

Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication

Michael J. Montague^{a,1}, Gang Li^{b,1}, Barbara Gandolfi^c, Razib Khan^d, Bronwen L. Aken^e, Steven M. J. Searle^e, Patrick Minx^a, LaDeana W. Hillier^a, Daniel C. Koboldt^a, Brian W. Davis^b, Carlos A. Driscoll^f, Christina S. Barr^f, Kevin Blackstone^g, Javier Quilez^g, Belen Lorente-Galdos^g, Tomas Marques-Bonet^{g,h}, Can Alkanⁱ, Gregg W. C. Thomas^j, Matthew W. Hahn^j, Marilyn Menotti-Raymond^k, Stephen J. O'Brien^{l,m}, Richard K. Wilson^a, Leslie A. Lyons^{c,2}, William J. Murphy^{b,2}, and Wesley C. Warren^{a,2}

^aThe Genome Institute, Washington University School of Medicine, St. Louis, MO 63108; ^bDepartment of Veterinary Integrative Biosciences, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843; ^cDepartment of Veterinary Medicine & Surgery, College of Veterinary Medicine, University of Missouri, Columbia, MO 65201; ^dPopulation Health & Reproduction, School of Veterinary Medicine, University of California, Davis, CA 95616; ^eWellcome Trust Sanger Institute, Hinxton CB10 1SA, United Kingdom; ^fNational Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD 20886; ^gCatalan Institution for Research and Advanced Studies, Institute of Evolutionary Biology, Pompeu Fabra University, 08003 Barcelona, Spain; ^hCentro de Analisis Genomico 08028, Barcelona, Spain; ⁱDepartment of Computer Engineering, Bilkent University, Ankara 06800, Turkey; ^jDepartment of Biology, Indiana University, Bloomington, IN 47405; ^kLaboratory of Genomic Diversity, Center for Cancer Research, Frederick, MD 21702; ^lDobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg 199178, Russia; and ^mOceanographic Center, Nova Southeastern University, Fort Lauderdale, FL 33314

Edited by James E. Womack, Texas A&M University, College Station, TX, and approved October 3, 2014 (received for review June 2, 2014)

Little is known about the genetic changes that distinguish domestic cat populations from their wild progenitors. Here we describe a high-quality domestic cat reference genome assembly and comparative inferences made with other cat breeds, wildcats, and other mammals. Based upon these comparisons, we identified positively selected genes enriched for genes involved in lipid metabolism that underpin adaptations to a hypercarnivorous diet. We also found positive selection signals within genes underlying sensory processes, especially those affecting vision and hearing in the carnivore lineage. We observed an evolutionary tradeoff between functional olfactory and vomeronasal receptor gene repertoires in the cat and dog genomes, with an expansion of the feline chemosensory system for detecting pheromones at the expense of odorant detection. Genomic regions harboring signatures of natural selection that distinguish domestic cats from their wild congeners are enriched in neural crest-related genes associated with behavior and reward in mouse models, as predicted by the domestication syndrome hypothesis. Our description of a previously unidentified allele for the glowing pigmentation pattern found in the Birman breed supports the hypothesis that cat breeds experienced strong selection on specific mutations drawn from random bred populations. Collectively, these findings provide insight into how the process of domestication altered the ancestral wildcat genome and build a resource for future disease mapping and phylogenomic studies across all members of the Felidae.

Felis catus | domestication | genome

The domestic cat (*Felis silvestris catus*) is a popular pet species, with as many as 600 million individuals worldwide (1). Cats and other members of Carnivora last shared a common ancestor with humans ~92 million years ago (2, 3). The cat family Felidae includes ~38 species that are widely distributed across the world, inhabiting diverse ecological niches that have resulted in divergent morphological and behavioral adaptations (4). The earliest archaeological evidence for human coexistence with cats dates to ~9.5 kya in Cyprus and ~5 kya in central China (5, 6), during periods when human populations adopted more agricultural lifestyles. Given their sustained beneficial role surrounding vermin control since the human transition to agriculture, any selective forces acting on cats may have been minimal subsequent to their domestication. Unlike many other domesticated mammals bred for food, herding, hunting, or security, most of the 30–40 cat breeds originated recently, within the past 150 y, largely due to selection for aesthetic rather than functional traits.

Previous studies have assessed breed differentiation (6, 7), phylogenetic origins of the domestic cat (8), and the extent of recent introgression between domestic cats and wildcats (9, 10). However, little is known regarding the impact of the domestication process within the genomes of modern cats and how this compares with genetic changes accompanying selection identified in other domesticated companion animal species. Here we describe, to our knowledge, the first high-quality annotation of the complete

Significance

We present highlights of the first complete domestic cat reference genome, to our knowledge. We provide evolutionary assessments of the feline protein-coding genome, population genetic discoveries surrounding domestication, and a resource of domestic cat genetic variants. These analyses span broadly, from carnivore adaptations for hunting behavior to comparative odorant and chemical detection abilities between cats and dogs. We describe how segregating genetic variation in pigmentation phenotypes has reached fixation within a single breed, and also highlight the genomic differences between domestic cats and wildcats. Specifically, the signatures of selection in the domestic cat genome are linked to genes associated with gene knockout models affecting memory, fear-conditioning behavior, and stimulus-reward learning, and potentially point to the processes by which cats became domesticated.

Author contributions: M.J.M., G.L., B.G., L.A.L., W.J.M., and W.C.W. designed research; M.J.M., G.L., B.G., P.M., L.W.H., D.C.K., B.W.D., C.A.D., C.S.B., K.B., G.W.C.T., M.W.H., M.M.-R., S.J.O., L.A.L., W.J.M., and W.C.W. performed research; M.J.M., G.L., B.G., B.L.A., S.M.J.S., D.C.K., B.W.D., C.A.D., J.Q., B.L.-G., T.M.-B., C.A., G.W.C.T., M.W.H., R.K.W., L.A.L., W.J.M., and W.C.W. contributed new reagents/analytic tools; M.J.M., G.L., B.G., R.K., B.W.D., J.Q., B.L.-G., T.M.-B., C.A., G.W.C.T., M.W.H., L.A.L., W.J.M., and W.C.W. analyzed data; and M.J.M., G.L., B.G., R.K., P.M., D.C.K., B.W.D., C.A.D., C.S.B., K.B., T.M.-B., M.W.H., L.A.L., W.J.M., and W.C.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. GU270865.1, KJ923925–KJ924979, SRX026946, SRX026943, SRX026929, SRX027004, SRX026944, SRX026941, SRX026909, SRX026901, SRX026955, SRX026947, SRX026911, SRX026910, SRX026948, SRX026928, SRX026912, SRX026942, SRX026930, SRX026913, SRX019549, SRX019524, SRX026956, SRX026945, and SRX026960).

¹M.J.M. and G.L. contributed equally to this work.

²To whom correspondence may be addressed. Email: wwarren@genome.wustl.edu, wmurphy@cvm.tamu.edu, or lyonsla@missouri.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1410083111/-DCSupplemental.

domestic cat genome and a comparative genomic analysis including whole-genome sequences from other felids and mammals to identify the molecular footprints of the domestication process within cats.

Results and Discussion

To identify molecular signatures underlying felid phenotypic innovations, we developed a higher-quality reference assembly for the domestic cat genome using whole-genome shotgun sequences (*Materials and Methods* and *SI Materials and Methods*). The assembly (FelCat5) comprises 2.35 gigabases (Gb) assigned to all 18 autosomes and the X chromosome relying on physical and linkage maps (11) with a further 11 megabases (Mb) in unplaced scaffolds. The assembly is represented by an N50 contig length of 20.6 kb and a scaffold N50 of 4.7 Mb, both of which show substantial improvement over previous light-coverage genome survey sequences that included only 60% of the genome (12, 13). The *Felis catus* genome is predicted to contain 19,493 protein-coding genes and 1,855 noncoding RNAs, similar to dog (14). Hundreds of feline traits and disease pathologies (15) offer novel opportunities to explore the genetic basis of simple and complex traits, host susceptibility to infectious diseases, as well as the distinctive genetic changes accompanying the evolution of carnivores from other mammals.

To identify signatures of natural selection along the lineages leading to the domestic cat, we identified rates of evolution using genome-wide analyses of the ratio of divergence at nonsynonymous and synonymous sites (d_N/d_S) (16) (*Materials and Methods* and *SI Materials and Methods*). We used the annotated gene set (19,493 protein-coding genes) to compare unambiguous mammalian gene orthologs shared between cat, tiger, dog, cow, and human ($n = 10,317$). Two-branch and branch-site models (17) collectively identified 467, 331, and 281 genes that were putatively shaped by positive selection in the carnivore, felid, and domestic cat (subfamily Felinae) ancestral lineages, respectively (S1.1–S1.3 in *Dataset S1*). We assessed the potential impact of amino acid changes using TreeSAAP (18) and PROVEAN (19). The majority of identified genes possess substitutions with significant predicted structural or biochemical effects based on one or both tests (Fig. S1 and S1.4 in *Dataset S1*). Although the inferences produced by our methods call for additional functional analyses, we highlight several positively selected genes to illustrate their importance to carnivore and feline biology.

Carnivores are endowed with extremely acute sensory adaptations, allowing them to effectively locate potential prey before being discovered (20). Within carnivores, cats have the broadest hearing range, allowing them to detect both ultrasonic communication by prey as well as their movement (21). We identified six positively selected genes (Fig. 1) that conceivably evolved to increase auditory acuity over a wider range of frequencies in the carnivore ancestor and within Felidae, as mutations within each gene have been associated with autosomal, nonsyndromic deafness or hearing loss (22, 23). Visual acuity is adaptive for hunting and catching prey, especially for crepuscular predators such as the cat and other carnivores. Accordingly, we identified elevated d_N/d_S values for 20 carnivore genes that, when mutated in humans, have well-described roles in a spectrum of visual pathologies (Fig. 1). For example, truncating mutations in human *CHM* cause the progressive disease choroideremia (24), beginning with a loss of night vision and peripheral vision and later a loss of central vision. Many carnivores have excellent night vision (20, 25), and we postulate that the acquisition of selectively advantageous amino acid substitutions within several genes increased visual acuity under low-light conditions. In one interesting dual-role example, *MYO7A* encodes a protein involved in the maintenance of both auditory and visual systems that, when mutated, results in loss of hearing and vision (26).

Cats differ from most other carnivores as a result of being obligately carnivorous. One outcome of this adaptive process is that cats are unable to synthesize certain essential fatty acids, specifically arachidonic acid, due to low Delta-6-desaturase activity (27). This has led to suggestions that cats use an alternate (yet unknown) pathway to generate this essential fatty acid for normal health and reproduction. Furthermore, cats fed a diet rich in

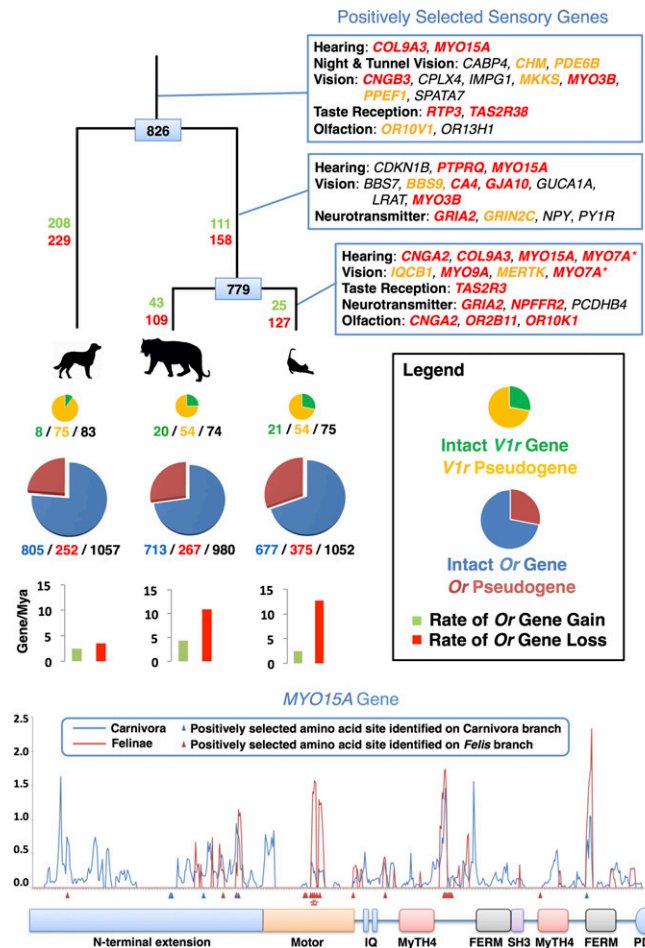


Fig. 1. Dynamic evolution of feline sensory repertoires (*Upper*). The phylogenetic tree depicts relationships scaled to time between dog, tiger, and domestic cat. Positively selected genes are listed (*Top Right*), with lines indicating genes identified on the ancestral branch of Carnivora (*Top*), Felidae (*Middle*), and Felinae (*Bottom*). Genes highlighted in red and orange were identified with significant structural or biochemical effects by two tests or one test, respectively (S1.4 in *Dataset S1*). *MYO7A* (*) expression is associated with hearing and vision. Numbers at each tree node represent the reconstructed ancestral functional olfactory receptor gene (*Or*) repertoire for carnivores and felids. Numbers labeling each branch are estimated *Or* gene gain (green) and loss (red). The pie charts refer to functional and nonfunctional (pseudogenic) vomeronasal (*V1r*; *Top*) and *Or* (*Bottom*) gene repertoires, with circles drawn in proportion to the size of each gene repertoire. *Or* genes are depicted in blue (functional) and red (nonfunctional), and *V1r* genes are depicted in green (functional) and yellow (nonfunctional). Beneath each pie chart are numbers of functional/nonfunctional/total genes identified in the current genome annotations of the three species. Bar graphs depict rates of *Or* gene gain and loss. Location of signatures of positive selection (*Lower*). Several genes encode members of the myosin gene family of mechanochemical proteins, with *MYO15A* notably under selection in all three branches tested. Curved lines represent the estimated d_N/d_S values (*y* axis) calculated in 90-bp sliding windows (step size of 18 bp) along the length of the gene alignment (*x* axis) for dog, cat, and tiger. Colored boxes indicate known functional domains. Arrowheads indicate the location of positively selected amino acid sites based on the results of the branch-site test. Stars indicate deleterious mutations in the domestic cat (*Materials and Methods*). Motifs and domains include the IQ calmodulin-binding motif (IQ); the myosin tail homology 4 domain (MyTH4); the FERM domain (FERM); the SRC homology 3 domain (SH3); and the PDZ domain (PDZ).

saturated and polyunsaturated fatty acids showed no effects on plasma lipid concentrations that in humans are risk factors for coronary heart disease and atherosclerosis (28). These aspects of feline biology are reflected in our positive selection results, where the notable classes of genes overrepresented in the Felinae list

are related to lipid metabolism (S1.5 in Dataset S1). For example, one of the positively selected genes, *ACOX2*, is critical for metabolism of branch-chain fatty acids and has been suggested to regulate triglyceride levels (29), whereas mutations in *PFAFH2* have been associated with risk for coronary heart disease and ischemia (30). The enrichment of genes related to lipid metabolism is likely a signature of adaptation for accommodating the hypercarnivorous diet of felids (31), and mirrors similar signs of selection on lipid metabolic pathways in the genomes of polar bears (32).

Gene duplication and gene loss events often play substantial roles in phenotypic differences between species. To identify protein families that rapidly evolved in the domestic cat, either by contraction or expansion, we examined gene family expansion along an established species tree (33) using tree orthology (34). Two extensive chemosensory gene families, coding for olfactory (*Or*) and vomeronasal (*V1r*) receptors, are responsible for small-molecule detection of odorants and other chemicals for mediating pheromone perception, respectively. Cats rely less on smell to hunt and locate prey in comparison with dogs, which are well-known for their olfactory prowess (35). These observations are confirmed by our analysis of the complete *Or* gene repertoires for cat, tiger, and dog (Fig. 1), illustrating smaller functional repertoires in felids relative to dogs (~700 genes versus >800, respectively). By contrast, the *V1r* gene repertoire is markedly reduced in dogs but expanded in the ancestor of the cat family (8 versus 21 functional genes, respectively), with evidence for species-specific gene loss in different felids (Fig. 1 and Figs. S2 and S3). A growing body of evidence cataloging *Or* gene repertoires in diverse mammals demonstrates common tradeoffs between functional *Or* repertoire size and other sensory systems involved in ecological niche specialization, such as loss of *Or* genes coinciding with gains in trichromatic color vision in primates (36) and chemosensation in platypus (37). These results add further evidence supporting cats' extensive reliance on pheromones for sociochemical communication (38), which is consistent with a genomic tradeoff between functional *Or* and *V1r* repertoires in response to uniquely evolved ecological strategies in the canid and felid lineages (4).

Cats are considered only a semidomesticated species, because many populations are not isolated from wildcats and humans do not control their food supply or breeding (39, 40). We therefore predicted a relatively modest effect of domestication on the cat genome based on recent divergence from and ongoing admixture with wildcats (8–10), a relatively short human cohabitation time

compared with dogs (5, 6), and the lack of clear morphological and behavioral differences from wildcats, with docility, gracility, and pigmentation being the exceptions. To identify genomic regions showing signatures of selection influenced by the domestication process, we used whole-genome analyses of cats from different domestic breeds and wildcats (i.e., other *F. silvestris* subspecies) using pooling methods that control for genetic drift (41). Detecting the genomic regions under putative selection during cat domestication can be complicated by random fixation due to genetic drift during the formation of breeds. We mitigated this effect by combining sequence data from a collection of 22 cats (~58× coverage) from six phylogenetically and geographically dispersed domestic breeds (42) before variant detection and performed selection analyses relative to variants detected within a pool of European (*F. silvestris silvestris*) and Near Eastern (*F. silvestris lybica*) wildcats (~7× coverage; Figs. S4 and S5 and S2.1 in Dataset S2).

After stringent filtering of resequencing data, we aligned sequences to the cat reference genome and identified 8,676,486 and 5,190,430 high-quality single-nucleotide variants (SNVs) among domestic breeds and wildcats, respectively, at a total of 10,975,197 sites (Fig. S3). We next identified 130 regions along cat autosomes with either pooled heterozygosity (H_p) 4 SDs below the mean or divergence (F_{ST}) greater than 4 SDs from the mean (Figs. S4 and S6, SI Materials and Methods, and S2.2 and S2.3 in Dataset S2). After parsing regions of high confidence displaying both low domestic H_p and high F_{ST} , we found 13 genes underlying five chromosomal regions (Fig. 2, Fig. S4, and S2.4 in Dataset S2). Genes within each of these regions play important roles in neural processes, notably pathways related to synaptic circuitry that influence behavior and contextual clues related to reward.

One putative region of selection along chromosome A1 (chrA1) (Fig. 3) is denoted by a pair of protocadherin genes (*PCDHA1* and *PCDHB4*), which establish and maintain specific neuronal connections and have implications for synaptic specificity, serotonergic innervation of the brain, and fear conditioning (43). *PCDHB4* was also identified in the d_N/d_S analyses. A second region, also on chrA1 (Fig. 3), overlaps with a glutamate receptor gene, *GRIA1*. Glutamate receptors are the predominant excitatory neurotransmitter receptors in the mammalian brain and play an important role in the expression of long-term potentiation and memory formation (44). *GRIA1* knockout mice exhibit defects in stimulus-reward learning, notably those related to food rewards (45). Two additional glutamate receptor genes,

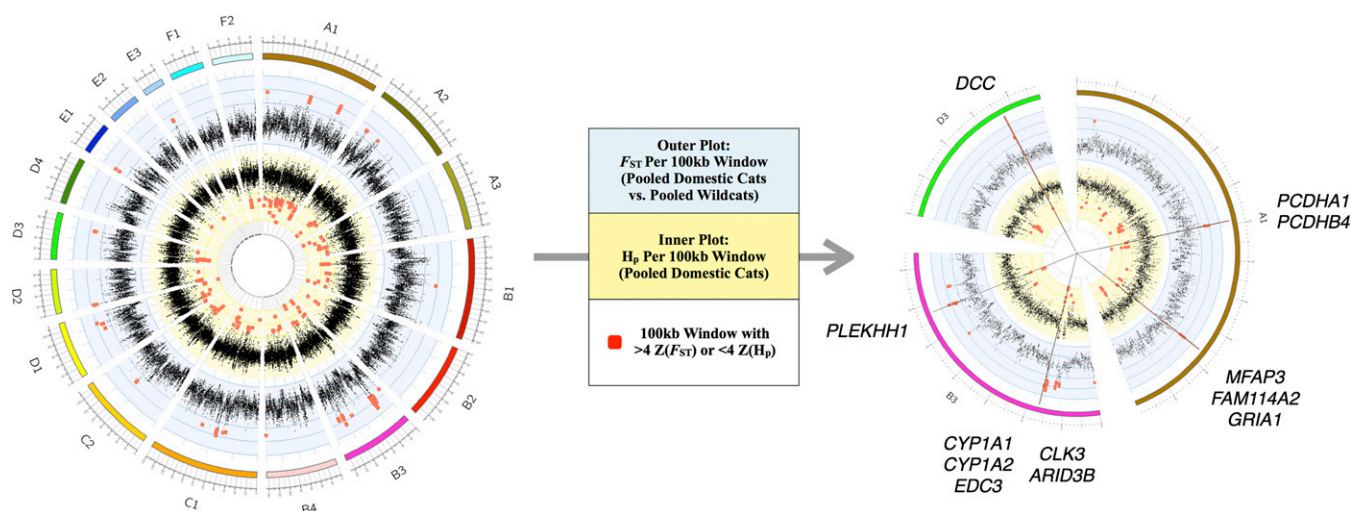


Fig. 2. Sliding window analyses identify five regions of putative selection in the domestic cat genome. Measurements of Z-transformed pooled heterozygosity in cat [inner plot; $Z(H_p)$] and the Z-transformed fixation index between pooled domestic cat and pooled wildcat [outer plot; $Z(F_{ST})$] for autosomal 100-kb windows across all 18 autosomes (Left). Red points indicate windows that passed the threshold for elevated divergence [$>4 Z(F_{ST})$] or low diversity [$<4 Z(H_p)$]. The five regions of putative selection are represented by the straight lines and include contiguous windows that passed both thresholds for elevated divergence and low diversity (Right). These regions, across cat autosomes A1, B3, and D3, contain 12 known genes.

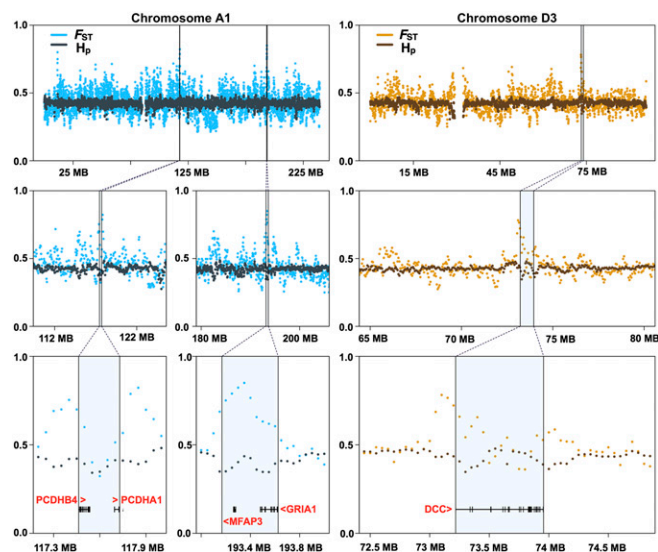


Fig. 3. Comparison between domestic cats and wildcats identifying genes within putative regions of selection in the domestic cat genome that are associated with pathways related to synaptic circuitry and contextual clues related to reward. We identified 130 regions along cat autosomes with either pooled domestic $Z(H_p) < -4$ or $Z(F_{ST}) > 4$, and 5 annotated regions met both criteria. A total of 12 genes was found within these regions, many of which are implicated in neural processes; for instance, genes within regions along chromosomes A1 and D3 are highlighted.

GRIA2 and *NPFFR2*, have elevated d_N/d_S rates within the domestic cat branch of the felid tree (Fig. 1). A third region on chromosome D3 (Fig. 3) encompasses a single gene, *DCC*, encoding the netrin receptor. This gene shows abundant expression in dopaminergic neurons, and behavioral studies of *DCC*-deficient mice show altered dopaminergic system organization, culminating in impaired memory, behavior, and reward responses (46, 47). Two additional regions on chromosome B3 harbor strong signatures of selection (Fig. S7). The first contains three genes, including *ARID3B* (AT rich interactive domain 3B), which plays a critical role in neural crest cell survival (48). The second region contains a single gene, *PLEKHH1*, which encodes a plekstrin homology domain expressed predominantly in human brain. Human genome-wide association studies link variants in *PLEKHH1* with sphingolipid concentrations that, when altered, lead to neurological and psychiatric disease (49).

The genetic signals from this analysis fall in line with the predictions of the domestication syndrome hypothesis (50), which posits that the morphological and physiological traits modified by mammalian domestication are explained by direct and indirect consequences of mild neural crest cell deficits during embryonic development. *ARID3B*, *DCC*, *PLEKHH1*, and protocadherins are all implicated in neural crest cell migration. *ARID3B* is induced in developing mouse embryos during the differentiation of neural crest cells to mature sympathetic ganglia cells (51). *DCC* directly interacts with the Myosin Tail Homology 4 (MyTH4) domain of *MYO10* (myosin X) (52), a gene critical for the migratory ability of neural crest cells. In this way, *DCC* regulates the function of *MYO10* to stimulate the formation and elongation of axons and cranial neural crest cells in developing mouse (53) and frog embryos (54). Like *MYO10*, *PLEKHH1* contains a MyTH4 domain and interacts with the transcription factor *MYC*, a regulator of neural crest cells, to activate transcription of growth-related genes (55). Taken together, we propose that changes in these neural crest-related genes underlie the evolution of tameness during cat domestication, in agreement with analyses of other domesticated genomes (56–58).

We also examined regions of high genetic differentiation between domestic cats and wildcats and observed enrichment in several Wiki and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (S2.5 in Dataset S2), including homologous recombination

and axon guidance. Divergence in regions harboring homologous recombination genes (*RAD51B*, *ZFYVE26*, *BRCA2*) may contribute to the high recombination rate reported for domestic cats relative to other mammals (59). Previous studies have suggested that domestication may select for an increase in recombination as a mechanism to generate diversity (60). Specifically, selection for a recombination driver allele may be favored when it is tightly linked to two or more genes with alleles under selection (61). We hypothesize that the close proximity (<350 kb) of two adjacent genes that regulate homologous recombination (*ZFYVE26* and *RAD51B*, which directly interact with *BRCA2*), two visual genes (*RDH11* and *RDH12*) related to retinol metabolism and dark adaptation (62), and one of our candidate domestication genes, *PLEKHH1* (S2.4 in Dataset S2), represents such a case of adaptive linkage.

Aesthetic qualities such as hair color, texture, and pattern strongly differentiate wildcats from domesticated populations and breeds; however, unlike other domesticated species, less than 30–40 genetically distinct breeds exist (63). At the beginning of the cat fancy ~200 y ago, only five different cat “breeds” were recognized, with each being akin to geographical isolates (64). Long hair and the Siamese coloration of “points” were the only diagnostic breed characteristics. Although most breeds were developed recently, following different breeding strategies and selection pressures, much of the color variation in cats developed during domestication, before breed development, and thus is known as “natural” or “ancient” mutations by cat fanciers.

White-spotting phenotypes are a hallmark of domestication, and in cats can range from a complete lack of pigmentation (white) to intermediate bicolor spotting phenotypes (spotting) to white at only the extremities (gloving). For instance, the Birman breed is characterized by point coloration, long hair, and gloving (Fig. 4). A recent study in several white-spotted cats localized the mutation responsible for the spotting pigmentation phenotype within *KIT* intron 1 (65). The *KIT* gene, located on cat chromosome B1 (66), is primarily involved in melanocyte migration, proliferation, and survival (67). Surprisingly, direct PCR and sequencing excluded the published dominant allele as being associated with the white coloration pattern in Birman (*SI Materials and Methods*). At the same time, whole-genome resequencing data from a pooled sample of Birman cats ($n = 4$; *SI Materials and Methods* and S2.6 in Dataset S2) identified the genomic region containing *KIT* as an outlier exhibiting unusually low genetic diversity (Fig. 4). We therefore resequenced *KIT* exons in a large cohort of domestic cats with various white-spotting phenotypes to genotype candidate SNVs (409 from 21 breeds, 5 Birman outcrosses, and 315 random bred cats). We identified just two adjacent missense mutations that were concordant with the gloving pattern in Birman cats (Fig. 4 and S2.7 in Dataset S2). Genotyping these SNPs in a larger sample including 150 Birman cats and 729 additional cats confirmed that all Birman cats were homozygous for both SNPs and that all first-generation outcrossed Birman cats with no gloving were carriers of the polymorphisms (S2.8 in Dataset S2).

Several lines of evidence indicate that the gloving phenotype in the Birman breed is the result of these two recessive mutations in *KIT*. Both mutations affect the fourth Ig domain of *KIT*, and mutations in this motif near the dimerization site have been shown to result in accelerated ligand dissociation and reduced downstream signal transduction events (68). Interestingly, the frequency of the Birman gloving haplotype in the Ragdoll breed, which shares an extremely similar white-spotting phenotype, was only 12.3%. We suggest that other genetic variants, including the endogenous retrovirus insertion in *KIT* intron 1 (65), likely contribute to the white-spotting phenotype in the Ragdoll breed. The frequency of the Birman gloving haplotype is just 10% in the random nonbreed population, thus illustrating a case where segregating genetic variation in ancestral nonbred populations has reached fixation within Birman cats through strong artificial selection in a remarkably short time frame.

In conclusion, our analyses have identified genetic signatures within feline genomes that match their unique biology and sensory skills. The number of genomic regions with strong signals of selection since cat domestication appears modest compared with those in

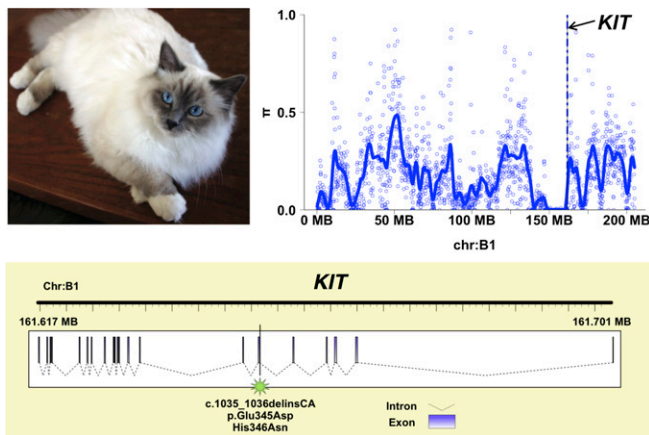


Fig. 4. Genetics of the gloving pigmentation pattern in the Birman cat. The paws of the Birman breed (Top Left) are distinguished by white gloving. The average nucleotide diversity adjacent to *KIT* was low (Top Right). Sequencing experiments identified two adjacent missense mutations within exon 6 of *KIT* that were concordant with the gloving pattern in Birman cats (Bottom).

the domestic dog (41), which is concordant with a more recent domestication history, the absence of strong selection for specific physical characteristics, as well as limited isolation from wild populations. Our results suggest that selection for docility, as a result of becoming accustomed to humans for food rewards, was most likely the major force that altered the first domesticated cat genomes.

Materials and Methods

A female Abyssinian cat, named Cinnamon, served as the DNA source for all sequencing reads (12). From this source we generated $\sim 14\times$ whole-genome shotgun coverage with Sanger and 454 technology. A BAC library was also constructed and all BACs were end-sequenced. We assembled the combined sequences using CABOG software (69) (*SI Materials and Methods*).

We estimated nonsynonymous and synonymous substitution rates using the software PAML 4.0 (17). The following pipeline was used to perform genome-wide selection analyses. (i) We identified 10,317 sets of 1:1:1:1 orthologs from the whole-genome annotations of human (GRCh37), cow (UMD3.1), dog (CanFam3.1), tiger (tiger.genome.org), and domestic cat using the Ensembl pipeline (70). We tested for signatures of natural selection assuming the species tree topology ((cat, tiger), dog), cow, human). (ii) We aligned the translated amino acid sequence of the coding region of each gene using MAFFT (71) with the slow and most accurate parameter settings. A locally developed Perl script pipeline was applied that removed poorly aligned or incorrectly annotated amino acid residues caused by obvious gene annotation errors within the domestic cat and tiger genome assemblies. Aligned amino acid sequences were used for guiding nucleotide-coding sequences by adding insertion gaps and removing poorly aligned regions. (iii) Model testing and likelihood ratio tests (LRTs) were performed using PAML 4.0. Paired models representing different hypotheses consisted of branch tests and branch-site tests (fixed $\omega = 1$ vs. variable ω). For the branch-specific tests, free ratio vs. one-ratio tests were used to identify putatively positively selected genes. These genes were subsequently tested by two-ratio and one-ratio models to identify genes with significant positive selection of one branch versus all other branches (two-branch test). Significance of LRT results used a threshold of $P < 0.05$. We also report the mean synonymous rates along the ancestral felid lineage as well as the tiger, cat, and dog lineages (Fig. S1). We assessed enrichment of gene functional clusters under positive natural selection using WebGestalt (72) (S1.5–S1.7 in Dataset S1). Entrez Gene IDs were input as gene symbols, with the organism of interest set to *Homo sapiens* using the genome as the reference set. Significant Gene Ontology categories (73), Pathway Commons categories

(74), WikiPathways (75), and KEGG Pathways (76) were reported using a hypergeometric test, and the significance level was set at 0.05. We implemented the Benjamini and Hochberg multiple test adjustment (77) to control for false discovery.

Using the whole-genome assembly of domestic cat (FelCat5) as a reference, we mapped Illumina raw sequences from a pool of four wildcat individuals [two European wildcats (*F. s. silvestris*) and two Eastern wildcats (*F. s. lybica*)]. Six additional domestic cat breeds from different worldwide regional populations were sequenced using the Illumina platform (*SI Materials and Methods*). Before sequencing, we pooled samples by breed for the following individuals: Maine Coon ($n = 5$), Norwegian Forest ($n = 4$), Birman ($n = 4$), Japanese Bobtail ($n = 4$), and Turkish Van ($n = 4$). Whole-genome sequencing was also performed on an Egyptian Mau cat ($n = 1$) and on the Abyssinian reference individual ($n = 1$).

We combined the raw reads from the following breed sequencing experiments (described above) before alignment and variant calling: Egyptian Mau, Maine Coon, Norwegian Forest, Birman, Japanese Bobtail, and Turkish Van. The domestic cat pool ($n = 22$) was sequenced to a genome coverage depth of ~ 58 -fold, whereas the wildcat pool was sequenced to a depth of ~ 7 -fold (S2.1 in Dataset S2). Base position differences were called using the convergent outcomes of the software SAMtools (78) and VarScan 2 (79). Parameters included a P value of 0.1, a map quality of 10, and parameters for filtering by false positives. A clustered variant filter was implemented to allow for a maximum of five variant sites in any 500-bp window. Variants were finally filtered using PoPoolation2 (80) to yield a high-confidence set of SNVs ($n = 6,534,957$; filtering steps included a minimum coverage of 8, a minimum variant count of 6, a maximum coverage of 500 for the domestic cat pool, and a maximum coverage of 200 for the wildcat pool).

We screened for positively selected candidate genes during cat domestication by parsing specific 100-kb windows that showed low diversity [low pooled heterozygosity (H_p)] in domestic cat breeds and had high divergence [a high fixation index (F_{ST})] between domestic cats and wildcats (41, 81). F_{ST} was calculated using PoPoolation2, and measurements of H_p were calculated using a custom script. A total of 6,534,957 high-quality SNV sites were used to calculate F_{ST} and H_p at each 100-kb window, and a step size of 50 kb was incorporated. All windows containing less than 10 variant sites were removed from the analysis, resulting in $n = 46,906$ 100-kb windows along cat autosomes, as represented in the FelCat5 assembly. We Z-transformed the autosomal H_p [$Z(H_p)$] and F_{ST} [$Z(F_{ST})$] distributions and designated as putatively selected regions those that fell at least 4 SDs away from the mean [$Z(H_p) < -4$ and $Z(F_{ST}) > 4$]. We applied a threshold of $Z(H_p) \leq -4$ and $Z(F_{ST}) \geq 4$ for putative selective sweeps, because windows below or above these thresholds represent the extreme lower and extreme upper ends of the respective distributions (Fig. S4). Windows with elevated F_{ST} or depressed H_p were annotated for gene content using the intersect tool in BEDTools (82). Enrichment analysis of underlying gene content was carried out using WebGestalt (72) using the same methods as described above, except only significant WikiPathways (75) and KEGG Pathways (76) were reported (S2.5 and S2.10–S2.11 in Dataset S2).

Primers to amplify *KIT* exons (ENSFCAG00000003112) were designed using Primer3Plus (83) and annealed to intronic regions flanking each exon. A PCR assay was performed to determine the presence or absence of the dominant, white-spotting retroviral insertion in *KIT* (65). An allele-specific PCR assay was designed for genotyping exon 6 SNPs (S2.9 in Dataset S2). See *SI Materials and Methods* for additional details.

ACKNOWLEDGMENTS. We thank The Genome Institute members Kim Kyung, Dave Larson, Karyn Meltz Steinburg, and Chad Tomlinson for providing assistance and advice on data analysis, and Tom Nicholas for manuscript review. We also thank NIH/National Institute on Alcohol Abuse and Alcoholism members David Goldman and Qiaoping Yuan. The cat genome and transcriptome sequencing was funded by NIH/National Human Genome Research Institute Grant U54HG003079 (to R.K.W.). Further research support included grants to M.V.H. (National Science Foundation Grant DBI-0845494), W.J.M. (Morris Animal Foundation Grants D06FE-063 and D12FE-019), T.M.-B. (European Research Council Starting Grant 260372 and Spanish Government Grant BFU2011-28549), and L.A.L. [National Center for Research Resources (R24 RR016094), Office of Research Infrastructure Programs/Office of the Director (R24 OD010928), and Winn Feline Foundation (W10-014, W09-009)].

1. American Pet Product Manufacturing Association (2008) *National Pet Owner's Survey* (Am Pet Prod Manuf Assoc, Greenwich, CT).
2. Meredith RW, et al. (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* 334(6055):521–524.
3. Hedges SB, Dudley J, Kumar S (2006) TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics* 22(23):2971–2972.

4. Sunquist M, Sunquist F (2002) *Wild Cats of the World* (Univ of Chicago Press, Chicago).
5. Vigne J-D, Guilaine J, Debue K, Haye L, Gérard P (2004) Early taming of the cat in Cyprus. *Science* 304(5668):259.
6. Hu Y, et al. (2014) Earliest evidence for commensal processes of cat domestication. *Proc Natl Acad Sci USA* 111(1):116–120.

7. Menotti-Raymond M, et al. (2008) Patterns of molecular genetic variation among cat breeds. *Genomics* 91(1):1–11.
8. Driscoll CA, et al. (2007) The Near Eastern origin of cat domestication. *Science* 317(5837):519–523.
9. Nussberger B, Greminger MP, Grossen C, Keller LF, Wandeler P (2013) Development of SNP markers identifying European wildcats, domestic cats, and their admixed progeny. *Mol Ecol Resour* 13(3):447–460.
10. Beaumont M, et al. (2001) Genetic diversity and introgression in the Scottish wildcat. *Mol Ecol* 10(2):319–336.
11. Bach LH, et al. (2012) A high-resolution 15,000_{rad} radiation hybrid panel for the domestic cat. *Cytogenet Genome Res* 137(1):7–14.
12. Pontius JU, et al.; Agencourt Sequencing Team; NISC Comparative Sequencing Program (2007) Initial sequence and comparative analysis of the cat genome. *Genome Res* 17(11):1675–1689.
13. Mullikin JC, et al.; NISC Comparative Sequencing Program (2010) Light whole genome sequence for SNP discovery across domestic cat breeds. *BMC Genomics* 11(1):406.
14. Lindblad-Toh K, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069):803–819.
15. Nicholas FW (2003) Online Mendelian Inheritance in Animals (OMIA): A comparative knowledgebase of genetic disorders and other familial traits in non-laboratory animals. *Nucleic Acids Res* 31(1):275–277.
16. Hill RE, Hastie ND (1987) Accelerated evolution in the reactive centre regions of serine protease inhibitors. *Nature* 326(6108):96–99.
17. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.
18. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19(5):671–672.
19. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* 7(10):e46688.
20. Savage RJG (1977) Evolution in carnivorous mammals. *Palaentology* 20:237–271.
21. Heffner RS, Heffner HE (1985) Hearing range of the domestic cat. *Hear Res* 19(1):85–88.
22. Riazuddin S, et al. (2006) Tricellulin is a tight-junction protein necessary for hearing. *Am J Hum Genet* 79(6):1040–1051.
23. Su C-C, et al. (2013) Mechanism of two novel human *GJC3* missense mutations in causing non-syndromic hearing loss. *Cell Biochem Biophys* 66(2):277–286.
24. Huang AS, Kim LA, Fawzi AA (2012) Clinical characteristics of a large choroideremia pedigree carrying a novel *CHM* mutation. *Arch Ophthalmol* 130(9):1184–1189.
25. Ever RF (1973) *The Carnivores* (Cornell Univ Press, New York).
26. Miller KA, et al. (2012) Inner ear morphology is perturbed in two novel mouse models of recessive deafness. *PLoS ONE* 7(12):e51284.
27. Bauer JE (2006) Metabolic basis for the essential nature of fatty acids and the unique dietary fatty acid requirements of cats. *J Am Vet Med Assoc* 229(11):1729–1732.
28. Butterwick RF, Salt C, Watson TDG (2012) Effects of increases in dietary fat intake on plasma lipid and lipoprotein cholesterol concentrations and associated enzyme activities in cats. *Am J Vet Res* 73(1):62–67.
29. Johansson A, et al. (2011) Identification of *ACOX2* as a shared genetic risk factor for preclampsia and cardiovascular disease. *Eur J Hum Genet* 19(7):796–800.
30. Unno N, et al. (2006) A single nucleotide polymorphism in the plasma PAF acetylhydrolase gene and risk of atherosclerosis in Japanese patients with peripheral artery occlusive disease. *J Surg Res* 134(1):36–43.
31. Cho YS, et al. (2013) The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat Commun* 4:2433.
32. Liu S, et al. (2014) Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* 157(4):785–794.
33. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol* 30(8):1987–1997.
34. De Bie T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* 22(10):1269–1271.
35. Kitchener AC (1991) *The Natural History of the Wild Cats* (Cornell Univ Press, New York).
36. Gilad Y, Przeworski M, Lancet D, Lancet D, Pääbo S (2004) Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol* 2(1):E5.
37. Warren WC, et al. (2008) Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453(7192):175–183.
38. Li G, Janacka JE, Murphy WJ (2011) Accelerated evolution of *CE57*, a gene encoding a novel major urinary protein in the cat family. *Mol Biol Evol* 28(2):911–920.
39. Cameron-Beaumont C, Lowe SE, Bradshaw CJA (2002) Evidence suggesting preadaptation to domestication throughout the small Felidae. *Biol J Linn Soc Lond* 75(3):361–366.
40. Driscoll CA, Macdonald DW, O'Brien SJ (2009) From wild animals to domestic pets, an evolutionary view of domestication. *Proc Natl Acad Sci USA* 106(Suppl 1):9971–9978.
41. Axelsson E, et al. (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.
42. Alhaddad H, et al. (2013) Extent of linkage disequilibrium in the domestic cat, *Felis silvestris catus*, and its breeds. *PLoS ONE* 8(1):e53537.
43. Fukuda E, et al. (2008) Down-regulation of protocadherin- α A isoforms in mice changes contextual fear conditioning and spatial working memory. *Eur J Neurosci* 28(7):1362–1376.
44. Mead AN, Stephens DN (2003) Selective disruption of stimulus-reward learning in glutamate receptor *gria1* knock-out mice. *J Neurosci* 23(3):1041–1048.
45. Mead AN, Brown G, Le Merrer J, Stephens DN (2005) Effects of deletion of *gria1* or *gria2* genes encoding glutamatergic AMPA-receptor subunits on place preference conditioning in mice. *Psychopharmacology (Berl)* 179(1):164–171.
46. Horn KE, et al. (2013) DCC expression by neurons regulates synaptic plasticity in the adult brain. *Cell Reports* 3(1):173–185.
47. Yetnikoff L, Almey A, Arvanitogiannis A, Flores C (2011) Abolition of the behavioral phenotype of adult netrin-1 receptor deficient mice by exposure to amphetamine during the juvenile period. *Psychopharmacology (Berl)* 217(4):505–514.
48. Takebe A, et al. (2006) Microarray analysis of *PDGFR α* populations in ES cell differentiation culture identifies genes involved in differentiation of mesoderm and mesenchyme including *ARID3b* that is essential for development of embryonic mesenchymal cells. *Dev Biol* 293(1):25–37.
49. Demirkan A, et al.; DIAGRAM Consortium; CARDIoGRAM Consortium; CHARGE Consortium; EUROSPAN Consortium (2012) Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS Genet* 8(2):e1002490.
50. Wilkins AS, Wrangham RW, Fitch WT (2014) The “domestication syndrome” in mammals: A unified explanation based on neural crest cell behavior and genetics. *Genetics* 197(3):795–808.
51. Kobayashi K, Jakt LM, Nishikawa SI (2013) Epigenetic regulation of the neuroblastoma genes, *Arid3b* and *Mycn*. *Oncogene* 32(21):2640–2648.
52. Wei Z, Yan J, Lu Q, Pan L, Zhang M (2011) Cargo recognition mechanism of myosin X revealed by the structure of its tail MyTH4-FERM tandem in complex with the DCC P3 domain. *Proc Natl Acad Sci USA* 108(9):3572–3577.
53. Zhu X-J, et al. (2007) Myosin X regulates netrin receptors and functions in axonal path-finding. *Nat Cell Biol* 9(2):184–192.
54. Hwang Y-S, Luo T, Xu Y, Sargent TD (2009) Myosin-X is required for cranial neural crest cell migration in *Xenopus laevis*. *Dev Dyn* 238(10):2522–2529.
55. Brown KR, Jurisica I (2007) Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol* 8(5):R95.
56. Hauswirth R, et al. (2012) Mutations in *MITF* and *PAX3* cause “splashed white” and other white spotting phenotypes in horses. *PLoS Genet* 8(4):e1002653.
57. Rubin CJ, et al. (2012) Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci* 109(48):19529–19536.
58. Reissmann M, Ludwig A (2013) Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. *Semin Cell Dev Biol* 24(6-7):576–586.
59. Menotti-Raymond M, et al. (2009) An autosomal genetic linkage map of the domestic cat, *Felis silvestris catus*. *Genomics* 93(4):305–313.
60. Ross-Ibarra J (2004) The evolution of recombination under domestication: A test of two hypotheses. *Am Nat* 163(1):105–112.
61. Coop G, Przeworski M (2007) An evolutionary view of human recombination. *Nat Rev Genet* 8(1):23–34.
62. Kanan Y, Wicker LD, Al-Ubaidi MR, Mandal NA, Kasus-Jacobi A (2008) Retinol dehydrogenases *RDH11* and *RDH12* in the mouse retina: Expression levels during development and regulation by oxidative stress. *Invest Ophthalmol Vis Sci* 49(3):1071–1078.
63. Kurushima JD, et al. (2013) Variation of cats under domestication: Genetic assignment of domestic cats to breeds and worldwide random-bred populations. *Anim Genet* 44(3):311–324.
64. Anonymous (July 22, 1871). Crystal Palace - Summer concert today, Cat Show on July 13. *Penny Illustrated Paper*. p 16.
65. David VA, et al. (2014) Endogenous retrovirus insertion in the *KIT* oncogene determines white and white spotting in domestic cats. *G3 (Bethesda)*, 10.1534/g3.114.013425.
66. Cooper MP, Fretwell N, Bailey SJ, Lyons LA (2006) White spotting in the domestic cat (*Felis catus*) maps near *KIT* on feline chromosome B1. *Anim Genet* 37(2):163–165.
67. Geissler EN, Ryan MA, Housman DE (1988) The dominant-white spotting (*W*) locus of the mouse encodes the c-kit proto-oncogene. *Cell* 55(1):185–192.
68. Blechman JM, et al. (1995) The fourth immunoglobulin domain of the stem cell factor receptor couples ligand binding to signal transduction. *Cell* 80(1):103–113.
69. Miller JR, et al. (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24(24):2818–2824.
70. Flicek P, et al. (2012) Ensembl 2012. *Nucleic Acids Res* 40(database issue):D84–D90.
71. Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26(15):1899–1900.
72. Wang J, Duncan D, Shi Z, Zhang B (2013) WEB-based GEne SeT Analysis Toolkit (WebGestalt): Update 2013. *Nucleic Acids Res* 41(web server issue):W77–W83.
73. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene Ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
74. Cerami EG, et al. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* 39(database issue):D685–D690.
75. Kelder T, et al. (2012) WikiPathways: Building research communities on biological pathways. *Nucleic Acids Res* 40(database issue):D1301–D1307.
76. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(database issue):D109–D114.
77. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57(1):289–300.
78. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
79. Kolobold DC, et al. (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22(3):568–576.
80. Kofler R, Pandey RV, Schlötterer C (2011) PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27(24):3435–3436.
81. Rubin C-J, et al. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464(7288):587–591.
82. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
83. Untergasser A, et al. (2012) Primer3—New capabilities and interfaces. *Nucleic Acids Res* 40(15):e115.