R*
FCA-4061	black	outbred, Israel	A. Boldo, G. Bar-Gal	Δ2/Δ2	+/+
FCA-4062	black	outbred, Israel	A. Boldo, G. Bar-Gal	Δ2/Δ2	NT
FCA-4063	black	outbred, Israel	A. Boldo, G. Bar-Gal	Δ2/Δ2	+/+
FCA-4064	black	outbred, Botswana	M. Roelke	Δ2/Δ2	+/+
FCA-4065	black	domestic shorthair, USA	M. Culver	Δ2/Δ2	NT
FCA-4071	black	outbred, Mongolia	M.Brown, OSU	Δ2/Δ2	NT
FCA-4075	black	outbred, Mongolia	M.Brown, OSU	Δ2/Δ2	NT
FCA-4076	black	outbred, Mongolia	M.Brown, OSU	Δ2/Δ2	NT
FCA-409	black	Bombay	LGD	Δ2/Δ2	+/+
FCA-4094	black	outbred, Mongolia	A.Fine,MSU	Δ2/Δ2	NT
FCA-475	black	Bombay	LGD	Δ2/Δ2	+/+
FCA-508	black	domestic shorthair, USA	LGD	Δ2/Δ2	NT
FCA-637*	brown	Havana Brown	LGD	Δ2/Δ2	NT
FCA-733*	brown	Havana Brown	LGD	Δ2/Δ2	NT
FCA-800	black	domestic shorthair, USA	LGD	Δ2/Δ2	+/+
FCA-802	blotched tabby	domestic shorthair, USA	LGD	+/+	NT
FCA-927	black	domestic shorthair, USA	LGD	Δ2/Δ2	NT
FCA-934	blotched tabby	domestic shorthair, USA	LGD	+/+	+/+
FCA-938	mackerel tabby	domestic shorthair, USA	LGD	+/+	NT

* samples not included in the genotyping for Figure 4-3; Havana Brown cats are expected to be black at the 'non-agouti' melanism locus (here identified as the *ASIP* gene), which is confirmed by these deletion data; nevertheless these individuals were excluded from the genotyping analysis since they were not exactly black.

LGD code	Color	Geographic Origin / Captive Institution	Contact / Institution	Genotype *ASIP*	Genotype *MC1R*
Jaguar (*Panthera onca*)					
PON-03	**Melanistic**	Ceri Zoo, USA	LGD	+/+	+ / Δ15
PON-13	**Melanistic**	Philadelphia Zoo, USA	LGD	+/+	+ / Δ15
PON-14*	**Melanistic**	Prescott Animal Park, USA	LGD	NT	Δ15/Δ15
PON-146	wild type	Brazil	D. Sana, A. Pro-Carnivoros	+/+	+/+
PON-147	wild type	Brazil	D. Sana, A. Pro-Carnivoros	+/+	+/+
PON-148	wild type	Brazil	D. Sana, A. Pro-Carnivoros	NT	+/+
PON-149	wild type	Brazil	D. Sana, A. Pro-Carnivoros	NT	+/+
PON-150	**Melanistic**	Brazil	D. Sana, A. Pro-Carnivoros	NT	+ / Δ15
PON-151	wild type	Brazil	D. Sana, A. Pro-Carnivoros	NT	+/+
PON-152	wild type	Brazil	D. Sana, A. Pro-Carnivoros	NT	+/+
PON-154	wild type	Brazil	D. Sana, A. Pro-Carnivoros	NT	+/+

LGD code	Color	Geographic Origin / Captive Institution	Contact / Institution	Genotype *ASIP*	Genotype *MC1R*
PON-155	**Melanistic**	Brazil	L.Cullen, IPE	NT	+/Δ15
PON-156	**Melanistic**	Sapucaia do Sul Zoo, Brazil	T. Freitas, T. Trigo	+/+	+/Δ15
PON-157	**Melanistic**	Brazil	R. Morato, A. Pro-Carnivoros	NT	+/Δ15
PON-16	wild type	Guatemala	ARCAS	NT	+/+
PON-18	**Melanistic**	San Diego Zoo, USA	LGD	NT	+/Δ15
PON-20	wild type	Panama	Summit Zoo	+/+	+/+
PON-21	wild type	Costa Rica	Las Pumas	NT	+/+
PON-22	wild type	Costa Rica	Simon Bolivar Zoo	NT	+/+
PON-23	wild type	Nicaragua	Juigalpa Zoo	NT	+/+
PON-25	wild type	Nicaragua	Managua Zoo	NT	+/+
PON-28	wild type	Probably Peru	Mendoza Zoo	NT	+/+
PON-31	wild type	Mexico	Leon Zoo	+/+	+/+
PON-43	wild type	Brazil	P. Crawshaw, CENAP/IBAMA	NT	+/+
PON-44	wild type	Brazil	P. Crawshaw, CENAP/IBAMA	NT	+/+
PON-46	wild type	Brazil	Goiania Zoo	NT	+/+
PON-48	**Melanistic**	Brazil	Goiania Zoo	NT	Δ15/Δ15
PON-52	wild type	Mexico	Idaho State University	NT	+/+
PON-54	wild type	Venezuela	Las Delicias	NT	+/+
PON-56	wild type	Venezuela	Las Delicias	NT	+/+
PON-57	wild type	Venezuela	Las Delicias	NT	+/+
PON-58	wild type	Venezuela	Las Delicias	NT	+/+
PON-62	wild type	Venezuela	Valencia Castellito	+/+	+/+
PON-63	wild type	Venezuela	Valencia Castellito	NT	+/+
PON-64	wild type	Bolivia	Santa Cruz Zoo	NT	+/+
PON-66	wild type	Bolivia	Santa Cruz Zoo	NT	+/+
PON-67	wild type	Venezuela	PROFAUNA	NT	+/+
PON-68	wild type	French Guiana	J.C. Vie	NT	+/+
PON-72**	wild type	Brazil	CENAP/IBAMA	+/+	+/+
PON-75	wild type	Mexico	B. Miller	NT	+/+
PON-76	wild type	Brazil	CIGS	NT	+/+
PON-85*	**Melanistic**	EFBC' Feline Conservation Center, CA, USA	J. Maynard	NT	+/Δ15
PON-86*	wild type	EFBC' Feline Conservation Center, CA, USA	J. Maynard	NT	+/+
PON-87*	wild type	Omaha Zoo, NE, USA	J. Maynard	NT	+/+
PON-88*	**Melanistic**	EFBC' Feline Conservation Center, CA, USA	J. Maynard	NT	+/Δ15
PON-89*	wild type	EFBC' Feline Conservation Center, CA, USA	J. Maynard	NT	+/+
PON-90*	wild type	Dallas World Aquarium / EFBC's FCC	J. Maynard	NT	+/+
PON-91	**Melanistic**	Montgomery Zoo, AL, USA	C. Hilton	NT	+/Δ15

LGD code	Color	Geographic Origin / Captive Institution	Contact / Institution	Genotype *ASIP*	Genotype *MC1R*
PON-92**	Melanistic	wild, Brazil	L.Cullen, IPE	+/+	+ / Δ15
PON-CRSB145	wild type	Costa Rica	R. Spindler / S. Bolivar Zoo	NT	+/+
PON-CRSB23	wild type	Costa Rica	R. Spindler / S. Bolivar Zoo	NT	+/+
PON-CRSB24	wild type	Costa Rica	R. Spindler / S. Bolivar Zoo	NT	+/+
PON-CRSB61	wild type	Costa Rica	R. Spindler / S. Bolivar Zoo	NT	+/+

*part of jaguar pedigree shown in Figure 4-6, along with PON-91; only PON-91 was included in the association study, representing this family.
** shown for full *MC1R* sequence in Figure 4-4 (*MC1R*-Δ15 allele cloned for PON-92)

LGD code	Color	Geographic Origin / Captive Institution	Contact / Institution	Genotype *ASIP*	Genotype *MC1R*
Jaguarundi (*Herpailurus yaguarondi*)					
HYA-01	?	Mexico	Blijdorp Zoo, Netherlands	NT	Δ24 /Δ24
HYA-05	?	?	Houston Zoo	NT	+ / Δ24
HYA-07	?	?	San Diego Zoo	NT	+ / Δ24
HYA-08	?	?	San Diego Zoo	NT	+/+
HYA-10*	reddish	Costa Rica	Las Pumas	NT	+/+
HYA-12**	dark brown/gray	northern Central America	Auto Safari Chapin, Guatemal	NT	+ / Δ24
HYA-13*	reddish	Guatemala	Auto Safari Chapin, Guatemal	+/+	+ / Δ24
HYA-14	?	Guatemala	Auto Safari Chapin, Guatemal	NT	+/+
HYA-15*	reddish	Guatemala	Auto Safari Chapin, Guatemal	+/+	+/+
HYA-16**	dark brown	Argentina	Buenos Aires Zoo	+/+	+ / Δ24
HYA-17**	dark brown/gray	Argentina	Buenos Aires Zoo	+/+	+ / Δ24
HYA-18***	very dark gray	Mexico	Leon Zoo	+/+	Δ24 /Δ24
HYA-19*	reddish	Mexico	Leon Zoo	+/+	+ / Δ24
HYA-20**	dark brown	Mexico	African Safari	+/+	Δ24 /Δ24
HYA-22*	reddish	Mexico	African Safari	+/+	+/+
HYA-23**	dark brown	Mexico	African Safari	NT	+ / Δ24
HYA-24***	very dark gray/brown	Mexico	Cd. Victoria Zoo	NT	Δ24 /Δ24
HYA-25	?	Mexico	Hermosillo	NT	Δ24 /Δ24
HYA-26***	very dark gray	Mexico	Aragon Zoo	+/+	Δ24 /Δ24
HYA-30**	dark	Brazil	Sao Paulo Zoo	+/+	Δ24 /Δ24
HYA-31*	reddish	Brazil	P.Crawshaw, CENAP/IBAMA	NT	+ / Δ24
HYA-32*	red	Brazil	Sorocaba Zoo	+/+	+/+
HYA-33*	red	Brazil	Goiania Zoo	+/+	+/+
HYA-34***	very dark gray	Brazil	Brasilia Zoo	+/+	Δ24 /Δ24
HYA-35***	very dark gray	Brazil	Brasilia Zoo	NT	+ / Δ24
HYA-36**	dark	Brazil	CASIB / W. Moraes	+/+	Δ24 /Δ24

LGD code	Color	Geographic Origin / Captive Institution	Contact / Institution	Genotype *ASIP*	Genotype *MC1R*
HYA-37*	reddish brown	Brazil	CASIB / W. Moraes	+/+	+ / Δ24
HYA-40	?	Argentina	Cordoba Zoo	NT	+/+
HYA-41	?	Argentina	Cordoba Zoo	NT	+/+
HYA-42	?	Argentina	Cordoba Zoo	NT	+/+
HYA-43	?	Argentina	Cordoba Zoo	NT	+/+
HYA-44**	dark brown	Venezuela	Leslie Pantin Zoo	NT	+ / Δ24
HYA-45**	dark brown	Venezuela	Leslie Pantin Zoo	NT	+ / Δ24
HYA-46*	reddish	Venezuela	Barquisimento	NT	+/+
HYA-47***	very dark gray	Venezuela	Barquisimento	NT	Δ24 /Δ24
HYA-48**	dark brown	Venezuela	Barquisimento	NT	+ / Δ24
HYA-49**	dark brown	Venezuela	Barquisimento	NT	+ / Δ24
HYA-50***	very dark gray	Bolivia	Santa Cruz Zoo	NT	Δ24 /Δ24
HYA-59	?	French Guiana	J.C. Vie	NT	Δ24 /Δ24
HYA-60**	dark brown	southern Brazil	Sapucaia do Sul Zoo	NT	+ / Δ24

* categorized as red/reddish in Figure 4-5
** categorized as dark brown/ gray in Figure 4-5
*** categorized as very dark in Figure 4-5
Note: HYA-15 (Hya-Red) and HYA-36 (Hya-Dark) are shown for full *MC1R* sequence in Figure 4-4.

LGD code	Color	Geographic Origin / Captive Institution	Contact / Institution	Genotype *ASIP*	Genotype *MC1R*
Leopard (*Panthera pardus*)					
PPA-30	wild type	Afghanistan	Lincoln Park Zoo	NT	+/+
PPA-47	wild type	Russia	St. Louis Zoo	NT	+/+
PPA-281	wild type	Thailand	Khao Kheo Open Zoo	NT	+/+
PPA-284	**Melanistic**	Malaysia	Khao Kheo Open Zoo	+/+	+/+
PPA-286	wild type	Thailand	Khao Kheo Open Zoo	NT	+/+
PPA-288	**Melanistic**	Thailand	Dusit Zoo	+/+	+/+
PPA-296	**Melanistic**	Thailand	Chiangmai Zoo	+/+	+/+
PPA-824	**Melanistic**	EFBC' Feline Conservation Center, CA, USA	J. Maynard	+/+	+/+
Geoffroy's cat (*Oncifelis geoffroyi*)					
OGE-26	wild type	Argentina	Cordoba Zoo	+/+	+/+
OGE-53	**Melanistic**	Brazil	Cach. Sul Zoo, RS /E.Salomai	+/+	+/+
OGE-54	**Melanistic**	Brazil	Cach. Sul Zoo, RS /E.Salomai	+/+	+/+
OGE-59	wild type	Bolivia	Santa Cruz Zoo	+/+	+/+
Tigrina (*Leopardus tigrinus*)					
LTI-41	wild type	Brazil	CASIB / W. Moraes	+/+	+/+
LTI-43	wild type	Brazil	CASIB / W. Moraes	+/+	+/+

LGD code	Color	Geographic Origin / Captive Institution	Contact / Institution	Genotype *ASIP*	Genotype *MC1R*
LTI-66	**Melanistic**	Brazil	D. Sana, A. Pro-Carnivoros	+/+	+/+
Pampas cat (*Lynchailurus colocolo*)					
LCO-02	wild type	Blijdorp Zoo, Netherlands	LGD	+/+	+/+
LCO-13	wild type	Brazil	P.Crawshaw, CENAP/IBAMA	+/+	+/+
LCO-16	wild type	Peru	LGD	NT	+/+
LCO-30	**Melanistic**	Cincinnatti Zoo, USA	W. Swanson	+/+	+/+
Asian golden cat (*Catopuma temminckl*)					
PTE-01	wild type	Ceri Zoo, USA	LGD	+/+	+/+
PTE-38	**Melanistic**	Thailand	Dusit Zoo	+/+	+/+
European wildcat (*Felis silvestris*)					
FSI-07	wild type	ISEC, USA	LGD	NT	+/+
Black-footed cat (*Felis nigripes*)					
FNI-14	wild type	South Africa	San Diego Zoo	+/+	+/+
Sand cat (*Felis margarita*)					
FMA-11	wild type	Washington Park Zoo, USA	LGD	+/+	+/+
Lion (*Panthera leo*)					
PLE-68	wild type	India	LGD	+/+	+/+
PLE-150	wild type	South Africa	LGD	+/+	+/+
Tiger (*Panthera tigris*)					
PTI-122	wild type	Russia	LGD	+/+	+/+
PTI-184	wild type	Sumatra	LGD	+/+	+/+
Snow leopard (*Panthera uncia*)					
PUN-20	wild type	Cheyenne Mountain Zoo, USA	LGD	+/+	+/+
Clouded leopard (*Neofelis nebulosa*)					
NNE-09	wild type	National Zoo, USA	LGD	NT	+/+
NNE-27	wild type	China	Cleveland MetroParks	+/+	+/+
Ocelot (*Leopardus pardalis*)					
LPA-70	wild type	Brazil	Brasilia Zoo	+/+	+/+

LGD code	Color	Geographic Origin / Captive Institution	Contact / Institution	Genotype *ASIP*	Genotype *MC1R*
Puma (*Puma concolor*)					
PCO-353	wild type	USA	J. Laundre/Idaho State Univ.	+/+	+/+
PCO-560	wild type	Argentina	Buenos Aires Zoo	NT	+/+
Cheetah (*Acinonyx jubatus*)					
AJU-78	wild type	San Diego Zoo, USA	LGD	+/+	+/+
AJU-138	wild type	Tanzania	T. Caro	NT	+/+
Asian leopard cat (*Prionailurus bengalensis*)					
PBE-30*	wild type	Tallinn Zoo, Estonia	LGD	NT	+/+
PBE-33	wild type	Tallinn Zoo, Estonia	LGD	+/+	+/+
Flat-headed cat (*Prionailurus planiceps*)					
IPL-04	wild type	Lincoln Park Zoo, USA	LGD	NT	+/+
Caracal (*Caracal caracal*)					
CCA-21	wild type	Central Florida Zoo, USA	LGD	+/+	+/+
African golden cat (*Profelis aurata*)					
PAU-02	wild type	Blijdorp Zoo, Netherlands	LGD	+/+	+/+
Bobcat (*Lynx rufus*)					
LRU-04	wild type	Brookfield Zoo, USA	LGD	+/+	+/+
Bornean Bay cat (*Catopuma badia*)					
PBA-01	wild type	Ceri Zoo, USA	LGD	NT	+/+
Serval (*Leptailurus serval*)					
LSE-18	wild type	Octagon Wildlife Park, USA	M. Roelke	NT	+/+

*excluded from Figure 4-5, as this individual is related to PBE-33.

APPENDIX 5

Pedigree of captive jaguars (*Panthera onca*) used to perform transmission analysis for melanism.

Identification numbers are from the North American regional studbook for the jaguar (G. McMillan, ed., Knoxville Zoo, 1996; with updated information provided by J. Maynard, L. Peek and S. Winslow), except for the following: ID's with an 'm' suffix are from the Montgomery Zoo, AL, USA; those with a 'br' suffix are from the Baton Rouge Zoo, LA, USA; those with names are from EFBC's Feline Conservation Center, CA, USA and from Omaha Zoo, NE, USA.

Color: WT indicates a non-melanistic (wild-type) individual.

Individuals with a 'Pon' ID in parentheses were analyzed at the molecular level in this study, and are listed in Appendix 4 under this number (and displayed in Figure 4-6).

Non-melanistic offspring from a pair of black jaguars are shown in bold, corroborating an autosomal dominant mode of inheritance.

ID	Sex	Sire	Dam	Color
288	M	0	0	WT
289	F	0	0	WT
307	F	0	0	WT
312	M	0	0	Black
381	F	0	0	WT
391	F	0	0	WT
405	F	288	289	WT
420	M	0	0	Black
529	F	0	0	WT
544	F	312	381	WT
545	F	312	381	WT
546	M	312	381	Black
595	M	312	381	Black
596	F	312	381	Black
602	M	312	381	Black
603	M	312	381	Black
615	M	312	381	Black
636	M	288	289	WT
657	M	312	381	Black
663	F	312	405	WT
664	F	312	405	WT
671	M	0	0	WT
686	F	312	381	WT
687	F	312	381	Black
688	F	312	381	Black
694	F	312	307	Black
695	F	312	307	Black
698	U	312	529	Black
699	U	312	529	WT
705	M	312	381	WT
706	F	312	381	Black
707	F	312	381	WT
714	F	312	529	WT
715	F	312	529	WT
717	M	420	544	Black
718	F	420	544	WT
723	F	0	0	WT

ID	Sex	Sire	Dam	Color
726	F	603	545	WT
727	F	603	545	Black
728	M	312	307	Black
729	M	312	307	Black
733	M	312	529	Black
734	F	312	529	Black
735	M	595	381	Black
750	F	595	596	Black
751	F	595	529	Black
752	F	595	529	WT
753	F	595	381	Black
754	M	595	381	WT
760	M	615	391	Black
763	F	603	545	WT
764	F	603	545	WT
778	M	603	545	WT
779	F	603	545	WT
784	M	312	381	WT
785	F	312	381	WT
786	U	312	381	WT
790	F	728	706	Black
791	F	728	706	Black
795	F	420	544	WT
797	F	420	544	Black
806	F	733	694	Black
808	U	735	734	Black
813	U	312	381	Black
814	M	717	726	Black
815	F	717	726	Black
816	F	735	734	Black
817	F	735	734	Black
823	M	735	734	Black
824	**F**	**735**	**734**	**WT**
827	F	657	663	Black
833	U	733	723	Black
837	**M**	**733**	**694**	**WT**
838	**F**	**733**	**694**	**WT**
839	F	733	694	Black
852	M	733	723	WT
853	M	733	723	Black
870	F	733	723	Black
876	M	657	663	Black
888	M	636	827	WT
896	F	636	827	Black
897	M	636	827	Black
898	M	733	723	Black
899	F	733	723	Black
900	F	657	663	Black
923	M	657	663	Black
924	F	657	663	Black
936 (Pon85)	F	898	896	Black
947	F	733	694	Black
952	F	657	663	Black

ID	Sex	Sire	Dam	Color
960 (Pon14)	M	898	896	Black
964 (Pon86)	M	0	0	WT
968	M	657	663	Black
970	F	671	806	Black
979	F	898	896	Black
980	M	898	896	Black
1000 (Pon91)	M	964	936	Black
1001	M	964	936	Black
1002	F	964	936	WT
1008 (Pon90)	M	964	936	WT
1016	F	898	896	Black
2497m	M	1000	1016	Black
2498m	F	1000	1016	Black
2704m	M	1000	1016	Black
2705m	**F**	**1000**	**1016**	**WT**
2851m	M	1000	1016	Black
2852m	**F**	**1000**	**1016**	**WT**
2924m	M	1000	1016	Black
2925m	F	1000	1016	Black
8669br	M	898	896	Black
9267br	**F**	**8669br**	**2498m**	**WT**
9268br	F	8669br	2498m	Black
Annie	F	964	936	WT
Calamity (Pon87)	F	964	936	WT
Cisco (Pon89)	M	964	936	WT
Doc (Pon88)	M	964	936	Black

LIST OF REFERENCES

Abdel-Malek Z.A., Scott M.C., Furumura M., Lamoreux M.L., Ollmann M., Barsh G.S., Hearing V.J. 2001. The melanocortin 1 receptor is the principal mediator of the effects of agouti signaling protein on mammalian melanocytes. **J. Cell. Sci. 114**: 1019-1024.

Adachi S., Morii E., Kim D., Ogihara H., Jippo T., Ito A., Lee Y.M., Kitamura Y. 2000. Involvement of mi-transcription factor in expression of alpha-melanocyte-stimulating hormone receptor in cultured mast cells of mice. **J. Immunol. 164**: 855-860.

Aksan I. & Goding C.R. 1998. Targeting the microphthalmia basic helix-loop-helix-leucine zipper transcription factor to a subset of E-box elements in vitro and in vivo. **Mol Cell Biol. 18**: 6930-6938.

Alberts B., Bray D., Lewis J., Raff. M., Roberts K., Watson J.D. 1994. **Molecular Biology of the Cell**, 3rd. ed. Garland, New York.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D. 1990. Basic local alignment search tool. **J. Mol. Biol. 215**: 403-410.

Angwin, D. 1975. Melanistic serval cats, melanistic wild cat and bongo. **Bull. E. Afr. Nat Hist. Soc. 3.**

Barsh, G.S. 1995. Pigmentation, pleiotropy, and genetic pathways in humans and mice. **Am. J. Hum. Genet. 57**: 743-747.

Barsh, G.S. 1996. The genetics of pigmentation: from fancy genes to complex traits. **Tr. in Genet. 12**: 299-305.

Bateson W. 1913. **Mendel's Principles of Heredity**. Cambridge Univ. Press, Cambridge.

Beck T.W., Menninger J., Voigt G., Newmann K., Nishigaki Y., Nash W.G., Stephens R.M., Wang Y., de Jong P.J., O'Brien S.J., Yuhki N. 2001. Comparative feline genomics: a BAC/PAC contig map of the major histocompatibility complex class II region. Genomics 71: 282-295.

Beddard F.E. 1895. **Animal Coloration**, 2nd ed. Swan Sonnenschein, London.

Benton M.J. 1993. **The Fossil Record 2.** Chapman & Hall, London.

Bergelson J., Kreitman M., Stahl E. A., Tian D. 2001. Evolutionary dynamics of plant R-genes. **Science 292**: 2281-2285.

Bertolotto C., Abbe P., Hemesath T.J., Bille K., Fisher D.E., Ortonne J.P., Ballotti R. 1998. Microphthalmia gene product as a signal transducer in cAMP-induced differentiation of melanocytes. **J. Cell Biol. 142**: 827-835.

Bininda-Emonds O.R.P., Decker-Flum D.M., Gittleman J.L. 2001. The utility of chemical signals as phylogenetic characters: an example from the Felidae. **Biol. J. of the Linnean Soc. 72**: 1-15.

Boehnke M., Lange K., Cox D.R. 1991. Statistical methods for multipoint radiation hybrid mapping. **Am. J. Hum. Genet. 49**: 1174-1188.

Box N.F., Wyeth J.R., O'Gorman L.E., Martin N.G., Sturm R.A. 1997. Characterization of melanocyte stimulating hormone receptor variant alleles in twins with red hair. **Hum. Mol. Genet. 6**: 1891-1897.

Bromham L., Penny D., Rambaut A., Hendy M.D. 2000. The power of relative rates tests depends on the data. **J. Mol. Evol. 50**: 296-301.

Bultman, S.J., Michaud E.J., Woychik R.P. 1992. Molecular characterization of the mouse agouti locus. **Cell 71**: 1195-1204.

Bultman S.J., Klebig M.L., Michaud E.J., Sweet H.O., Davisson M.T., Woychik R.P. 1994. Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a^t) and white-bellied agouti (A^w) reveals alternative forms of agouti transcripts. **Genes and Dev. 8**: 481-490.

Burghes A.H., Logan C., Hu X., Belfall B., Worton R.G., Ray P.N. 1987. A cDNA clone from the Duchenne/Becker muscular dystrophy gene. **Nature 328**: 434-437.

Burton, R.G. 1928. Black tigers. **Field 152**: 656-657.

Castle W.E. 1919. Siamese, an albinistic color variation in cats. **Am. Nat. 53**: 265-268.

Castle W.E. & Allen G.M. 1903. The heredity of albinism. **Proc. Amer. Acad. Arts Sci. 38**: 603-622.

Castle W.E. & Wright S. 1916. **Studies of Inheritance in Guinea-Pigs and Rats.** Carnegie Inst. Wash. Publ. No. 241.

Coleman L.E. 1974. Melanistic phases of Felidae in captivity. Preliminary survey results. **Carnivore Genet. Newsl. 2**: 209-211.

Collier G.E. & O'Brien S.J. 1985. A molecular phylogeny of the Felidae: immunological distance. **Evolution 39**: 473-487.

Cone R.D., Lu D., Koppula S., Vage D.I., Klungland H., Boston B., Chen W., Orth D.N., Pouton C., Kesterson R.A. 1996. The melanocortin receptors: agonists, antagonists, and the hormonal control of pigmentation. **Recent Prog Horm Res. 51**: 287-317.

Cott H.B. 1940. **Adaptive Coloration in Animals**. Methuen, London.

Cottingham R.W. Jr., Idury R.M., Schäffer A.A. 1993. Faster Sequential Genetic Linkage Computations. **Am. J. Human Genet. 53**: 252-263.

Darwin C. 1859. **The Origin of Species**, re-printed in 1993 (paperback edition), Random House, New York.

Darwin C. 1883. **The Variation of Animals and Plants under Domestication**, 2nd ed., re-printed in 1998 (paperback edition). Johns Hopkins Univ. Press, Baltimore.

Dean M., Carrington M., O'Brien S.J. 2002. Balanced polymorphism selected by genetic versus infectious human disease. **Annu. Rev. Genomics Hum. Genet. 3**: 263-292.

Dittrich L. 1979. Die vererbung des melanismus beim jaguar (*Panthera onca*). **Zool. Garten 49**: 417-428.

Doncaster L. 1904. On the inheritance of tortoiseshell and related colors in cats. **Proc. Camb. Phil. Soc. 13**: 35-38.

Dubchak I., Brudno M., Loots G.G., Mayor C., Pachter L., Rubin E.M., Frazer K.A. 2000. Active Conservation of Noncoding Sequences Revealed by 3-way Species Comparisons. **Genome Res. 10**: 1304-1306.

Duhl D.M., Vrieling H., Miller K.A., Wolff G.L., Barsh G.S. 1994. Neomorphic agouti mutations in obese yellow mice. **Nat. Genet. 8**: 59-65

Dunstone N., Durbin L., Wyllie I., Rose S., Acosta G. 1998. Ecology of the kodkod in Laguna San Rafael National Park, Chile. **Cat News 29**: 18-20.

Echols H. 2001. **Operators and Promoters: the Story of Molecular Biology and its Creators**. Univ. California Press, Berkeley.

Eizirik, E., Bonatto S.L., Johnson W.E., Crawshaw P.G. Jr., Vié J.C., Brousset D.M., O'Brien S.J., Salzano F.M. 1998. Phylogeographic patterns and mitochondrial DNA control region evolution in two Neotropical cats (Mammalia, Felidae). **J. Mol. Evol. 47**: 613-624.

Eizirik E., Kim J.H., Menotti-Raymond M., Crawshaw P.G. Jr., O'Brien S.J., Johnson W.E. 2001a. Phylogeography, population history and conservation genetics of jaguars (*Panthera onca*, Mammalia, Felidae). **Mol Ecol. 10**: 65-79.

Eizirik E., Murphy W.J., O'Brien S.J. 2001b. Molecular dating and biogeography of the early placental mammal radiation. **J. Hered. 92**: 212-219.

Eizirik E., Murphy W.J., Springer M.S., O'Brien S.J. (in press). Molecular Phylogeny and Dating of Early Primate Divergences. In: Ross C. & Kay R. (eds.) **Anthropoid Origins: New Visions**. Kluwer Academic/Plenum, New York.

Everts R.E., Rothuizen J., van Oost B.A. 2000. Identification of a premature stop codon in the melanocyte-stimulating hormone receptor gene (*MC1R*) in Labrador and Golden retrievers with yellow coat colour. **Anim Genet. 31**: 194-199.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Res. 8**: 175-185.

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

Fleischman R.A. 1993. From white spots to stem cells: the role of the Kit receptor in mammalian development. **Tr. Genet. 9**: 285-290.

Fisher R.A. 1930. **The Genetical Theory of Natural Selection** [complete variorum edition, published in 1999]. Oxford Univ. Pres, Oxford..

Fu Y.-X. 1996. New statistical tests of neutrality for DNA samples from a population. Genetics **143**: 557-570.

Fu Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. **Genetics 147**: 915-925.

Fu Y.-X. & Li W.-H. 1993. Statistical tests of neutrality of mutations. **Genetics 133**: 693-709.

Galton F. 1889. **Natural Inheritance**. MacMillan, London.

Garcia-Perea R. 1994. The pampas cat group (Genus *Lynchailurus* Severtzov, 1858) (Carnivora: Felidae), a systematic and biogeographic review. **American Museum Novitates 3096**: 1-36.

Gitschier J., Drayna D., Tuddenham E.G., White R.L., Lawn R.M.. 1985. Genetic mapping and diagnosis of haemophilia A achieved through a BclI polymorphism in the factor VIII gene. **Nature 314**: 738-740.

Golding G.B. & Dean A.M. 1998. The structural basis of molecular adaptation. **Mol Biol Evol. 15**: 355-369.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. **Genome Res. 8**: 195-202.

Gumucio D.L., Shelton D.A., Zhu W., Millinoff D., Gray T., Bock J.H., Slightom J.L., Goodman M. 1996. Evolutionary strategies for the elucidation of *cis* and *trans* factors that regulate the developmental switching programs of the beta-like globin genes. **Mol. Phylogenet. Evol. 5**: 18-32.

Gunn T.M., Miller K.A., He L., Hyman R.W., Davis R.W., Azarani A., Schlossman S.F., Duke-Cohan J.S., Barsh G.S. 1999. The mouse mahogany locus encodes a transmembrane form of human attractin. **Nature 398**: 152-156.

Gunn T.M., Inui T., Kitada K., Ito S., Wakamatsu K., He L., Bouley D.M., Serikawa T., Barsh G.S. 2001. Molecular and phenotypic analysis of Attractin mutant mice. **Genetics 158**: 1683-1695.

Haldane J.B.S. 1927. The comparative genetics of colour in rodents and carnivora. **Biol. Rev. and Biol. Proc. Cambridge Phil. Soc. 2**: 201-212.

Haldane J.B.S., Sprunt A.D., Haldane N.M. 1915. Reduplication in mice. **J. Genet. 5**: 133-135.

Harding R.M., Healy E., Ray A.J., Ellis N.S., Flanagan N., Todd C., Dixon C., Sajantila A., Jackson I.J., Birch-Machin M.A., Rees J.L. 2000. Evidence for variable selective pressures at *MC1R*. **Am. J. Hum. Genet. 66**: 1351-1361.

Hartl D.L. & Clark A.G. 1997. **Principles of Population Genetics**. 3rd ed. Sinauer, Sunderland.

He L., Gunn T.M., Bouley D.M., Lu X.Y., Watson S.J., Schlossman S.F., Duke-Cohan J.S., Barsh G.S. 2001. A biochemical function for attractin in agouti-induced pigmentation and obesity. **Nat. Genet. 27**: 40-47.

Herrington S.J. 1986. **Phylogenetic Relationships of the Wild Cats of the World.** Ph.D. Dissertation, Univ. Kansas.

Higaki Y., Schullery D., Kawata Y., Shnyreva M., Abrass C., Bomsztyk K. 2002. Synergistic activation of the rat laminin gamma1 chain promoter by the gut-enriched Kruppel-like factor (GKLF/KLF4) and Sp1. **Nucleic Acids Res 30**: 2270-2279.

Hill A.V.S., Elvin J., Willis A.C., Aidoo M., Allsopp C.E.M., Gotch F.M., Gao X.M., Takiguchi M., Greenwood B.M., Townsend A.R.M., McMichael A.J., Whittle H.C. 1992. Molecular analysis of the association of HLA-B53 and resistance to severe malaria. **Nature 360**: 434-439.

Hough R.B., Avivi A., Davis J., Joel A., Nevo E., Piatigorsky J. 2002. Adaptive evolution of small heat shock protein/alpha B-crystallin promoter activity of the blind subterranean mole rat, *Spalax ehrenbergi*. **Proc. Natl. Acad. Sci. USA 99**: 8145-8150.

Hromas R., Davis B., Rauscher F.J. 3rd, Klemsz M., Tenen D., Hoffman S., Xu D., Morris J.F. 1996 . Hematopoietic transcriptional regulation by the myeloid zinc finger gene, MZF-1. **Curr. Top. Microbiol. Immunol. 211**: 159-164.

Hudson R.R., Kreitman M., Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. **Genetics 116**:153-159.

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

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

Hustad C.M., Perry W.L., Siracusa L.D., Rasberry C., Cobb L., Cattanach B.M., Kovatch R., Copeland N.G., Jenkins N.A. 1995. Molecular genetic characterization of six recessive viable alleles of the mouse agouti locus. **Genetics 140**: 255-265.

Jackson, I. J. 1993. Colour-coded switches. **Nature 362**: 587-588.

Jackson, I. J. 1994. Molecular and developmental genetics of mouse coat color. **Annu. Rev. Genet. 28**: 189-217.

Joerg H., Fries H. R., Meijerink E., Stranzinger G. F. 1996. Red coat color in Holstein cattle is associated with a deletion in the *MSHR* gene. **Mamm. Genome 7**: 317-318.

Johnson W. E. & O'Brien S. J. 1997. Phylogenetic reconstruction of the Felidae using *16S* rRNA and *NADH-5* mitochondrial genes. **J. Mol. Evol. 44**: S98-S116.

Johnson W.E., Culver M., Iriarte A., Eizirik E., Seymour K., O'Brien S.J. 1998. Tracking the evolution of the elusive Andean mountain cat with mitochondrial DNA sequences. **J. Hered. 89**: 227-232.

Johnson W.E., Slattery J.P., Eizirik E., Kim J.H., Raymond M.M., Bonacic C., Cambre R., Crawshaw P., Nunes A., Seuanez H.N., Moreira M.A., Seymour K.L., Simon F., Swanson W., O'Brien S.J. 1999. Disparate phylogeographic patterns of molecular genetic variation in four closely related South American small cat species. **Molecular Ecology 8 (Suppl)**: S79-94.

Junge, C. 1975. Chillan Viejo, Chile. **Int. Zoo News 21**: 38.

Kanetsky P.A., Swoyer J., Panossian S., Holmes R., Guerry D., Rebbeck T.R. 2002. A polymorphism in the agouti signaling protein gene is associated with human pigmentation. **Am. J. Hum. Genet. 70**: 770-775.

Kaufman C.K., Sinha S., Bolotin D., Fan J., Fuchs E. 2002. Dissection of a complex enhancer element: Maintenance of Keratinocyte specificity but loss of differentiation specificity. **Mol. Cell. Biol. 22**: 4293-4308.

Kiefer L.L., Ittoop O.R., Bunce K., Truesdale A.T., Willard D.H., Nichols J.S., Blanchard S.G., Mountjoy K., Chen W.J., Wilkison W.O. 1997. Mutations in the carboxyl terminus of the agouti protein decrease agouti inhibition of ligand binding to the melanocortin receptors. **Biochemistry 36**: 2084-2090.

Kiefer L.L., Veal J.M., Mountjoy K.G., Wilkison W.O. 1998. Melanocortin receptor binding determinants in the agouti protein. **Biochemistry 37**: 991-997.

Kijas, J.M.H., Wales R., Törnstein A., Chardon P., Moller M., Andersson L. 1998. Melanocortin receptor 1 (*MC1R*) mutations and coat color in pigs. **Genetics 150**: 1177-1185.

Kijas J.M., Moller M., Plastow G., Andersson L. 2001. A frameshift mutation in MC1R and a high frequency of somatic reversions cause black spotting in pigs. **Genetics 158**: 779-785.

Kim J.-H., Eizirik E., O'Brien S.J., Johnson W.E. 2001. Structure and patterns of sequence variation in the mitochondrial DNA control region of the great cats. **Mitochondrion 1**: 279-292.

Kimura M. 1983. **The Neutral Theory of Molecular Evolution**. Cambridge Univ. press, Cambridge.

Kitchener A. 1991. **The Natural History of Wild Cats.** Cornell Univ. Press, Ithaca.

Klungland H, Våge DI, Gomez-Raya L, Adalsteinsson S, Lien S. 1995. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. **Mamm. Genome 6**: 636-639.

Klungland H., Roed K.H., Nesbo C.L., Jakobsen K.S., Våge D.I. 1999. The melanocyte-stimulating hormone receptor (MC1-R) gene as a tool in evolutionary studies of artiodactyles. **Hereditas 131**: 39-46.

Kornegay J.R., Schilling J.W., Wilson A.C. 1994. Molecular adaptation of a leaf-eating bird: stomach lysozyme of the hoatzin. **Mol. Biol. Evol. 11**: 921-928.

Kreitman M. 2000. Methods to detect selection in populations with applications to the human. **Annu. Rev. Genomics Hum. Genet. 1**: 539-559.

Kumar S., Tamura K., Jakobsen I.B., Nei M. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. **Bioinformatics 17**: 1244-1245.

Kuramoto T., Nomoto T., Sugimura T., Ushijima T. 2001a. Cloning of the rat agouti gene and identification of the rat nonagouti mutation. **Mamm. Genome 12**: 469-471.

Kuramoto T., Kitada K., Inui T., Sasaki Y., Ito K., Hase T., Kawagachi S., Ogawa Y., Nakao K., Barsh G.S., Nagao M., Ushijima T., Serikawa T. 2001b. Attractin / mahogany / zitter plays a critical role in myelination of the central nervous system. **Proc. Natl. Acad. Sci. USA 98**: 559-564.

Kwon H.Y., Bultman S.J., Löffler C., Chen W.J., Furdon P.J., Powell J.G., Usala A.L., Wilkison W., Hansmann I., Woychik R.P. 1994. Molecular structure and chromosomal mapping of the human homolog of the agouti gene. **Proc. Natl. Acad. Sci. USA 91**: 9760-9764.

Lamotte M. 1942. La fauna mammalogique du Mont Nimba (Haute Guinée). **Mammalia 6**: 114-119.

Lathrop G.M., Lalouel J.M., Julier C., Ott J. 1984. Strategies for multilocus linkage analysis in humans. **Proc. Natl. Acad. Sci. USA 81**: 3443-3446.

Leeb T., Doppe A., Kriegesmann B., Brenig B. 2000. Genomic structure and nucleotide polymorphisms of the porcine agouti signaling protein gene(*ASIP*). **Anim. Genet. 31**: 335-336.

Lewis E.B. 1978. A gene complex controlling segmentation in *Drosophila*. **Nature 276**: 565-570.

Li W.-H. 1993.Unbiased estimation of the rates of synonymous and nonsynonymous substitution. **J. Mol. Evol. 36**: 96-99.

Li W.-H. 1997. **Molecular Evolution.** Sinauer, Sunderland.

Liberles D.A. & Wayne M.L. 2002. Tracking adaptive evolutionary events in genomic sequences. **Genome Biol. 3**: reviews1018.

Lönnberg, E. 1898. Über eine melanistische Varietät vom Serval nebst Bemerkungen über andere melanistische Säugetiere. **Zool. Jb. 10**: 569-595.

Loots G. G., Ovcharenko I., Pachter L., Dubchak I., Rubin E. M. 2002. rVista for comparative sequence-based discovery of functional transcription factor binding sites. **Genome Res. 12**: 832-839.

Lopez-Rodriguez C., Botella L., Corbi A.L. 1997. CCAAT-enhancer-binding proteins (C/EBP) regulate the tissue specific activity of the CD11c integrin gene promoter through functional interactions with Sp1 proteins. **J. Biol. Chem. 272**: 29120-29126.

Lu D., Willard D., Patel I.R., Kadwell S., Overton L., Kost T., Luther M., Chen W., Woychik R.P., Wilkison W.O., Cone R.D. 1994. Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor. **Nature 371**: 799-802.

Lu D., Haskell-Luevano C., Vage D.I., Cone R.D. 1998. Functional variants of the *MSH receptor* (*MC1-R*), *agouti*, and their effects on mammalian pigmentation. Pp. 231-259 in: Spiegel A.M. **Contemporary Endocrinology: G Proteins, Receptors, and Disease**. Humana Press, Totowa, NJ, USA.

Lyon, M.F. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). **Nature 190**: 372-373.

Majerus M.E.N. 1998. **Melanism: Evolution in Action**. Oxford Univ. Press, Oxford.

Makalowski W. & Boguski M.S. 1998. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. **Proc. Natl. Acad. Sci. USA 95**: 9407-9412.

Makova K.D., Ramsay M., Jenkins T., Li W.-H. 2001. Human DNA sequence variation in a 6.6-kb region containing the melanocortin 1 receptor promoter. **Genetics 158**: 1253-1268.

Manga, P., Kromberg J.G.R., Box N.F., Sturm R.A., Jenkins T., Ramsay M. 1997. Rufous oculocutaneous albinism in southern African blacks is caused by mutations in the *Tyrp1* gene. **Am. J. Hum. Genet. 61**: 1095-1101.

Marklund L., Moller M.J., Sandberg K., Andersson L. 1996. A missense mutation in the gene for melanocyte-stimulating hormone receptor (MC1R) is associated with the chestnut coat color in horses. **Mamm. Genome 7**: 895-899.

Marklund S., Moller M., Sandberg K., Andersson L. 1999. Close association between sequence polymorphism in the *KIT* gene and the roan coat color in horses. **Mamm. Genome 10**: 283-288.

Mattern M.Y. & McLennan D.A. 2000. Phylogeny and speciation of felids. **Cladistics 16**: 232-253.

May R.M. 1985. Evolution of pesticide resistance. **Nature 315**: 12-13.

Mayor C., Brudno M., Schwartz J.R., Poliakov A., Rubin E.M., Frazer K.A., Pachter L.S., Dubchak I. 2000. VISTA: Visualizing Global DNA Sequence Alignments of Arbitrary Length. **Bioinformatics 16**: 1046-1047.

Mayr E. 1973. The recent historiography of genetics. **J. of the Hist. of Biology 6**: 125-154. (reprinted in: Mayr E. 1976. Evolution and the Diversity of Life: selected essays. Harvard Univ. Press, Cambridge.).

McClintock B. 1950. The origins and behavior of mutable loci in maize. **Proc. Natl. Acad. Sci USA 36**: 344-355 (reprinted in: Peters J.A. 1959. Classic papers in genetics. Prentice-Hall, Englewood Cliffs, NJ, USA).

McDonald J.H. & Kreitman M.1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351: 652-654.

Mendel G. 1865. Experiments in plant hybridisation. English Translation published in: Bateson W. 1913. **Mendel's Principles of Heredity**. Cambridge Univ. Press, Cambridge.

Menotti-Raymond M., David V.A., Lyons L.A., Schäffer A.A., Tomlin J.F., Hutton M.K., O'Brien S.J. 1999. A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). **Genomics 57**: 9-23.

Menotti-Raymond M., David V.A., Chen Z.Q., Menotti K.A., Sun S., Schaffer A.A., Agarwala R., Tomlin J.F., O'Brien S.J., Murphy W.J. (in press) Second Generation Genetic Linkage and Radiation Hybrid Maps of the Domestic Cat (*Felis catus*). **J. Hered.**

Metallinos, D.L., Bowling A.T., Rine J. 1998. A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung Disease. **Mamm. Genome 9**: 426-431.

Michaud E.J., van Vugt M.J., Bultman S.J., Sweet H.O., Davisson M.T., Woychik R.P. 1994. Differential expression of a new dominant agouti allele (A^{iapy}) is correlated with methylation state and is influenced by parental lineage. **Genes Dev. 8**: 1463-1472.

Millar S.E., Miller M.W., Stevens M.E., Barsh G.S. 1995. Expression and transgenic studies of the mouse *agouti* gene provide insight into the mechanisms by which mammalian coat color patterns are generated. **Development 121**: 3223-3232.

Miller M.W., Duhl D.M., Vrieling H., Cordes S.P., Ollmann M.M., Winkes B.M., Barsh G.S. 1993. Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation. **Genes Dev. 7**: 454-467.

Miller M.W., Duhl D.M., Winkes B.M., Arredondo-Vega F., Saxon P.J., Wolff G.L.,Epstein C.J., Hershfield M.S., Barsh G.S. 1994. The mouse lethal nonagouti (a^x) mutation deletes the S-adenosylhomocysteine hydrolase (*Ahcy*) gene. **EMBO J. 13**: 1806-1816.

Miltenberger R.J., Wakamatsu K., Ito S., Woychik R.P., Russell L.B., Michaud E.J. 2002. Molecular and phenotypic analysis of 25 recessive, homozygous-viable alleles at the mouse agouti locus. **Genetics 160**: 659-674.

Moro O., Ideta R., Ifuku O. 1999. Characterization of the promoter region of the human melanocortin-1 receptor (*MC1R*) gene. **Biochem. Biophys. Res. Commun. 262**: 452-460.

Mountjoy K.G, Robbins L.S, Mortrud M.T, Cone R.D. 1992. The cloning of a family of genes that encode the melanocortin receptors. **Science. 257**: 1248-1251.

Muller H.J. 1922. Variation due to change in the individual gene. **Am. Nat. 56**: 32-50 (reprinted in: Peters J.A. 1959. Classic papers in genetics. Prentice-Hall, Englewood Cliffs, NJ, USA).

Mullis K.B. & Faloona F. 1987. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. **Meth. Enzymol. 155**: 335-350.

Murphy W.J., Sun S., Chen Z., Yuhki N., Hirschmann D., Menotti-Raymond M., O'Brien S.J. 2000. A radiation hybrid map of the cat genome: implications for comparative mapping. **Genome Res. 10**: 691-702.

Murphy W.J., Eizirik E., O'Brien S.J., Madsen O., Scally M., Douady C., Teeling E., Ryder O.A., Stanhope M., de Jong W.W., Springer M.S. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. **Science 294**: 2348-2351.

Nagle D.L., McGrail S.H., Vitale J., Woolf E.A., Dussault B.J. Jr, DiRocco L., Holmgren L., Montagno J., Bork P., Huszar D., Fairchild-Huntress V., Ge P., Keilty J., Ebeling C., Baldini L., Gilchrist J., Burn P., Carlson G.A., Moore K.J. 1999. The mahogany protein is a receptor involved in suppression of obesity. **Nature 398**: 148-152.

Nei M. & Kumar S. 2000. **Molecular Evolution and Phylogenetics**. Oxford Univ. Press, New York.

Newton J. M., Wilkie A.L., He L., Jordan S. A., Metallinos D.L., Holmes N.G., Jackson I.J., Barsh G.S. 2000. Melanocortin 1 receptor variation in the domestic dog. **Mamm. Genome 11**: 24-30.

Ng W.L., Schummer M., Cirisano F.D., Baldwin R.L., Karlan B.Y., Hood L. 1996. High-throughput plasmid mini preparations facilitated by micro-mixing. **Nucleic Acids Res. 24**: 5045-5047.

Nowak R. M. 1999. **Walker's Mammals of the World**, 6th ed. Johns Hopkins Univ. Press, Baltimore.

Nowell K. & Jackson P. 1996. **Wild Cats: Status Survey and Conservation Action Plan**. IUCN/SSC Cat Specialist Group, Gland / Cambridge.

O'Brien S.J., Haskins M.E., Winkler C.A., Nash W.G., Patterson D.F. 1986. Chromosomal mapping of beta-globin and albino loci in the domestic cat: a conserved mammalian chromosome group. **J. Hered. 77**: 374-378.

O'Connell J.R., Weeks D.E. 1998. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. **Am. J. Hum. Genet. 63**: 259-266.

Ortolani A. & Caro T. M. 1996. The adaptive significance of color patterns in carnivores: phylogenetic test of classic hypotheses. Pp. 132-188 in Gittleman J.L. (ed.) **Carnivore Behavior, Ecology an Evolution**, vol. 2. Cornell Univ. Press, New York.

Ott J. 1991. **Analysis of Human Genetic Linkage**, 2nd ed. Johns Hopkins Univ. Press, Baltimore.

Pamilo P., Bianchi N.O. 1993. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. **Mol. Biol. Evol. 10**: 271-281.

Pecon-Slattery J. & O'Brien S. J. 1998. Patterns of Y and X chromosome DNA sequence divergence during the Felidae radiation. **Genetics 148**: 1245-1255.

Perry W.L., Nakamura T., Swing D.A., Secrest L., Eagleson B., Hustad C.M., Copeland N.G., Jenkins N.A. 1996. Coupled site-directed mutagenesis / trangenesis identifies important functional domains of the mouse agouti protein. **Genetics 144**: 255-264.

Perry W.L., Hustad C.M., Swing D.A., O'Sullivan T.N., Jenkins N.A., Copeland N.G. 1998. The itchy locus encodes a novel ubiquitin protein ligase that is disrupted in a18H mice. **Nat. Genet. 18**:143-146.

Perutz M.F. 1983. Species adaptation in a protein molecule. **Mol. Biol. Evol. 1**: 1-28.

Pocock R.I. 1917. The classification of existing Felidae. **Ann. Mag. Nat. Hist. 20**: 329-350.

Pocock R.I. 1940. The races of Geoffroy's cat (*Oncifelis geoffroyi*). **Ann. Mag. Nat. Hist. London 11**: 350-355.

Raich N., Clegg C.H., Grofti J., Romeo P.H., Stamatoyannopoulos G. 1995. GATA1 and YY1 are developmental repressors of the human epsilon-globin gene. **EMBO J. 14**: 801-809.

Rana B. K., Hewett-Emmett D., Jin L., Chang B.H.-J., Sambuughin N., Lin M., Watkins S., Bamshad M., Jorde L.B., Ramsay M., Jenkins T., Li W.-H. 1999. High polymorphism at the human melanocortin 1 receptor locus. **Genetics 151**: 1547-1557.

Raymond M., Berticat C., Weill M., Pasteur N., Chevillon C. 2001. Insecticide resistance in the mosquito *Culex pipiens*: what have we learned about adaptation? **Genetica 112-113**: 287-296.

Rees J.L. 2000. The melanocortin 1 receptor (MC1R): more than just red hair. **Pigment Cell Res. 13**: 135-140.

Rieder S., Taourit S., Mariat D., Langlois B., Guerin G. 2001. Mutations in the agouti (*ASIP*), the extension (*MC1R*), and the brown (*TYRP1*) loci and their association to coat color phenotypes in horses (*Equus caballus*). **Mamm. Genome 12**: 450-455.

Ritland K., Newton C., Marshall H.D. 2001. Inheritance and population structure of the white-phased "Kermode" black bear. **Curr. Biol. 11**: 1468-1472.

Robbins L.S., Nadeau J.H., Johnson K.R., Kelly M.A., Roselli-Rehfuss L., Baack E., Mountjoy K.G., Cone R.D. 1993. Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. **Cell 72**: 827-834.

Robinson R. 1959. Genetics of the domestic cat. **Bibliographia Genetica XVIII**: 273-362.

Robinson R. 1969. The breeding of spotted and black leopards. **J. Bombay Nat. Hist. Soc. 66**: 423-429.

Robinson R. 1970a. Homologous mutants in mammalian coat colour variation. **Symp. Zool. Soc. Lond. 26**: 251-269.

Robinson R. 1970b. Inheritance of the black form of the leopard *Panthera pardus*. **Genetica 41**: 190-197.

Robinson R. 1976. Homologous genetic variation in the *Felidae*. 1976. **Genetica 46**: 1-31.

Robinson R. 1991. **Genetics for Cat Breeders**, 3rd ed. Butterworth-Heinemann.

Robson K.J., Chandra T., MacGillivray R.T., Woo S.L. 1982. Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA. **Proc. Natl. Acad. Sci. USA. 79**: 4701-4705.

Rozas J. & Rozas R. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. **Bioinformatics 15**: 174-175.

Sabeti P.C., Reich D.E., Higgins J.M., Levine H.Z., Richter D.J., Schaffner S.F., Gabriel S.B., Platko J.V., Patterson N.J., McDonald G.J., Ackerman H.C., Campbell S.J., Altshuler D., Cooper R., Kwiatkowski D., Ward R., Lander E.S. 2002. Detecting recent positive selection in the human genome from haplotype structure. **Nature 419**: 832-837.

Saitou N. & Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Mol. Biol. Evol. 4**: 406-425.

Salles L.O. 1992. Felid phylogenetics: extant taxa and skull morphology (Felidae, Aeluroidea). **American Museum Novitates 3047**: 1-67.

Sambrook J., Fritsch E.F., Maniatis T. 1989. **Molecular Cloning: a Laboratory Manual**, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Searle A.G. 1968. **Comparative Genetics of Coat Colour in Mammals**. Logos Press, London.

Sharp P.M., Tuohy T.M.F., Mosurski K.R. 1986. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. **Nucleic Acids Res. 14**: 5125-5143.

Shyue S.K., Hewett-Emmett D., Sperling H.G., Hunt D.M., Bowmaker J.K., Mollon J.D., Li W.-H. 1995. Adaptive evolution of color vision genes in higher primates. **Science 269**: 1265-1267.

Sinha S., Degenstein L., Copenhaver C., Fuchs E. 2000. Defining the regulatory factors required for epidermal gene expression. **Mol. Cell. Biol. 20**: 2543-2555.

Silvers W.K. 1979. **The Coat Colors of Mice: a Model for Mammalian Gene Action and Interaction**. Springer-Verlag, New York.

Siracusa L.D. 1994. The agouti gene: turned on to yellow. **Tr. Genet. 10**: 423-428.

Slatkin M. & Rannala B. 2000. Estimating allele age. **Annu. Rev. Genomics Hum. Genet. 1**: 225-249.

Smith A.G., Box N.F., Marks L.H., Chen W., Smit D.J., Wyeth J.R., Huttley G.A., Easteal S., Sturm R.A. 2001. The human melanocortin-1 receptor locus: analysis of transcription unit, locus polymorphism and haplotype evolution. **Gene 281**: 81-94.

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

Sokal R.R. & Rohlf F.J. 1995. **Biometry**, 3rd. ed. Freeman, New York.

Stephens J.C., Reich D.E., Goldstein D.B., Shin H.D., Smith M.W., Carrington M., Winkler C., Huttley G.A., Allikmets R., Schriml L., Gerrard B., Malasky M., Ramos M.D., Morlot S., Tzetis M., Oddoux C., di Giovine F.S., Nasioulas G., Chandler D., Aseev M., Hanson M., Kalaydjieva L., Glavac D., Gasparini P., Kanavakis E., Claustres M., Kambouris M., Ostrer H., Duff G., Baranov V., Sibul H., Metspalu A., Goldman D., Martin N., Duffy D., Schmidtke J., Estivill X., O'Brien S.J., Dean M. 1998. Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. **Am. J. Hum. Genet. 62**: 1507-1515.

Stewart C.B. & Wilson A.C. 1987. Sequence convergence and functional adaptation of stomach lysozymes from foregut fermenters. **Cold Spring Harb. Symp. Quant. Biol. 52**: 891-899.

Struhl K. 1983. The new yeast genetics. **Nature 305**: 391-397.

Sturm R.A., Teasdale R.D., Box N.F. 2001. Human pigmentation genes: identification, structure and consequences of polymorphic variation. **Gene 277**: 49-62.

Sunquist M. & Sanderson J. 1998. Ecology and behavior of the kodkod in a highly-fragmented, human-dominated landscape. **Cat News 28**:17-18.

Swanson W.J. & Vacquier V.D. 2002. The rapid evolution of reproductive proteins. **Nat. Rev. Genet. 3**: 137-144.

Sweetser M.T., Hoey T., Sun Y.L., Weaver W.M., Price G.A., Wilson C.B. 1998. The roles of nuclear factor of activated T cells and ying-yang 1 in activation-induced expression of the interferon-gamma promoter in T cells. **J. Biol. Chem. 273**: 34775-34783

Swofford D.L. 1998. **PAUP*: 'Phylogenetic Analysis Using Parsimony and Other Methods'**, Computer Program. Sinauer, Sunderland.

Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, Jones RT. 1988. Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. **J. Mol. Biol. 203**: 439-455.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. **Genetics 105**: 437-460.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics 123**: 585-595.

Takeuchi S., Suzuki S., Hirose S., Yabuuchi M., Sato C., Yamamoto H., Takahashi S. 1996. Molecular cloning and sequence analysis of the chick melanocortin 1-receptor gene. **Biochim. Biophys. Acta 1306**: 122-126.

Takeuchi S., Suzuki H., Yabuuchi M., Takahashi S. 1996. A possible involvement of melanocortin 1-receptor in regulating feather color pigmentation in the chicken. **Biochim. Biophys. Acta 1308**: 164-168.

Tan C.P., McKee K.K., Weinberg D.H., MacNeil T., Palyha O.C., Feighner S.D., Hreniuk D.L., Van Der Ploeg L.H., MacNeil D.J., Howard A.D. 1999. Molecular analysis of a new splice variant of the human melanocortin-1 receptor. **FEBS Lett. 451**: 137-141.

Theron E., Hawkins K., Bermingham E., Ricklefs R.E., Mundy N.I. 2001. The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. **Curr. Biol. 11**: 550-557.

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

Tishkoff S.A., Varkonyi R., Cahinhinan N., Abbes S., Argyropoulos G., Destro-Bisol G., Drousiotou A., Dangerfield B., Lefranc G., Loiselet J., Piro A., Stoneking M., Tagarelli A., Tagarelli G., Touma E.H., Williams S.M., Clark A.G. 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. **Science 293**: 455-462.

Ulmer F.A. Jr. 1941. Melanism in the Felidae, with special reference to the genus *Lynx*. **J. Mamm. 22**: 285-288.

Våge D.I., Lu D., Klungland H., Lien S., Adalsteinsson S., Cone R.D. 1997. A non-epistatic interaction of agouti and extension in the fox, *Vulpes vulpes*. **Nat. Genet. 15**: 311-315.

Våge D.I., Klungland H., Lu D., Cone R.D. 1999. Molecular and pharmacological characterization of dominant black coat color in sheep. **Mamm. Genome 10**: 39-43.

Valverde P., Healy E., Jackson I., Rees J.L., Thody A. J. 1995. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. **Nat. Genet. 11**: 328-330.

Verrelli B.C., McDonald J.H., Argyropoulos G., Destro-Bisol G., Froment A., Drousiotou A., Lefranc G., Helal A.N., Loiselet J., Tishkoff S.A. 2002. Evidence for balancing selection from nucleotide sequence analyses of human G6PD. **Am. J. Hum. Genet. 71**: 1112-1128.

Virador V.M., Santis C., Furumura M., Kalbacher H., Hearing V.J. 2000. Bioactive motifs of agouti signal protein. **Exp. Cell. Res. 259**: 54-63.

Voisey J., Box N.F., van Daal A. 2001. A polymorphism study of the human *Agouti* gene and its association with *MC1R*. **Pigment Cell Res. 14**: 264-267.

Voisey J. & van Daal A. 2002. *Agouti*: from mouse to man, from skin to fat. **Pigment Cell Res. 15**: 10-18.

Vrieling H., Duhl D.M.J., Millar S.E., Miller K.A., Barsh G.S. 1994. Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. **Proc. Natl. Acad. Sci. USA 91**: 5667-5671.

Weigel I. 1961. Das Fellmuster der wildlebenden Katzenarten und der Hauskatze in vergleichender und stammesgeschichtlicher Hinsicht. **Säugetierkdl. Mitt. 9 (Sonderheft)**: 1-120.

Whelan S., Lio P., Goldman N. 2001. Molecular phylogenetics: state-of-the-art methods for looking into the past. **Tr. Genet. 17**: 262-272.

Whiting P.W. 1918. Inheritance of coat colour in cats. **J. Exp. Zool. 25**: 539-570.

Wilson B.D., Ollmann M.M., Kang L., Stoffel M., Bell G.I., Barsh G.S. 1995. Structure and function of *ASP*, the human homolog of the mouse *agouti* gene. **Hum. Mol. Genet. 4**: 223-230.

Wozencraft W.C. 1989. Classification of the recent Carnivora. Pp. 569-593 in: Gittleman J.L. (ed.). **Carnivore Behavior, Ecology and Evolution**. Cornell University Press, New York.

Wozencraft W.C. 1993. Order Carnivora, Family Felidae. Pp. 288-300 in: Wilson D.E. & Reeder D.M. (eds.). **Mammal Species of the World: a Taxonomic and Geographic Reference**. 2nd. ed. Smithsonian Institution Press / American Society of Mammalogists, Washington.

Wright S. 1917. Color inheritance in mammals. **J. Hered. 8**: 224-235.

Wright S. 1918. Color inheritance in mammals - X. The cat. **J. Hered. 9**: 139-144.

Wright S. 1968. **Evolution and the Genetics of Populations, vol. 1: Genetics and Biometric Foundations**. Univ. Chicago Press, Chicago.

Wright S. 1977. **Evolution and the Genetics of Populations, vol. 3: Experimental Results and Evolutionary deductions**. Univ. Chicago Press, Chicago.

Wright S. 1978. **Evolution and the Genetics of Populations, vol. 4: Variability within and among Natural Populations**. Univ. Chicago Press, Chicago.

Yang Z. & Bielawski J.P. 2000. Statistical methods for detecting molecular adaptation. **Tr. Ecol. Evol. 15**: 496-503.

Yokoyama S. 1997. Molecular genetic basis of adaptive selection: examples from color vision in vertebrates. **Annu. Rev. Genet. 31**: 315-336.

Yokoyama S. & Yokoyama R. 1989. Molecular evolution of human visual pigment genes. **Mol. Biol. Evol. 6**: 186-197.

Young, S.P. & Goldman E.A. 1946. **The Puma, Mysterious American Cat**. Stockpole, Harrisburg.

Yuhki N., Beck T., Stephens R.M., Nishigaki Y., Newmann K., O'Brien S.J. (in press) Comparative genome organization of human, murine and feline MHC class II region. **Genome Res.**

Zhang J., Zhang Y.P., Rosenberg H.F. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. **Nat. Genet. 30**: 411-415.

Zhu J., Liu J.S., Lawrence C.E. 1998. Bayesian adaptive sequence alignment algorithms. **Bioinformatics 14**: 25-39.