Supporting Information

Montague et al. 10.1073/pnas.1410083111

SI Materials and Methods

Genome Assembly. The current draft assembly is referred to as FelCat5 or *Felis catus* 6.2. There are ∼2.35 Gb (including Ns in gaps) on ordered/oriented chromosomes, ∼15.4 Mb on the chr*_random, and ∼11.74 Mb on chromosome Un. Initially, we ran CABOG 6.1 (1) with default parameters (2). To evaluate changes in contiguity, we altered a small set of the default parameters to obtain the best assembly possible. CABOG settings, including parameters used, are available upon request.

To create an initial chromosomal version of the assembly, we aligned marker sequences associated with a radiation hybrid (RH) map (3) to the assembled genome sequence. The chromosomal index file (.agp) contains the ordered/oriented bases for each chromosome (named after the respective linkage group).

Once scaffolds were ordered and oriented along the cat chromosomes using the RH map marker content (3), the assembled cat genome was broken into 1-kb segments and aligned against the dog genome (CanFam2) and human genome (hg19) using BLASTZ (4) to align and score nonrepetitive cat regions against repeat-masked dog and human sequences, respectively. BLASTZ (4) and BLAT (5) alignments with the dog and human genomes were then used to refine the order and orientation information as well as to insert additional scaffolds into the conditional scaffold framework provided by the marker assignments. Alignment chains differentiated all orthologous and paralogous alignments, and breakpoint identification confirmed a false join within the genome assembly. Only "reciprocal best" alignments were retained in the alignment set. Finally, satellite sequences were identified in the genome, and centromeres were placed along each chromosome using localization data (3) in combination with the localization of the satellite sequences. In the last step, finished cat BACs (n = 86; totaling ∼14.92 Mb) were integrated into the assembly, using the BLAT (5) aligner for accurate coordinates.

Gene Family Expansions and Contractions. To explore gene family expansions and contractions, we obtained peptides from cat, dog, ferret, panda, cow, pig, horse, human, and elephant from Ensembl (6). We clustered these into protein families by performing an all-against-all BLAST (7) search using the OrthoMCL clustering program (8). The clusters were converted to CAFE (9) format, and families were filtered out based on the following groupings: (*i*) at least one protein must be present in (elephant, human, horse), and (*ii*) at least one protein must be present in (cat, dog, ferret, panda, cow, pig), or the family is filtered out. We used www.timetree.org to obtain divergence times for all species to construct the following tree: (elephant:101.7, (human:94.2, (horse:82.4, ((pig:63.1, cow:63.1):14.3, (cat:55.1, (dog:42.6, (ferret:38, panda:38):4.6):12.5):22.3):5):11.8):7.5).

From phylogenetic inference, we found 50 expanded gene families in the cat genome, of which 28 have known homologs in other mammals (Fig. S3 and S1.8 in Dataset S1). Analyses using CAFE 3.0 (9) confirmed contraction in multiple *Or* gene families and expansion in the *V1r* gene family in the ancestor of modern felids (S1.8 in Dataset S1), with differential gene gain and loss within the cat family (Fig. S2).

In addition to the chemosensory *Or* and *V1r* gene families mentioned, we found evidence for expansion of genes related to processes of mechanotransduction (10) (*PIEZO2*), T-cell receptors (*TRAV8*), melanocyte development (11) (*SOX10*), and meiotic processes (*SYCP1*). Four gene families were complete losses along the cat lineage; the annotations for these entries for other species include gene families related to reproduction (spermatogenesis-associated protein 31D1 and precursor acrosomal vesicle 1), secretory proteins (precursor lipophilin), and hair fibers (high sulfur keratin associated).

Segmental Duplication, Copy Number (CNV) Discovery, and Structural Variation.

Sequencing data. For the domestic cat (Abyssinian), sequenced with Illumina technology, bam files resulting from mapping 100-bp reads were used to recover the original fastq reads, which were clipped into 36-bp reads after trimming the first 10 bp to avoid lower-quality positions. That is, we used a total of 1,485,609,004 reads for mapping (coverage 21.8×).

Reference assembly. We downloaded the FelCat5 assembly from the UCSC Genome Browser (12). The 5,480 scaffolds either unplaced or labeled as random were concatenated into a single artificial chromosome. In addition to the repeats already masked in FelCat5 with RepeatMasker (www.repeatmasker.org) and Tandem repeats finder (13), we sought to identify and mask potential hidden repeats in the assembly. To do so, chromosomes were partitioned into 36-bp K-mers (with adjacent K-mers overlapping 5 bp), and these were mapped against FelCat5 using mrsFAST (14). Next, we masked positions in the assembly mapped by K-mers with more than 20 placements in the genome, resulting in 5,942,755 bp additionally masked compared with the original masked assembly (Fig. S3).

Mapping and copy number estimation from read depth. In the domestic cat, the 36-bp reads resulting from clipping the original fastq reads (see above) were mapped to the prepared reference assembly using mrFAST (15). mrCaNaVaR (version 0.41) (15) was used to estimate the copy number along the genome from the mapping read depth. Briefly, mean read depth per base pair is calculated in 1-kbp nonoverlapping windows of nonmasked sequence (that is, the size of a window will include any repeat or gap, and thus the real window size may be larger than 1 kbp). Importantly, because reads will not map to positions covering regions masked in the reference assembly, read depth will be lower at the edges of these regions, which could underestimate the copy number in the subsequent step. To avoid this, the 36 bp flanking any masked region or gap were masked as well and thus are not included within the defined windows. In addition, gaps >10 kbp were not included within the defined windows. A read depth distribution was obtained through iteratively excluding windows with extreme read depth values relative to the normal distribution, and the remaining windows were defined as control regions (Fig. S3 and S1.9 in Dataset S1). The mean read depth in these control regions was considered to correspond to a copy number equal to two and was used to convert the read depth value in each window into a GC-corrected absolute copy number. Note that the control/noncontrol status was determined based on the read depth distribution, making this step critical for further copy number calls. Of the 993,102 control windows, none aligned to the artificial chromosome (see above), and 37,123 (3.7%) were on chromosome X in the sample.

Calling of duplications and deletions. The copy number distribution in the control regions was used to define specific gain/loss cutoffs as the mean copy number plus/minus three units of SD (calculated not considering those windows exceeding the 1% highest copy number value). Note that because the mean copy number in the control regions was equal to two (by definition), the gain/loss cutoffs were largely influenced by the SD.

We used two methods to call duplications: M1, the circular binary segmentation (CBS) method (16), was used to combine 1-kbp windows that represent segments with significantly the same copy number. Segments with copy number (defined as the median copy number of the 1-kbp windows comprising the segment) exceeding the gain/loss cutoffs defined above (but lower than 100 copies in the case of duplications) were merged and called as duplications or deletions if comprising more than 10 1-kbp windows (~10 kbp); finally, only duplications with >85% of their size not overlapping with repeats were retained for the analyses.

As a second method (M2), we also called duplications avoiding the segmentation step with the CBS method by merging 1-kbp windows with copy number larger than sample-specific gain cutoff (but lower than 100 copies) and then selecting those regions comprising at least five 1-kbp windows and >10 kbp; similarly, only duplications with >85% of their size not overlapping with repeats were retained for the analyses.

In M1, the copy number distribution in the control regions was used to define sample-specific gain/loss cutoffs as the mean copy number plus/minus three units of SD (calculated not considering those windows exceeding the 1% highest copy number value). Note that because the mean copy number in the control regions is equal to two by definition, the gain/loss cutoffs will be largely influenced by the SD. Then, we merged 1-kbp windows with copy number larger than sample-specific gain cutoff (but lower than 100 copies) and identified as duplications the regions that comprised at least five 1-kbp windows and >10 kbp. Finally, only duplications with >85% of their size not overlapping with repeats were retained. This method is highly restrictive (conservative), so we used an alternative method (M2) similar to what had been previously done with Sanger capillary reads (17). We performed a 5-kbp sliding window approach and required six out of seven windows with a significantly higher read depth, relative to the control regions, to consider a region as duplicated.

Several categories were significantly overrepresented in regions of expanded CNV (Fig. S3 and S1.10–S1.12 in Dataset S1), some of which overlap those identified in other CNV studies for other taxa (18–22). In the cat, we note that an expanded CNV region on chromosome B2 contained a pair of genes that transcribe an MHC class I antigen and an MHC class I antigen precursor. The MHC class I molecules present self-antigens to cytotoxic $CD8^+$ T lymphocytes and regulate natural killer cell activity. Investigations of MHC genes in other domesticated animals, including pig (23), sheep (24), and cow (25, 26), have shown that MHCs in these groups are affected by CNV. These results suggest that CNV is an additional common source of disease resistance or susceptibility variability in the MHC of the cat as well.

***V1r/Or* Identification and Annotation.** Published *V1r* and *Or* sequences from human, mouse, rat, cow, dog, and opossum were used as the query sequences for BLAST (7) searches against the domestic cat genome. All query sequences were previously shown as belonging to *V1r* (27, 28) and *Or* (29) subfamilies, thus ensuring identification of the most complete gene repertoires. We enforced an *E*-value threshold of 10^{-5} for filtering BLAST results. All identified sequences were extended 1.5 kb on either side for open reading identification and assessment of functionality. If multiple start codons were found, the alignment results of known intact mammalian *V1r* and *Or* amino acid sequences were used as guidance for determining the most appropriate one. Any putative genes containing early stop codons, frameshift mutations, and/or incomplete gene structure (i.e., not containing three extracellular regions, seven transmembrane regions, and three intracellular regions) were designated as pseudogenes. To confirm orthology, we aligned all members of the *V1r* and *Or* gene families and constructed maximum likelihood trees rooted with appropriate outgroup taxa, such as *V2r* and taste receptor gene families. Assembled whole-sequencing data were obtained from the Ensembl database (6) [domestic cat: vFelCat5; domestic dog: vCanFam; domestic horse: vEquCab2; human: vGRCh39; domestic cow: vBosTau7; great panda: vAilMel1; and tiger (tigergenome.org)]. *V1r* gene clusters were defined as all identified functional genes and pseudogenes within a 2-Mb window. Synteny blocks of different mammals were identified using the software SyntenyTracker (30).

Felid V1r sequencing. The following felid taxa were used for *V1r* PCR and sequencing: *Felis catus* (domestic cat; FCA), *Felis nigripes* (Black-footed cat; FNI), *Prionailurus bengalensis* (Leopard cat; PBE), *Prionailurus viverrinus* (Fishing cat; PVI), *Puma concolor* (Cougar; PCO), *Puma yagouaroundi* (Jaguarundi; PYA), *Acinonyx jubatus* (Cheetah; AJU), *Lynx canadensis* (Canadian Lynx; LCA), *Lynx lynx* (Eurasian Lynx; LLY), *Lynx pardinus* (Iberian Lynx; LPA), *Lynx rufus* (Bobcat; LRU), *Leopardus pardalis* (Ocelot; LPA), *Leopardus wiedii* (Margay; LWI), *Leopardus geoffroyi* (Geoffroy's cat; LGE), *Leopardus colocolo* (Pampas cat; LCO), *Leopardus tigrinus* (Tiger cat; LTI), *Profelis serval* (Serval; PSE), *Profelis caracal* (Caracal; PCL), *Pardofelis temminckii* (Asian Golden cat; PTE), *Pardofelis marmorata* (Marbled cat; PMA), *Neofelis nebulosa* (Clouded Leopard; NNE), *Panthera leo* (Lion; PLE), *Panthera onca* (Jaguar; PON), *Panthera pardus* (Leopard; PPA), *Panthera tigris* (Tiger; PTI), and *Panthera uncia* (Snow Leopard; PUN). Forty-three pairs of primers for *V1r* amplification were designed using several versions of the domestic cat whole-genome assembly (FelCat1–FelCat5). Target amplicons were designed to be longer than 1.1 kb to ensure amplification of the complete coding region sequence. PCR was performed using PlatinumTaq DNA polymerase using a touchdown profile of 60–55 °C, as described (31). All amplicons were sequenced using Sanger sequencing on an ABI 3700 (Applied Biosystems). A total of 1,055 sequences of intact *V1r* genes and pseudogenes from 27 cat species were submitted to GenBank under accession numbers KJ923925–KJ924979.

Sequence alignment and phylogenetic reconstruction. We aligned our previously unidentified *V1r* sequences with known published *V1r* sequences using MAFFT (32) with stringent parameter settings. Coding sequences were aligned under the guidance of the translated amino acid alignment results. Poorly aligned 5′ and 3′ flanking regions were trimmed before tree building. MODELTEST (33) was used to estimate the best nucleotide substitution models and parameters for sequence data. Maximum likelihood trees (with 500 bootstrap replicates) were constructed with RAxML7.0.0 (34).

Estimation of gene gain and loss within V1r and Or gene families. We compared the *Or* and *V1r* gene trees generated above with a mammalian species tree (35) to estimate gene gain and loss using the software NOTUNG (36). We examined variation in *V1r* and *Or* gene family repertoire size among different domestic cat breeds by aligning Illumina reads to the cat assembly using BWA (37). Mapping results were analyzed with CNVnator (38). We reestimated the tiger *Or* and *V1r* repertoires by remapping all of the raw tiger Illumina reads to the Siberian tiger assembly (tigergenome.org) as well as the current domestic cat version 6.2 assembly.

Natural Selection Tests.

Phylogenetic analyses by maximum likelihood. Four sets of models were applied for null hypothesis and alternative hypothesis comparisons. Set 1 involved a comparison between the free-ratio model and the one-ratio model, whereas set 2 compared the two-ratio model with the one-ratio model. These two comparisons are classified as branch-specific tests, which were used to identify accelerated rates of genes on specific branches of an evolutionary tree. In addition, we performed site-specific tests, which detected natural selection acting on specific amino acid sites of the protein. For this step, we performed model tests within sets 3 and 4, which involved model 1a (nearly neutral) versus model 2a (positive selection) and model 7 (gamma) versus model 8

(gamma and ω) to evaluate and identify specific amino acid sites that were potentially under positive selection.

To evaluate the structural influence of domestic cat nonsynonymous substitutions from the common ancestor of felids, we used TreeSAAP (39) to measure 31 structural and biochemical amino acid properties while applying the tree topology (human, (cow, (dog, (cat, tiger)))). We used a significance threshold of $P < 0.001$ to report structural or biochemical properties of amino acid substitutions likely to affect protein function. We also used PROVEAN (40) to predict the potential functional impact of domestic cat-specific amino acid substitutions and indels. We considered amino acid substitutions as "deleterious" if the PROVEAN score was ≤−2.5. We considered amino acid substitutions as "neutral replacements" if the PROVEAN score was >−2.5 (Fig. S1).

To explore the heterogeneous selection pressure across positively selected genes, peaks of high d_N/d_S were visualized using sliding window analyses performed across alignments of the full coding sequence. Sliding windows of ω values were estimated using the Nei and Gojobori method (41) with a default window size of 90 bp and a step size of 18 bp.

Many positively selected genes appear to have played a role in the sensory evolution of felines, as highlighted above. For instance, chemosensory genes with significant signatures of positive selection in the Felinae include two gustducin-coupled bitter taste receptors, *TAS2R1* and *TAS2R3*, as well as a cofactor, *RTP3* (S1.3 in Dataset S1). We speculate that selection at these loci increased sensitivity to and avoidance of toxic prey items in the hypercarnivorous ancestor of cats (42). Other positively selected genes appear to have played a role in the morphological evolution of carnivores. For instance, all carnivores have robust claws (except where they are secondarily lost) that serve as critical adaptations to capture and disarticulate prey. The *RSPO4* gene (S1.1 in Dataset S1) plays a crucial role in nail morphogenesis across mammals, and its expression is restricted to the developing nail mesenchyme (43). Further, the recessive human disorders anonychia/hyponychia congenita result from mutations in *RPSO4* (44), and are characterized by absence of or severe reduction in fingernails and toenails. Evidence of positive selection within the *RSPO4* gene in the ancestral carnivore lineage likely reflects molecular adaptations driving enhanced nail morphology.

Genome mapping and variant analysis. We next performed whole-genome analyses of cats from different domestic breeds [Maine Coon (SRX026946, SRX026943, SRX026929), Norwegian Forest (SRX027004, SRX026944, SRX026941, SRX026909, SRX026901), Birman (SRX026955, SRX026947, SRX026911, SRX026910), Japanese Bobtail (SRX026948, SRX026928, SRX026912), Turkish Van (SRX026942, SRX026930, SRX026913), and Egyptian Mau (SRX019549, SRX019524, SRX026956, SRX026945)] and wildcats [i.e., other *F. silvestris* subspecies (SRX026960)] using pooling methods that control for genetic drift (45). All reads were preprocessed by removing duplicate reads and only properly paired reads were aligned to the FelCat5 reference using BWA (37) (n = 2,332,398,473 reads from the pooled domestic cats combined; n = 189,543,907 reads from pooled wildcats). A total of 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, passed the thresholds using our initial variant-calling methods with SAMtools (46) and VarScan (47). Because SNVs for the domestic and wildcat pools were called separately, variants ascertained in one may not be present in the other. This can be due to homozygosity for the reference allele or inadequate data at the locus. We therefore implemented a consensus-calling analysis for the combined variant set to categorize each SNV as high-quality passing, low-quality failure, or no sequence coverage within each pool for all 10,975,197 passing sites. To do this, we generated a two-sample mpileup using SAMtools (46) for every site that was called a variant. We next implemented the mpileup2cns command in VarScan (47) with the minimum read depth set to three. Because every site in the mpileup passed the initial false positive filtering in at least one pool, we were able to determine the percentage of variant overlap between the pool of domestic cats and the pool of wildcats. This revealed 9,010,197 shared variant alleles between the domestic cats and wildcats, indicating that 1.7% and 10.3% of sites with variant alleles were unique to domestic cats and wildcats, respectively (Fig. S4). As expected, due to the coverage differences between the pools, a total of 3,121 and 745,091 sites, in the pooled domestic cats and pooled wildcats, respectively, contained low coverage (fewer than three aligned reads) or missing coverage.

We next used VCFtools (48) to explore the extent of overlap between the different variant callers. For the domestic cat pool, SAMtools (46) called 11,119,091 variants and VarScan (47) called 10,138,788 variants. A total of 9,683,549 variants overlapped, revealing that 4.5% and 12.9% of the original VarScan (47) and original SAMtools (46) calls, respectively, were undetected by the other variant caller. For the wildcat pool, SAMtools (46) called 9,860,972 variants and VarScan (47) called 9,098,242 variants. A total of 7,848,268 variants overlapped, revealing that 13.7% and 20.4% of the original VarScan (47) and original SAMtools (46) calls, respectively, were undetected by the other variant caller.

SNV validation. We verified our high-quality set of SNVs by comparing the list of markers with those of an SNP array developed previously (49). To accomplish this, we used BLAST (7) to locate the best-hit coordinates along the *F. catus* 6.2 reference assembly for each of the array variants. We then parsed our pooled domestic cat variant file for matching coordinates and discovered 184 out of 384 variants (47.9%). The calls made by our pipeline matched the variant on the chip in 183 out 184 cases (99.5%).

Breed differentiation. We verified the genetic relationships among the breeds using multidimensional scaling (MDS) and a population stratification analysis. Seven populations of 26 domestic cats were analyzed, including the breeds described above as well as a population of Eastern Random Bred cats (n = 4; SRX026993). Genome mapping and variant calling was performed on a per-breed basis using described variant-calling methods (above). After aligning the short reads to FelCat5, we identified 77,749 high-quality variants that were shared among all seven breeds. The pedigree genotype file was quality-controlled with PLINK (50) to remove all individuals with more than 80% missing genotype data, all SNVs missing in more than 5% of cases, and all SNVs with less than 5% minor allele frequency (MAF). Following quality control filtering, a total of 44,377 autosomal SNVs remained. MDS was implemented using PLINK (50) to produce an output file with identity by state values, and genetic distances of the first four principal coordinates were visualized (Fig. S5). Model-based clustering was performed with ADMIXTURE (51). A total of 20 replicates of K = 2 to K = 20 was run in unsupervised mode, each with random seeds and fivefold cross-validation. The replicates of each Q file for each K were merged using the LargeKGreedy method (with random input orders) using the program CLUMPP (52). The merged Q files were then visualized in DISTRUCT (53) to output plots of estimated membership coefficients for each individual according to each K, with K = 5 offering the highest support (Fig. S5).

Discovering putative regions of selection in the domestic cat genome. As a quality control assessment, the average H_p and F_{ST} of all autosomal 100-kb windows were plotted against the corresponding number of segregating sites per window. In line with our expectations, H_p was positively correlated with the number of segregating sites (rho = 0.021, $P < 0.001$, Spearman; Fig. S5) whereas F_{ST} was negatively correlated (rho = 0.225, $P < 0.001$, Spearman; Fig. S5), suggesting that the number of variants per window was lower in our putative regions of selection due to the loss of linked variation following an adaptive sweep. We also compared the depth of coverage at variant sites within the putative regions of selection with the depth of coverage at variants

found within all other genomic regions. The average read depth among the 3,265 variant positions for pooled domestic cats within the five regions of putative selection was relatively equivalent to the average read depth of all 8,676,486 variants across all autosomes for pooled domestic cats (53.82 versus 53.65, respectively).

Although accurate detection of heterozygosity is dependent on coverage, similar depths among the breed pools and members of each pool were not obtained for this study. Further, individual cats were not indexed when pooling by breed. Although equal numbers of samples among pools and subsets were difficult to obtain, we tested whether unequal representation between domestic cat and wildcat contributed to variance of the divergence statistics across the genome by reperforming the F_{ST} analysis based on a random subsample of the domestic cat data where the average coverage (6.81×) approximated the original coverage for the wildcat pool (6.84×), with ∼1.1× coverage contributed by each domestic breed pool. First, the variant-calling pipeline identified 3,494,488 total variants in the subsampled data. A final variant set consisting of 1,274,175 autosomal variants was then used for a sliding window analysis of F_{ST} using the same methods as the original analysis. When analyzing the subsampled data, all windows that passed the threshold under the original analysis were found within the 99th percentile of highest F_{ST} using the subsampled domestic cat data (Fig. S5). All of the original windows were thus identified as windows with high divergence using the subsampled data. These results suggest that the unequal sample sizes of domestic cats and wildcats likely had little effect on the overall results of our sliding window analyses.

Analysis of the X chromosome. To not confound the results of the autosomal analyses, we analyzed the X chromosome separately, using the method as described previously for the autosomes. We found that the average pooled heterozygosity, H_p, is higher (H_pX: 0.496 vs. H_pA: 0.385) and the average fixation index, F_{ST}, is higher (F_{ST}X: 0.674 vs. F_{ST}A: 0.429) on X compared with on autosomes. We also note that the SDs of the H_p (σX: 0.049 vs. σA: 0.029) and F_{ST} (σX: 0.183 vs. σA: 0.074) distributions are larger on the X chromosome relative to the autosomes. No windows passed the thresholds of significance [$Z(H_p) < -4$ or $Z(F_{ST}) > 4$] used for the autosomal analyses. We instead applied a lower threshold of 1.5 SDs from the mean of both the H_p and F_{ST} distributions. A total of 54 windows, representing 36 unique regions, passed this cutoff in the F_{ST} analysis (S2.12 in Dataset S2). A total of 210 windows representing 72 unique regions passed this threshold for the H_p analysis (S2.13 in Dataset S2). Known genes underlying regions of low domestic H_p and high F_{ST} (Fig. S6 and S2.14 in Dataset S2) include cyclin B3 (*CCNB3*), Cdc42 guanine nucleotide exchange factor 9 (*ARHGEF9*), zinc finger C4H2 domain containing (*ZC4H2*), family with sequence similarity 155, member B (*FAM155B*), protocadherin 19 (*PCDH19*), annexin A2 (*ANXA2*), and brain expressed X-linked 5 (*BEX5*). Our sliding window analysis along the autosomes revealed a strong trend associating genomic signatures of selection in domestic cats with genes influencing memory, fear-conditioning behavior, and stimulus-reward learning, particularly those predicted to underlie the evolution of tameness (54). This analysis of the X chromosome reveals similar functional trends, with four of six regions containing genes associated with neurological diseases and aberrant synaptic activity, including an additional protocadherin locus.

The Z-transformation technique, outlined above, resulted in a skewed (i.e., not normal) distribution (Fig. S6), so the conclusions must be viewed cautiously. By applying a percentile approach, we found that no genes underlie windows that met thresholds for either the 99th percentile or the 95th percentile for both F_{ST} and domestic H_p. Only a single window met the 99th percentile for F_{ST} and the 99th percentile for domestic H_p. This window (X:23800000–23900000), although noncoding, is within the X-linked *MAGE* gene family complex. The protocadherin gene that we highlighted above (*PCDH19*) was found within the 95th percentile threshold for F_{ST} and the 90th percentile for domestic H_p. The annexin gene (*ANXA2*), which is located within an adjacent window to *PCDH19*, met the 95th percentile threshold for F_{ST} and the 85th percentile for domestic H_p. Only one other gene displayed a higher F_{ST} value than *PCDH19* and *ANXA2*: *BEX5*, also highlighted by our Z-transformation analysis, met the 99th percentile threshold for F_{ST} and the 90th percentile for domestic H_p.

Pigmentation Patterns in Domestic Cat Breeds. Several breeds represent random bred populations of cats that do not have strong selection on a specific trait, such as Maine Coon and Norwegian Forest; however, the vast majority of cat breeds, including Japanese Bobtail, Birman, Egyptian Mau, and Turkish Van, likely experienced strong selection on novel and specific mutations (i.e., morphological traits and pigmentation patterns), as individuals were selected from random bred populations.

The genomic sequence data from the pooled Birman breed revealed an ∼10-Mb homozygous block located directly upstream of *KIT*. The average nucleotide diversity for 100-kb windows adjacent to *KIT* was lower (ChrB1: 161.5–161.9 Mb; pi = 0.0011) than the average nucleotide diversity for 100-kb windows across all autosomes (pi = 0.2185) or the average nucleotide diversity for 100-kb windows across ChrB1 (pi = 0.1762) (Fig. 4). An additional analysis of 63K single-nucleotide variants in individual Birman cats revealed an ∼5-Mb homozygous block located directly upstream of *KIT*. This loss of variation could be explained by genetic drift (e.g., inbreeding, the small founder population of the breed) or as a consequence of selection (e.g., the white gloving trait is fixed and recessive). We hypothesized that an extensive homozygous block is a measure of the selection on the gloving trait because Birman is highly selected for coat color, and we discovered a unique pair of fixed SNVs within the Birman breed that are associated with amino acid changes in *KIT*.

Samples and genotyping. We noninvasively collected DNA samples from all domestic cats by buccal swabs using a cytological brush or cotton tip applicator. DNA was isolated using the QIAamp DNA Mini Kit (Qiagen). The previous linkage analysis pedigree from the Waltham Centre for Pet Nutrition (55) was extended from 114 to 147 cats to refine the linkage region. Phenotypes were determined as in the previous study (55). Two previously published short tandem repeats (STRs) (56) (*FCA097* and *FCA149*) and four previously unidentified feline-derived STRs (*UCDC259b*, *UCDC443*, *UCDC487*, and *UCDC489*) (S2.9 in Dataset S2), flanking *KIT* on feline chromosome B1, were genotyped. Genotyping for the markers and two-point linkage between the microsatellite genotypes and the spotting phenotype was conducted using the LINKAGE (57) and FASTLINK (58) programs as in previous studies (59).

Genomic analysis of KIT. To identify *KIT* exons, publicly available (in GenBank) sequences from various species were aligned, including *Homo sapiens* (NM_000222.2), *Canis familiaris* (NM_001003181.1), *Mus musculus* (NM_021099.3), and *Equus caballus* (NM_001163866.1) and a partial sequence for the domestic cat, *F. catus* (NM_001009837.3), because *F. catus KIT* was located on the previous version of the assembly (60) (GeneScaffold_3098:168,162–233,592). Primers (Operon) were tested for efficient product amplification, and the final magnesium and temperature conditions for each primer pair are presented in Dataset S2 (S2.9). PCR and thermocycling conditions were conducted as previously described (61). The PCR products with the appropriate lengths were purified using the ExoSap (USB) enzyme per the manufacturer's recommendations. Purified genomic products were sequenced using BigDye Terminator Sequencing Kit version 3.1 (Applied Biosystems), purified with Illustra Sephadex G-50 (GE Healthcare) according to the manufacturer's recommendations, and electrophoretically separated on an ABI 3730 DNA Analyzer (Applied Biosystems). Sequences were verified and aligned using the software Sequencher version 4.8 (Gene Codes). The complete coding sequence for *F. catus KIT* was submitted to GenBank under accession number GU270865.1.

KIT mRNA analysis. RNA from a nonwhite control cat was isolated from whole blood using the PAXgene Blood RNA Kit (Qiagen) following the manufacturer's directions. The 5′ UTR amplification and the PCR analysis were conducted as previously described (62). The 5′ RACE used the cDNA pool generated by the KIT-specific primers (S2.9 in Dataset S2). The 5′ RACE PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen) before sequencing. Five 5′ RACE cDNA clones from the control cat were selected and sequenced. Genomic primers (S2.9 in Dataset S2) were then designed in the 5′ UTR region to sequence the cats used for the genomic analysis of *KIT* (S2.7 in Dataset S2).

KIT SNP genotyping. An allele-specific PCR (AS-PCR) assay was designed for genotyping exon 6 SNPs (S2.9 in Dataset S2). Both allele-specific primer pairs annealed at the 2-nt primer–template mismatch (c.1035_1036delinsCA; p.Glu345Asp; His346Asn; S2.9 in Dataset S2). The AS-PCR assay used 1× buffer, 1.5 mM $MgCl_2$, 200 μM each dNTP, and 0.1 U Taq (Denville), per 15 μL of reaction mixture. The primer concentrations in each PCR were 0.67 μM KITgloA-FAM, 0.67 μM KITgloB-VIC, and 0.67 μM KITR. PCR conditions were: initial denaturation at 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s, and a final extension step of 72 °C for 7 min. The amplified products were separated on an ABI 3730 DNA Analyzer (Applied Biosystems). The genotypes were scored based on fluorescence intensity using the software STRand (63). The variants and exons within *KIT* were schematically presented with FancyGene (64).

Exploring other potential regulatory variation within KIT. We initially planned to investigate only the exonic regions of *KIT*, even though flanking or intronic regions often regulate gene expression. Along this line of reasoning, an ∼7-kb retroviral (FERV1) insertion within *KIT* intron 1 was recently identified as the causative factor for white spotting among different cat breeds (65); however, the Birman breed was not surveyed for the insertion. We therefore searched for the dominant, white-spotting FERV1 insertion sequence in the pooled Birman genomic sequence data (with estimated 4× coverage). To do this, we aligned all ∼190 million 50-bp reads from the Birman pool to the 7,296-bp FERV1 insertion sequence and generated a consensus to compare with the FERV1 reference. A total of 778 reads aligned using BWA (37), but the result was ambiguous due to the following observations: (*i*) 1,169 bp (16%) of the FERV1 reference were missing across 23 regions, with an average of 50.8 bp missing per region; and (*ii*) there were regions of 275, 173, and 296 bp within the FERV1 reference with no read coverage. Instead, we designed a long-range PCR experiment (for primers and conditions, see ref. 65) to capture the white-spotting alleles in mitted and bicolor Ragdoll ($n = 10$), Birman ($n = 10$), and other white-spotted ($n = 5$) and solid ($n = 5$) cats. Whereas the FERV1 insertion was confirmed in spotted Ragdolls and other spotted cats, we found no evidence for the insertion in all Birman cats and solid cats. These results demonstrate a second mechanism for white spotting in the Birman breed while also confirming a separate mode of inheritance. Future experiments will investigate how the fixed mutations in *KIT* exon 6 interact with *KIT* regulatory elements during expression.

1. Miller JR, et al. (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24(24):2818–2824.
2. Salzberg SL, et al. (2012) GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Res* 22(3):557–567.
3. Davis BW, et al. (2009) A high-resolution cat radiation hybrid and integrated FISH mapping resource for phylogenomic studies across Felidae. *Genomics* 93(4): 299–304.
4. Schwartz S, et al. (2003) Human-mouse alignments with BLASTZ. *Genome Res* 13(1): 103–107.
5. Kent WJ (2002) BLAT—The BLAST-like alignment tool. *Genome Res* 12(4):656–664.
6. Flicek P, et al. (2012) Ensembl 2012. *Nucleic Acids Res* 40(database issue):D84–D90.
7. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
8. Li L, Stoeckert CJ, Jr, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13(9):2178–2189.
9. 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.
10. Coste B, et al. (2010) *Piezo1* and *Piezo2* are essential components of distinct mechanically activated cation channels. *Science* 330(6000):55–60.
11. Hou L, Arnheiter H, Pavan WJ (2006) Interspecies difference in the regulation of melanocyte development by *SOX10* and *MITF*. *Proc Natl Acad Sci USA* 103(24): 9081–9085.
12. Karolchik D, et al. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32(Suppl 1):D493–D496.
13. Benson G (1999) Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res* 27(2):573–580.
14. Hach F, et al. (2010) mrsFAST: A cache-oblivious algorithm for short-read mapping. *Nat Methods* 7(8):576–577.
15. Alkan C, et al. (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* 41(10):1061–1067.
16. Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5(4): 557–572.
17. Bailey JA, et al. (2002) Recent segmental duplications in the human genome. *Science* 297(5583):1003–1007.
18. Fontanesi L, et al. (2012) Exploring copy number variation in the rabbit (*Oryctolagus cuniculus*) genome by array comparative genome hybridization. *Genomics* 100(4):245–251.
19. Graubert TA, et al. (2007) A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet* 3(1):e3.
20. Chen W-K, Swartz JD, Rush LJ, Alvarez CE (2009) Mapping DNA structural variation in dogs. *Genome Res* 19(3):500–509.
21. Nicholas TJ, et al. (2009) The genomic architecture of segmental duplications and associated copy number variants in dogs. *Genome Res* 19(3):491–499.
22. Fontanesi L, et al. (2010) An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genomics* 11:639.
23. Tanaka-Matsuda M, Ando A, Rogel-Gaillard C, Chardon P, Uenishi H (2009) Difference in number of loci of swine leukocyte antigen classical class I genes among haplotypes. *Genomics* 93(3):261–273.
24. Fontanesi L, et al. (2011) A first comparative map of copy number variations in the sheep genome. *Genomics* 97(3):158–165.
25. Liu GE, et al. (2010) Analysis of copy number variations among diverse cattle breeds. *Genome Res* 20(5):693–703.
26. Fadista J, Thomsen B, Holm L-E, Bendixen C (2010) Copy number variation in the bovine genome. *BMC Genomics* 11:284.
27. Shi P, Bielawski JP, Yang H, Zhang Y-P (2005) Adaptive diversification of vomeronasal receptor 1 genes in rodents. *J Mol Evol* 60(5):566–576.
28. Young JM, Massa HF, Hsu L, Trask BJ (2010) Extreme variability among mammalian *V1R* gene families. *Genome Res* 20(1):10–18.
29. Niimura Y, Nei M (2007) Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* 2(8):e708.
30. Donthu R, Lewin HA, Larkin DM (2009) SyntenyTracker: A tool for defining homologous synteny blocks using radiation hybrid maps and whole-genome sequence. *BMC Res Notes* 2:148.
31. Murphy WJ, O'Brien SJ (2007) Designing and optimizing comparative anchor primers for comparative gene mapping and phylogenetic inference. *Nat Protoc* 2(11): 3022–3030.
32. Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26(15):1899–1900.
33. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14(9):817–818.
34. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.
35. Murphy WJ, et al. (2001) Molecular phylogenetics and the origins of placental mammals. *Nature* 409(6820):614–618.
36. Chen K, Durand D, Farach-Colton M (2000) NOTUNG: A program for dating gene duplications and optimizing gene family trees. *J Comput Biol* 7(3-4):429–447.
37. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
38. Abyzov A, Urban AE, Snyder M, Gerstein M (2011) CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 21(6):974–984.
39. 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.
40. 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.
41. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3(5):418–426.
42. Shi P, Zhang J (2006) Contrasting modes of evolution between vertebrate sweet/-umami receptor genes and bitter receptor genes. *Mol Biol Evol* 23(2):292–300.
43. Ishii Y, et al. (2008) Mutations in R-spondin 4 (*RSPO4*) underlie inherited anonychia. *J Invest Dermatol* 128(4):867–870.

44. Khan TN, et al. (2012) Novel missense mutation in the *RSPO4* gene in congenital hyponychia and evidence for a polymorphic initiation codon (p.M1I). *BMC Med Genet* 13:120.
45. Axelsson E, et al. (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.
46. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
47. Koboldt DC, et al. (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22(3):568–576.
48. Danecek P, et al.; 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
49. Alhaddad H, et al. (2013) Extent of linkage disequilibrium in the domestic cat, *Felis silvestris catus*, and its breeds. *PLoS ONE* 8(1):e53537.
50. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
51. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.
52. Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23(14):1801–1806.
53. Rosenberg NA (2003) DISTRUCT: A program for the graphical display of population structure. *Mol Ecol Notes* 4(1):137–138.
54. Albert FW, et al. (2009) Genetic architecture of tameness in a rat model of animal domestication. *Genetics* 182(2):541–554.
55. 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.
56. Menotti-Raymond M, et al. (1999) A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). *Genomics* 57(1):9–23.
57. Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81(11):3443–3446.
58. Schäffer AA (1996) Faster linkage analysis computations for pedigrees with loops or unused alleles. *Hum Hered* 46(4):226–235.
59. Young AE, Biller DS, Herrgesell EJ, Roberts HR, Lyons LA (2005) Feline polycystic kidney disease is linked to the *PKD1* region. *Mamm Genome* 16(1):59–65.
60. 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.
61. Bighignoli B, et al. (2007) Cytidine monophospho-N-acetylneuraminic acid hydroxylase (CMAH) mutations associated with the domestic cat AB blood group. *BMC Genet* 8:27.
62. Gandolfi B, et al. (2012) First *WNK4*-hypokalemia animal model identified by genome-wide association in Burmese cats. *PLoS ONE* 7(12):e53173.
63. Toonen RJ, Hughes S (2001) Increased throughput for fragment analysis on an ABI PRISM 377 automated sequencer using a membrane comb and STRand software. *Biotechniques* 31(6):1320–1324.
64. Rambaldi D, Ciccarelli FD (2009) FancyGene: Dynamic visualization of gene structures and protein domain architectures on genomic loci. *Bioinformatics* 25(17):2281–2282.
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.

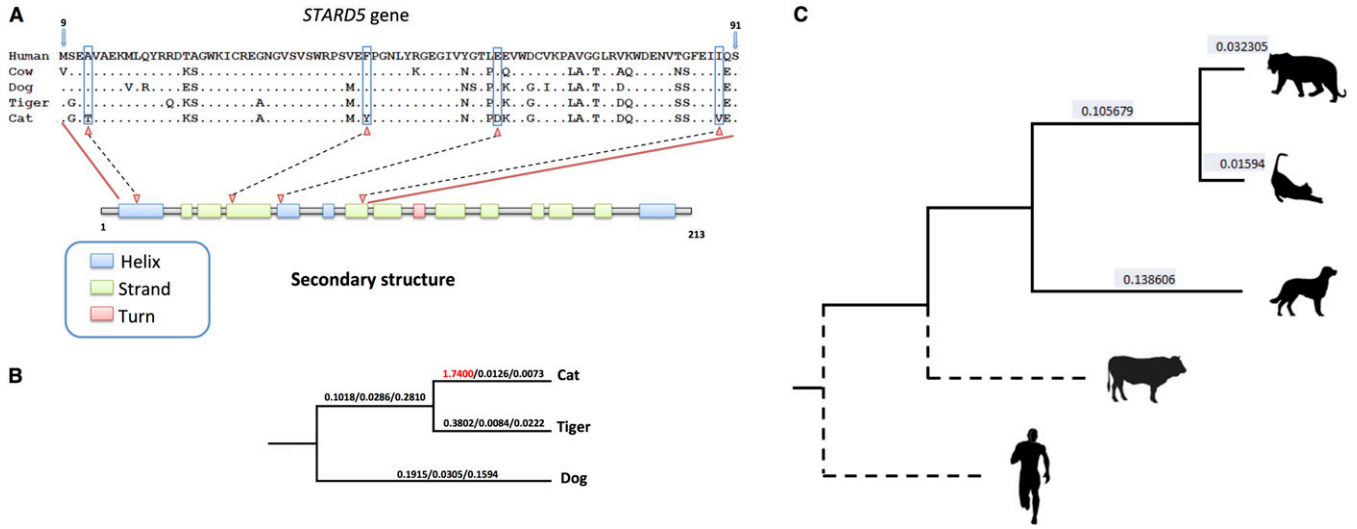

Fig. S1. (*A*) Predicted structure of the domestic cat *STARD5* gene. Positively selected amino acid sites are indicated with red arrowheads. (*B*) Results of the d_N/d_S test suggest an accelerated evolutionary rate of the *STARD5* gene on the domestic cat branch. Numbers on each branch are scores of the estimated d_N/d_S, d_N, and d_S. (*C*) Average synonymous mutation rates along branches used for assessments of positive selection. Dashed lines indicate relationships since rates are not reported for cow and human.

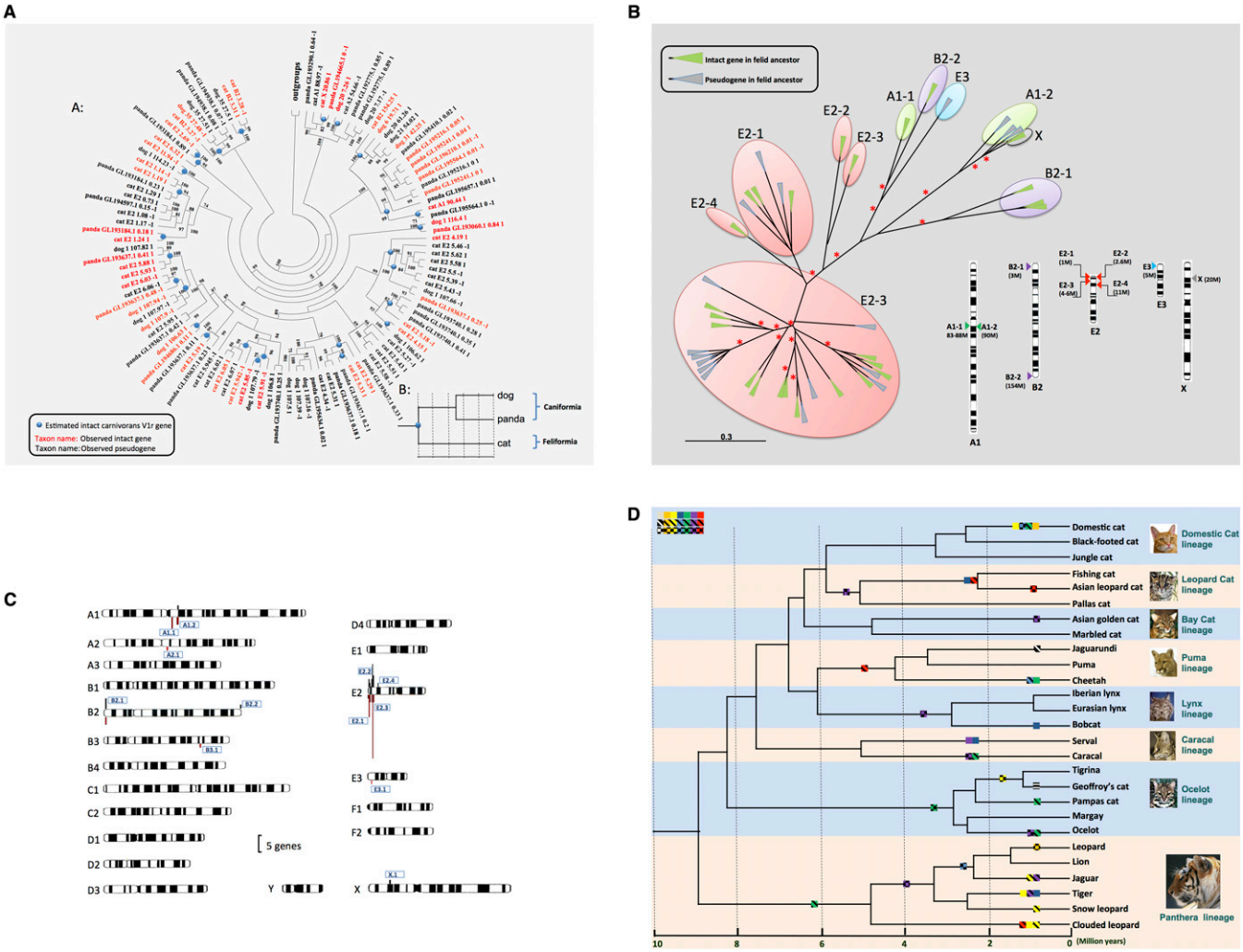

Fig. S2. (*A*) Maximum likelihood phylogenetic tree of functional *V1r* genes and long pseudogenes from the domestic cat, giant panda, and dog genomes. (*B*) Gene tree of the *V1r* supergene family determined for 35 feline species suggesting the early expansion of *V1r* genes in the common ancestor of felids. (*C*) Distribution of the *V1r* gene family in the domestic cat genome. Intact genes are denoted in black; pseudogenes are denoted in red. (*D*) Detected *V1r* gene loss among different lineages of the Felidae. Colored boxes in the tree indicate gene loss events based on 48 putatively intact *V1r* genes present in the common ancestor of all current felids.

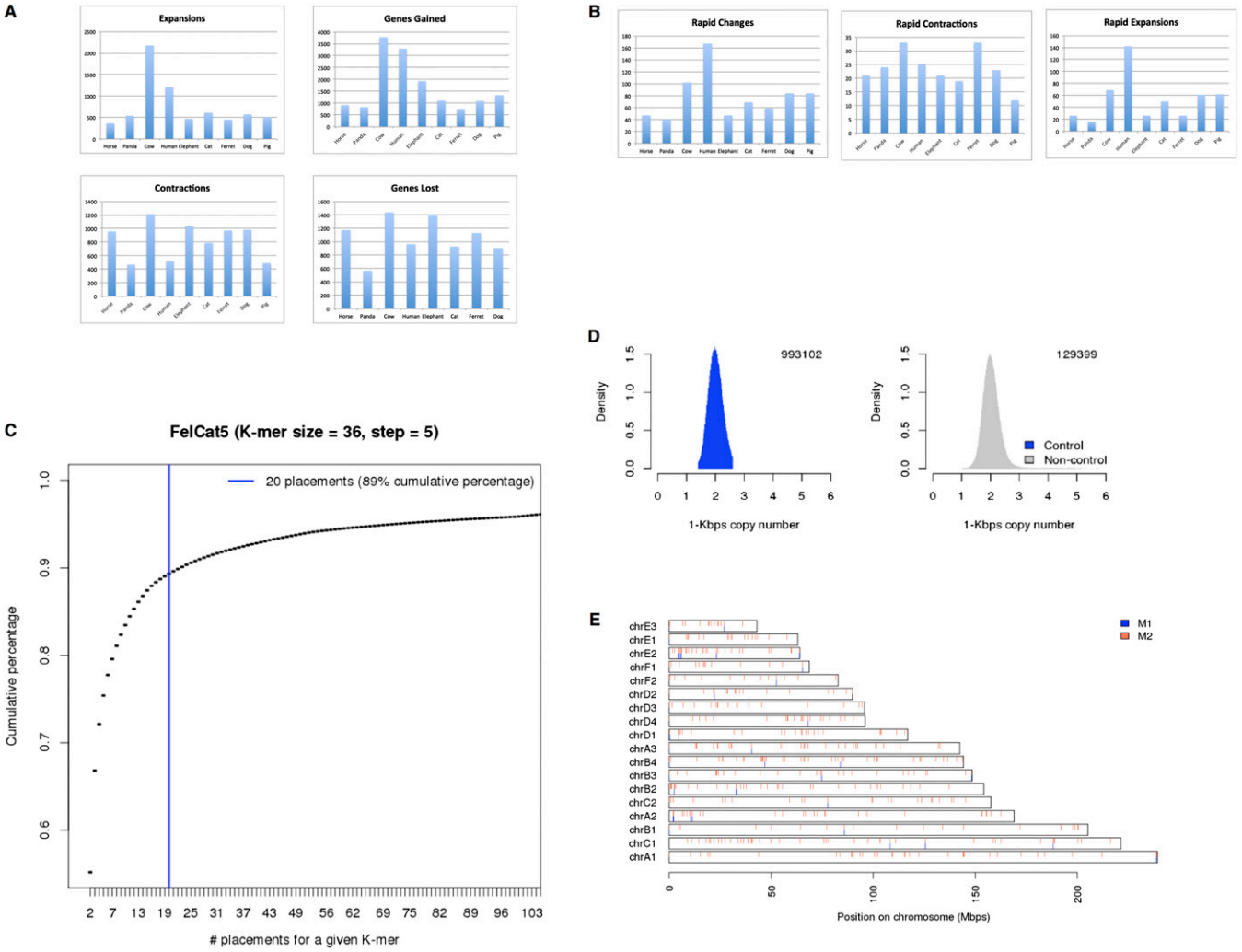

Fig. S3. (*A*) Raw counts for the number of gene family expansions, number of genes gained, number of gene family contractions, and number of genes lost among horse, panda, cow, human, elephant, cat, ferret, dog, and pig. (*B*) Significant results for the number of rapid gene family expansions and the number of rapid gene family contractions among horse, panda, cow, human, elephant, cat, ferret, dog, and pig. (*C*) Cumulative distribution of additional masking achieved by masking overrepresented K-mers in Fca 6.2 (FelCat5 in UCSC). (*D*) Distribution of 1-kbp copy number values in control and noncontrol regions. The number of windows in each distribution is indicated. (*E*) CNV map of expansions on domestic cat autosomes.

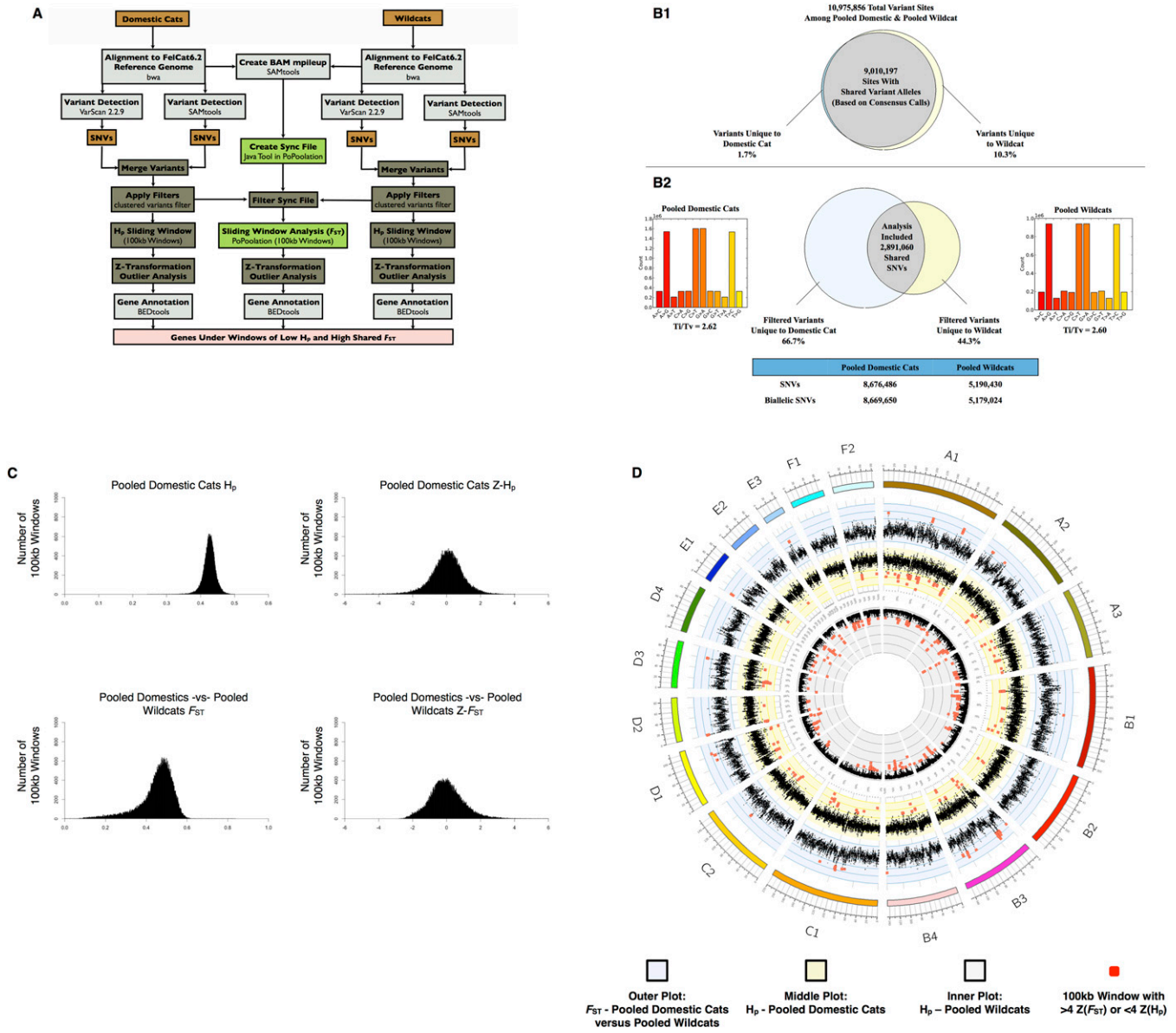

Fig. S4. (*A*) Analysis pipeline for determining putative regions of selection using variant data. (*B*) Summary results of variant calling for pooled domestic cats ($n = 22$) and pooled wildcats ($n = 4$). (*B1*) Percentage of overlapping variant alleles at each of the sites where a high-quality variant was detected. (*B2*) Percentage of unique and overlapping variant sites included in the sliding window analysis comparing domestic cats with wildcats based on stringent filtering parameters. Also included are transition:transversion ratios per pool as well as counts of variant types per pool. (*C*) Distribution of pooled heterozygosity, H_P, and average fixation index, F_{ST}, and corresponding Z transformations, Z(H_P) and Z(F_{ST}), estimated in 100-kb windows across all cat autosomes. (*D*) Circos plot of (*i*) pooled domestic cat versus pooled wildcat F_{ST}, (*ii*) pooled domestic cat H_p, and (*iii*) pooled wildcat H_p results for each 100-kb window (with a step size of 50 kb) along each chromosome. Windows with elevated F_{ST} or depressed H_p are depicted as red dots, whereas all other windows are depicted as black dots.

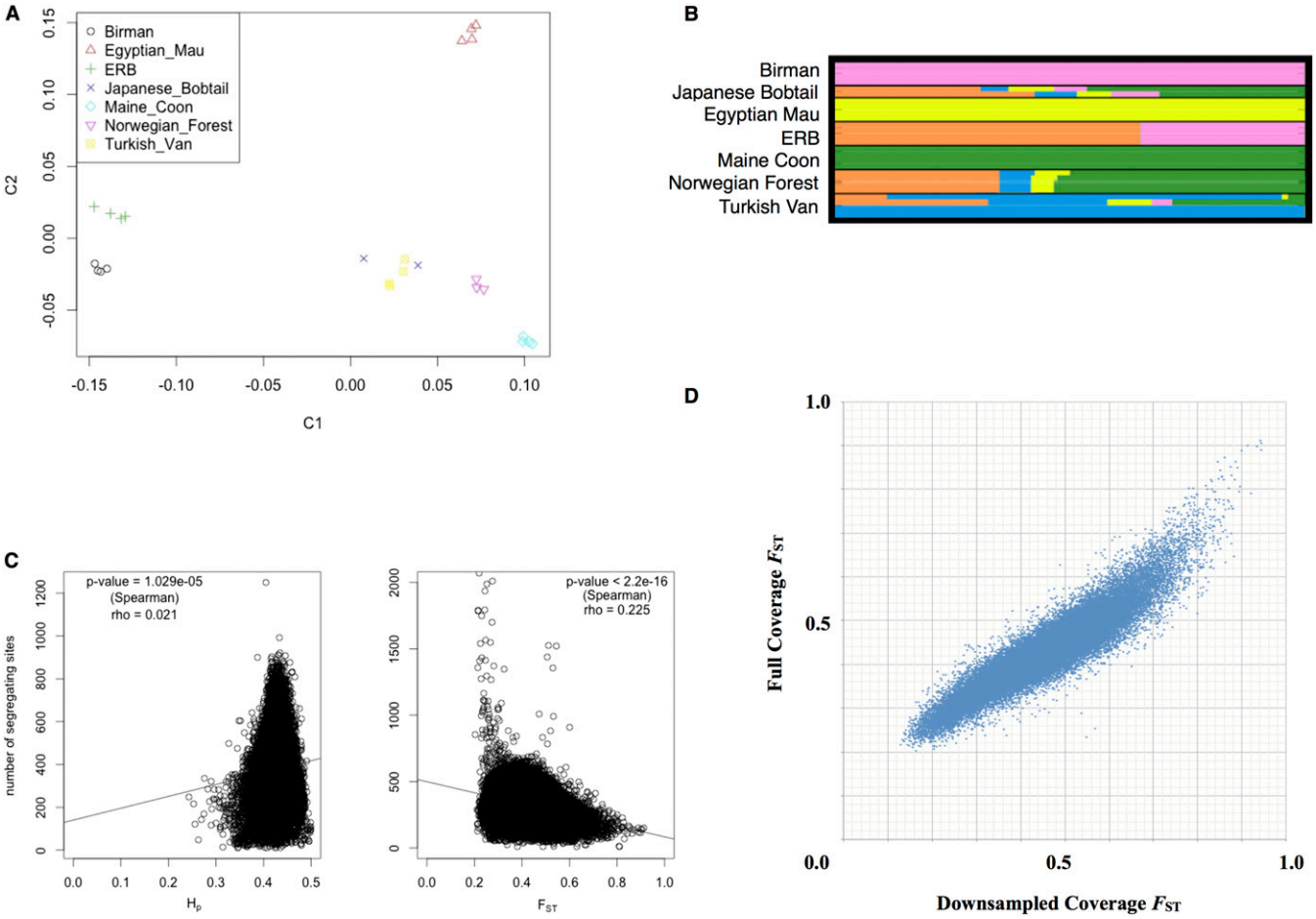

Fig. S5. (*A*) MDS plot depicting the relationship between individuals within the seven domestic cat pools used for the analysis of breed differentiation. (*B*) Admixture results for K = 5 showing genetic differentiation between eastern (Birman) and western (Maine Coon) populations, with moderate admixture between other breeds, including eastern random bred (ERB) individuals. (*C*) The average H_P and F_{ST} of all autosomal 100-kb windows plotted against the corresponding number of segregating sites per window. H_P is positively correlated with the number of segregating sites, whereas F_{ST} is negatively correlated. (*D*) The F_{ST} results for all autosomal 100-kb windows for the full coverage (~55×) pooled domestic analysis (*x* axis) are plotted against the F_{ST} results for all autosomal 100-kb windows for the subsampled (~7×) pooled domestic analysis (*y* axis).

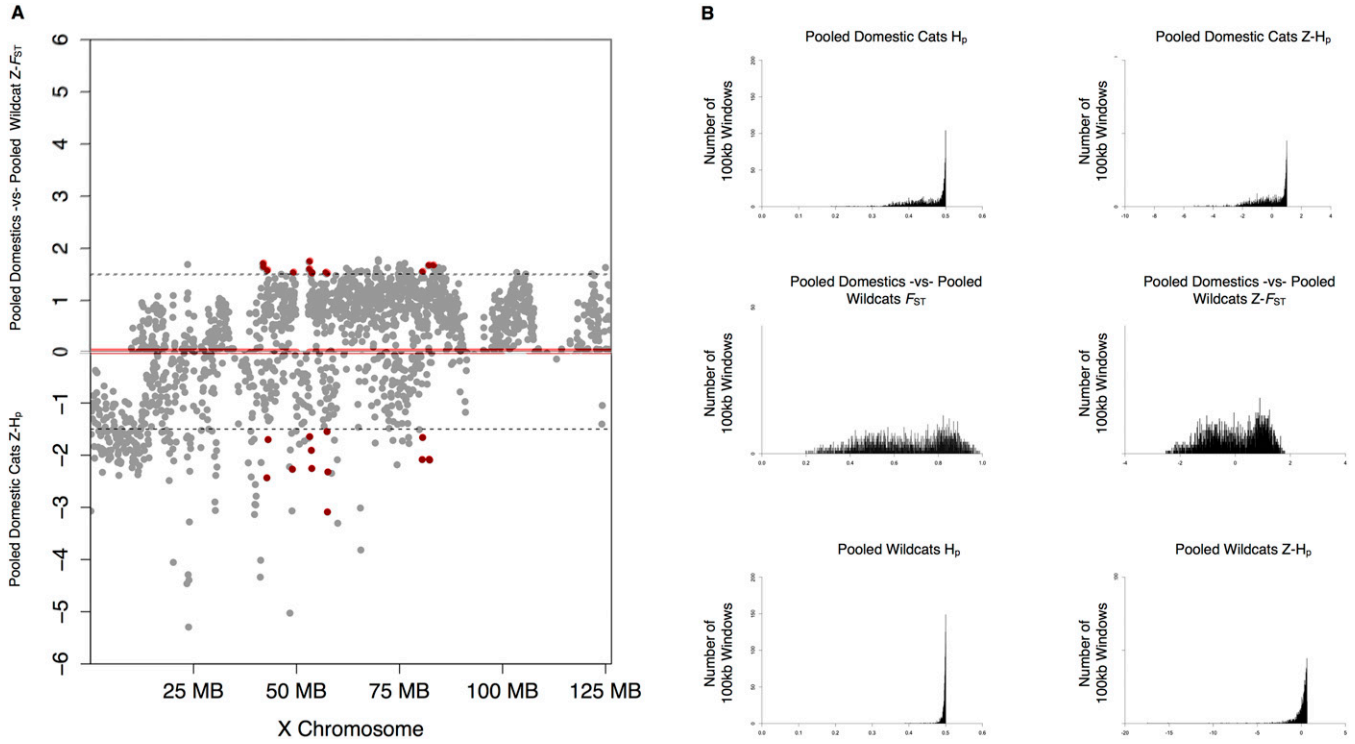

Fig. S6. (*A*) Z-transformed average fixation index (only positive values are shown), Z(F_{ST}), and pooled heterozygosity (only negative values are shown), Z(H_p), in 100-kb windows across chromosome X. Red dots indicate windows with (*i*) high F_{ST} and low H_p along with (*ii*) underlying gene content. (*B*) Distribution of pooled heterozygosity, H_p, and average fixation index, F_{ST}, and corresponding Z transformations, Z(H_p) and Z(F_{ST}), estimated in 100-kb windows across chromosome X.

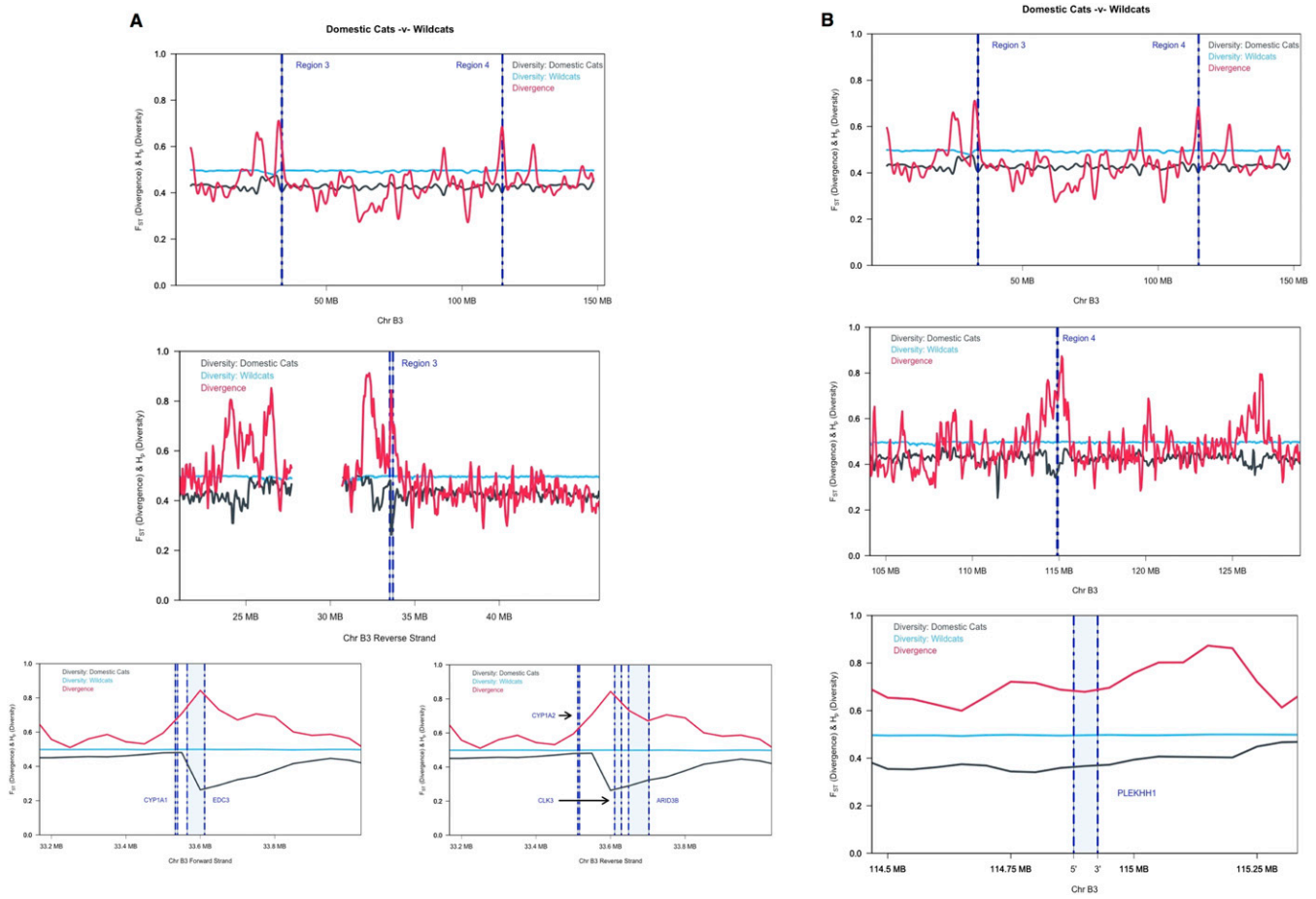

Fig. S7. (*A*) Plots of (*i*) pooled domestic cat versus pooled wildcat F_{ST} results (red), (*ii*) pooled domestic cat H_p results (light blue), and (*iii*) pooled wildcat H_p results (dark blue) for each 100-kb window (with a step size of 50 kb) along chromosome B3 with increasing resolution to the genes underlying region 3. (*B*) Plots of (*i*) pooled domestic cat versus pooled wildcat F_{ST} results (red), (*ii*) pooled domestic cat H_p results (light blue), and (*iii*) pooled wildcat H_p results (dark blue) for each 100-kb window (with a step size of 50 kb) along chromosome B3 with increasing resolution to the genes underlying region 4.

Other Supporting Information Files

Dataset S1 (PDF)
Dataset S2 (PDF)