NSU Florida
NOVA SOUTHEASTERN UNIVERSITY

Nova Southeastern University
NSUWorks

Biology Faculty Articles

Department of Biological Sciences

6-1-2016

A High-Resolution SNP Array-Based Linkage Map Anchors a New Domestic Cat Draft Genome Assembly and Provides Detailed Patterns of Recombination

Gang Li
Texas A&M University

LaDeana W. Hillier
Washington University School of Medicine in St. Louis

Robert A. Grahn
University of Missouri-Columbia

Aleksey V. Zimin
University of Maryland at College Park

Victor A. David
National Cancer Institute at Frederick

See next page for additional authors

Follow this and additional works at: https://nsuworks.nova.edu/cnso_bio_facarticles

Part of the Genetics and Genomics Commons

NSUWorks Citation

Li, Gang; LaDeana W. Hillier; Robert A. Grahn; Aleksey V. Zimin; Victor A. David; Marilyn Menotti-Raymond; Rondo Middleton; Steven S. Hannah; Sher L. Hendrickson; Alexey Makunin; Stephen J. O'Brien; Patrick Minx; Richard K. Wilson; Leslie A. Lyons; Wesley C. Warren; and William J. Murphy. 2016. "A High-Resolution SNP Array-Based Linkage Map Anchors a New Domestic Cat Draft Genome Assembly and Provides Detailed Patterns of Recombination." *G3* 6, (6): 1607-1616. doi:g3.116.028746v1.

This Article is brought to you for free and open access by the Department of Biological Sciences at NSUWorks. It has been accepted for inclusion in Biology Faculty Articles by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.

Authors
Gang Li, LaDeana W. Hillier, Robert A. Grahn, Aleksey V. Zimin, Victor A. David, Marilyn Menotti-Raymond, Rondo Middleton, Steven S. Hannah, Sher L. Hendrickson, Alexey Makunin, Stephen J. O'Brien, Patrick Minx, Richard K. Wilson, Leslie A. Lyons, Wesley C. Warren, and William J. Murphy

This article is available at NSUWorks: https://nsuworks.nova.edu/cnso_bio_facarticles/827

A High-Resolution SNP Array-Based Linkage Map Anchors a New Domestic Cat Draft Genome Assembly and Provides Detailed Patterns of Recombination

Gang Li,[*,1] LaDeana W. Hillier,[†,1] Robert A. Grahn,[‡,§] Aleksey V. Zimin,[**] Victor A. David,[††] Marilyn Menotti-Raymond,[††] Rondo Middleton,[‡‡] Steven Hannah,[‡‡] Sher Hendrickson,[§§,*] Alex Makunin,[***] Stephen J. O'Brien,[††,***] Pat Minx,[†] Richard K. Wilson,[†] Leslie A. Lyons,[‡,§] Wesley C. Warren,[†,2] and William J. Murphy[*,2]**

*Department of Veterinary Integrative Biosciences, Interdisciplinary Program in Genetics, Texas A&M University, College Station, Texas 77843, †McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri 63108, ‡Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, University of Missouri-Columbia, Missouri 65201, §Population Health and Reproduction, University of California-Davis, California 95616, **Institute for Physical Sciences and Technology, University of Maryland, College Park, Maryland 20742, ††National Cancer Institute-Frederick, National Institutes of Health, Maryland 21702, ‡‡Nestlé Purina PetCare Company, St. Louis, Missouri 63134, §§Department of Biology, Shepherd University, Shepherdstown, West Virginia 25443, and ***Theodosius Dobzhansky Center for Genome Bioinformatics, Saint Petersburg State University, 199004 St. Petersburg, Russia

ABSTRACT High-resolution genetic and physical maps are invaluable tools for building accurate genome assemblies, and interpreting results of genome-wide association studies (GWAS). Previous genetic and physical maps anchored good quality draft assemblies of the domestic cat genome, enabling the discovery of numerous genes underlying hereditary disease and phenotypes of interest to the biomedical science and breeding communities. However, these maps lacked sufficient marker density to order thousands of shorter scaffolds in earlier assemblies, which instead relied heavily on comparative mapping with related species. A high-resolution map would aid in validating and ordering chromosome scaffolds from existing and new genome assemblies. Here, we describe a high-resolution genetic linkage map of the domestic cat genome based on genotyping 453 domestic cats from several multi-generational pedigrees on the Illumina 63K SNP array. The final maps include 58,055 SNP markers placed relative to 6637 markers with unique positions, distributed across all autosomes and the X chromosome. Our final sex-averaged maps span a total autosomal length of 4464 cM, the longest described linkage map for any mammal, confirming length estimates from a previous microsatellite-based map. The linkage map was used to order and orient the scaffolds from a substantially more contiguous domestic cat genome assembly (*Felis catus* v8.0), which incorporated ~20 × coverage of Illumina fragment reads. The new genome assembly shows substantial improvements in contiguity, with a nearly fourfold increase in N50 scaffold size to 18 Mb. We use this map to report probable structural errors in previous maps and assemblies, and to describe features of the recombination landscape, including a massive (~50 Mb) recombination desert (of virtually zero recombination) on the X chromosome that parallels a similar desert on the porcine X chromosome in both size and physical location.

KEYWORDS
Felis catus
Illumina 63K SNP array
genetic map
recombination
domestic cat

During the past decade, increasingly detailed gene maps and sequence assemblies of the domestic cat genome have provided tools for mapping and discerning the heritable basis of morphological variation and genetic diseases (Murphy *et al.* 2007; Pontius *et al.* 2007; Bach *et al.* 20012; Davis *et al.* 2009; Menotti-Raymond *et al.* 2009; Mullikin *et al.* 2010; Montague *et al.* 2014). These include mutations in functional genes that control coat morphology (*e.g.*, Lyons *et al.* 2005; Schmidt-Küntzel *et al.* 2005; Ishida *et al.* 2006; Gandolfi *et al.* 2013a,b; Montague *et al.* 2014; Lyons 2015), as well as those that are causative for monogenic diseases such as hypertrophic cardiomyopathy (Meurs *et al.* 2005), progressive

retinal atrophy in different breeds (Menotti-Raymond *et al.* 2007; Alhaddad *et al.* 2014), and many others (see Gandolfi and Alhaddad 2015; Lyons 2015 for current reviews). The ability to narrow candidate regions and identify causative mutations in the feline model is currently limited by the quality of the domestic cat genome assembly and an inadequate genetic variant database.

The recent development of a moderate-density (~63,000 marker) Illumina SNP array has transitioned the mapping of feline traits from family-based linkage approaches to GWAS for simple and polygenic traits (Gandolfi *et al.* 2012, 2013a,b; Alhaddad *et al.* 2014; Davis *et al.* 2015). Here, we demonstrate the utility of the 63K SNP array for generating a high-density genetic linkage map based on genotyping data from several large multigenerational domestic cat pedigrees. High-density linkage maps have been produced for a number of mammals, including human (Kong *et al.* 2002), mouse (Shifman *et al.* 2006; Cox *et al.* 2009), cattle (Arias *et al.* 2009), pig (Tortereau *et al.* 2012), and dog (Wong *et al.* 2010). These resources provide valuable recombination information that complement genome sequence data, allowing for an improved understanding of the landscape of genetic linkage in mapping and association studies.

Prior feline genetic maps have been constructed from genotyping microsatellite markers in both interspecies backcross pedigrees (Menotti-Raymond *et al.* 1999, 2003) and domestic cat multigenerational pedigrees (Grahn *et al.* 2005; Menotti-Raymond *et al.* 2009; Schmidt-Küntzel *et al.* 2005). The average marker density of the latter map was relatively low (483 microsatellite markers, ~1 marker/9 Mb), and possessed some discrepancies with feline radiation hybrid maps (Menotti-Raymond *et al.* 2009; Davis *et al.* 2009). Nonetheless, the existing resources have provided crucial information for the initial genetic mapping of most feline phenotypic traits and diseases. Here, we describe the production of high-density sex-averaged recombination maps for each domestic cat chromosome based on genotyping data from the Illumina 63K feline SNP array. These high-resolution maps were then used in the generation of an improved whole genome assembly that was not heavily reliant on human–dog conserved gene order for scaffolding. The new assembly incorporated 20 × coverage of Illumina reads into the data used for the previous version (v6.2, Montague *et al.* 2014) genome assembly (felCat5). The new maps and assembly correct a number of structural errors in the previous assembly, and allow for the first detailed examination of recombination rate patterns across the domestic cat genome.

MATERIALS AND METHODS

Pedigrees and genotyping

Three independent pedigrees of domestic cats were used in the calculation of the recombination maps. The first is a five-generation extended pedigree of cats maintained by Nestlé Purina Company, consisting of 309 nonbreed cats (Menotti-Raymond *et al.* 1999), 305 of which were successfully genotyped on the array. The second pedigree had founders of Persian, Oriental Shorthair, Bengal, Japanese Bobtail, American Curl, and Somali that were crossbred to maximize diversity while retaining diseases and traits of interest. The colony was bred from 1999–2013, producing a nine-generation crossbreed pedigree. We included 165 genotyped cats from this second colony in our analysis. Genomic DNA was extracted from whole blood or tissue using either standard phenol chloroform extraction or the DNeasy Blood and Tissue Kit (Qiagen). DNA from all cats were genotyped on the Illumina feline 63K SNP BeadChip. Only SNPs with call rates > 0.95 and no replication errors were included in the construction of the final maps. Samples with < 0.90 call rate were removed from the analysis. Following Mendelian inheritance checks, 453 final cats were evaluated in the final linkage analysis.

Linkage analysis

We used the software package CRI-MAP (Green *et al.* 1990) v2.504 to build linkage maps for each chromosome using the combined Illumina 63K SNP genotype data derived from each of the 453 domestic cats. We removed 2365 markers that were included in the array design that assay felid lineage-specific apomorphies and phenotype-associated mutations, were uninformative, or possessed poor genotyping quality across samples. The *two-point* and *autogroup* options were used to cluster SNP markers into linkage groups based on pairwise LOD scores among all informative markers, enforcing a LOD score threshold of 8.0 to ensure no spurious cross-chromosome instances of linkage. The chromosomal assignments for each SNP marker were cross-checked and validated by BLAST (Altschul *et al.* 1990) comparison to the v6.2 genome assembly (Davis *et al.* 2015; Li *et al.* 2016).

Given the large number of markers assayed on the SNP array for each chromosome, we constructed an initial framework map that was then used to successively position more markers at less stringent thresholds of marker inclusion and number of informative meiosis (IM). The *build* option of CRI-MAP was used to construct a framework map for each chromosome/linkage group using only markers with > 300 IM and a LOD threshold of 3.0. Enforcing this LOD 3.0 framework marker order, we then used a sliding window-based approach to position additional markers from each chromosome relative to the framework marker order, because it was computationally prohibitive to position all remaining markers across an entire chromosome in a single analysis. Using the marker coordinates from the v6.2 assembly, we selected 500 markers in consecutive windows moving from the p terminus to the q terminus of each chromosome, for positioning within the LOD-3 chromosome framework maps. Successive windows on each chromosome contained 100 overlapping markers to ensure contiguity of marker order. For each window, markers were added relative to the framework map using successively less-stringent LOD thresholds of 2.0, 1.0, and 0.5. The minimum IM for the latter stages was 100. Markers that could not be positioned at LOD > 0.5 were positioned into their most likely marker order (minimum LOD = 0.1) but were not given a position in the final map. We used the *flips* option (6 marker windows per calculation) to evaluate and refine local marker ordering at each step of the process. We also iteratively examined maps for putative "expander" markers (those that inflated maps > 3 cM), identified by larger-than-average flanking marker intervals. Local maps were rebuilt following removal of each expander marker, and map distances recalculated and compared with the previous map. In the final stage, we placed all remaining candidate markers relative to their most likely marker interval. The final maps and source data may be found in Supplemental Material, File S1 and Table S1.

Copyright © 2016 Li *et al.*
doi: 10.1534/g3.116.028746
Manuscript received December 22, 2015; accepted for publication March 23, 2016; published Early Online March 29, 2016.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Supplemental material is available online at www.g3journal.org/lookup/suppl/doi:10.1534/g3.116.028746/-/DC1
[1]These authors contributed equally to this work.
[2]Corresponding authors: Veterinary Integrative Biosciences, Texas A&M University, 4458 TAMU, College Station, TX 77843-4458. E-mail: wmurphy@cvm.tamu.edu; McDonnell Genome Institute, Washington University School of Medicine, 4444 Forest Park Ave., St Louis, MO 63108. E-mail: wwarren@genome.wustl.edu

Table 1 Comparison of previous domestic cat radiation hybrid and linkage maps with the new SNP array-based linkage map (see Table S1)

	Radiation Hybrid Map (Davis *et al.* 2009)				Microsatellite Genetic Linkage Map (Menotti-Raymond *et al.* 1999; Schmidt-Küntzel *et al.* 2005)			63K SNP Array Genetic Linkage Map (Present Study)			
Chromosome	v8.0 Physical Length (Mb)	Markers	Unique Map Positions	Marker/Mb[a]	Markers	Map Length (cM)	cM/Mb[a]	Total Markers/Framework Markers	Total Markers/Mb	Map Length (cM)	cM/Mb[b]
A1	246	246	129	1.0	40	337.5	1.4	6219/217	25.3	414.6	1.7
A2	180	205	127	0.9	32	235.7	1.1	4529/208	25.2	315.8	1.8
A3	143	155	97	0.9	26	230.2	1.5	3162/175	22.1	275.5	1.9
B1	198	190	104	1.0	48	367.2	1.9	4636/210	23.4	305.4	1.5
B2	148	140	90	1.1	25	218.8	1.6	3612/186	24.4	267.2	1.8
B3	143	144	96	1.0	20	218.8	1.5	2934/172	20.5	278.0	1.9
B4	138	148	88	0.9	27	293.5	2.0	3633/194	26.3	276.0	2.0
C1	220	197	125	1.1	41	430.2	2.2	5677/208	25.8	424.1	1.9
C2	148	158	98	0.9	28	230.4	1.5	3726/175	25.2	244.0	1.6
D1	123	133	83	0.9	31	228.8	1.7	3129/172	25.4	231.5	1.9
D2	103	105	58	1.0	27	221.4	2.1	2279/167	22.1	208.5	2.0
D3	103	114	80	0.9	19	146.4	1.3	2685/144	26.1	199.6	1.9
D4	95	109	69	0.9	18	167.2	1.5	2334/172	24.6	216.5	2.3
E1	95	109	62	0.9	18	194.4	1.8	1659/128	17.5	204.9	2.2
E2	78	82	52	0.9	21	186.9	2.3	1736/130	22.3	159.6	2.0
E3	60	69	50	0.9	9	112.4	1.6	1259/85	21.0	126.9	2.1
F1	75	102	56	0.7	17	150.3	1.5	1865/125	24.9	158.3	2.1
F2	75	84	47	0.9	34	172.4	2.1	2248/156	30.0	157.1	2.1
X	128	172	91	0.7	46	228.4	1.3	733/82	5.7	161.0/109[c]	1.3/0.85[c]
Total/average	2493	2662	1602	0.9	527	4370.9	1.6	58,055/2204	24.0[d]	4624.5	1.9

SNP, single nucleotide polymorphism.
[a] Based on comparison to the felCat3 (1.9X) genome assembly.
[b] Based on comparison to the v8.0 genome assembly.
[c] Excluding the pseudoautosomal region.
[d] Autosomes only.

Table 2 Domestic cat v8.0 genome assembly statistics compared to v6.2

Assembly Version	Assembled Size	N50 Contig Size	N50 Scaffold Size	Total Gap Size	Reference
F. catus v6.2	2.36 Gb	21 kb	4.7 Mb	40 Mb	Montague *et al.* 2014
F. catus v8.0	2.64 Gb	45 kb	18 Mb	41 Mb	Present study

Construction of the v8.0 genome assembly

To construct a new genome assembly we utilized all existing genomic data generated from Cinnamon, a female Abyssinian cat used for all prior genome assemblies (Pontius *et al.* 2007; Montague *et al.* 2014), with the exception of felCat4 (Mullikin *et al.* 2010), which also included reads from multiple breeds and a wild cat (*Felis silvestris lybica*). The data from prior maps included ~2 × whole genome coverage of Sanger-based sequencing (6.7 million plasmid and 1.3 million 40 kb fosmid end reads; Pontius *et al.* 2007), ~12 × whole genome coverage with 454 sequencing (6 × fragment and 6 × of 3 kb paired-end reads), and Sanger-based end-sequenced BACs (Amplicon Express *Felis catus* FSCC library) (Montague *et al.* 2014). Here, we generated ~20 × coverage of nonoverlapping 100 bp paired-end reads from a single Illumina short-insert (avg. length = 350 bp) library, prepared from Cinnamon's DNA, on the HiSeq2000 (SRA Accession numbers SRX478589 and SRX478590). The new Illumina reads were combined with all prior sequence reads from Cinnamon (described above) to generate a new *de novo* assembly (v7.0) using the MaSuRCA assembler version 1.9.2 (Zimin *et al.* 2013). To ensure minimal loss of prior assembly sequence data, we merged the v6.2 assembly with the new v7.0 MaSuRCA assembly to create the final assembly, v8, using GAA (Yao *et al.* 2012), deposited under accession number GCA_000181335.3. A final processing step was deployed to fill small (< 1 kb) within-scaffold gaps using custom scripts to align and assemble Illumina reads.

To build individual chromosomes, we used BLASTN (E-value of 1e-9, with percent identity threshold = 85%) to align flanking probe sequences for the Illumina SNP markers (used in the linkage map construction) to the v8.0 genome assembly scaffolds. In addition, the assembled cat genome was broken into 1 kb segments and aligned against the v6.2 domestic cat assembly, the dog genome (canFam3), and the human genome (hg19) using BLAT (Kent 2002) with identity threshold = 85%, to precisely identify uniquely aligning segments of the cat genome that aided in identifying syntenic breakpoints down to very fine resolution. The new v8.0 genome scaffolds were ordered and oriented along 1) the v6.2 cat chromosomes, to check for larger scale scaffolding errors and or alignment inconsistencies, and 2) the new linkage map to validate, and in some cases modify, their order and orientation along the entire chromosome.

BLASTZ and BLAT alignments of the v8.0 assembly to the dog and human genomes were then used to refine the order and orientation information, as well as to insert additional scaffolds that may not have been tagged by markers on the linkage map. These new scaffolds were placed into the conditional scaffold framework provided by the initial v6.2/v8.0 cat alignments and marker assignments. Alignments between the cat, dog, and human genomes were used to more precisely identify the breakpoints within scaffolds when marker and alignment information confirmed a false join within the genome assembly. In addition, 98 finished BAC clone sequences (primarily selected from the FCAB library made from Cinnamon, with 8 MHC-spanning clones from the RPCI-86 library) were also incorporated into the genome assembly. Satellite sequences were identified in the v8.0 assembly, and centromeres were placed along each chromosome using the localization data from the radiation hybrid (RH) map of Davis *et al.* (2009).

Data availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

RESULTS

Recombination map

We removed 35 markers with Mendelian inheritance violations, and 2478 markers with fewer than 100 IM during the marker ordering step. The final maps included 58,055 markers, including 8402 unique map positions at LOD ≥ 0.5 that were ordered relative to an initial framework map of 2204 LOD > 3.0 markers distributed across the 18 autosomes and X chromosome. 6659 of the 8402 markers were positioned at LOD > 1.0 and 4978 at LOD > 2.0. The final map provided an ordered and positioned density of ~24 markers/Mb (Table 1).

The sex-averaged length of the autosomal map was 4464 cM, just 2% longer than the 4370 cM map length based on microsatellite genotyping (Menotti-Raymond *et al.* 2009). Calculating map length from only 2204 LOD > 3.0 framework markers also resulted in a similarly long map estimate of 4048 cM, suggesting our results were not heavily biased by inclusion of more than 6000 additional markers positioned at lower LOD thresholds (Table 1). The overall map lengths and chromosome-wide recombination rates were very similar to previous estimates based on comparison to a lower quality genome assembly (Menotti-Raymond *et al.* 2009). Female autosomal map length (5857 cM) was 1.7 × longer than the male map length (3415 cM), consistent with expectations from previous feline linkage maps (Menotti-Raymond *et al.* 2009).

The sex-averaged recombination rate across the genome was 1.9 cM/Mb, with rates varying from a high of 2.3 cM/Mb on chromosome D4, to a low of 0.85 cM/Mb on the X chromosome outside of the pseudoautosomal region (PAR) (Table 1). Recombination rates were highest near telomeric and subtelomeric regions when compared to centromeres, which typically contained the lowest recombination rates for each biarmed chromosome (Figure 2). Similar patterns of increased recombination near chromosome ends have been reported in other mammals (*e.g.*, Paigen *et al.* 2008; Wong *et al.* 2010). However, this pattern was not as apparent on the acrocentric cat chromosomes F1 and F2 (Figure 3).

The cat X chromosome possessed a large recombination coldspot that spans ~50 Mb, including the centromere. Notably, this same coldspot was "painted" in FISH mapping experiments that hybridized multiple, individual BAC clones selected from within this general region of the X chromosome to feline metaphase spreads (Davis *et al.* 2009). In contrast, FISH results from more than 100 individual BAC clones from outside of the X coldspot and on other autosomes, produced only discrete FISH signals on individual autosomes or distal portions of Xp and Xq, but not a paint-like FISH hybridization signal (Davis *et al.* 2009) (Figure 3). We therefore interpret this painting effect to be due to the presence of an unidentified repetitive gene family or enrichment of a class of repetitive elements within these X-specific BAC clones, which are unique to or highly enriched within the central portion of the X chromosome.

A new domestic cat genome assembly

The new v8.0 genome assembly incorporates ~2.56 Gb of sequence (excluding Ns in gaps) that were ordered and oriented on the 18 autosomes

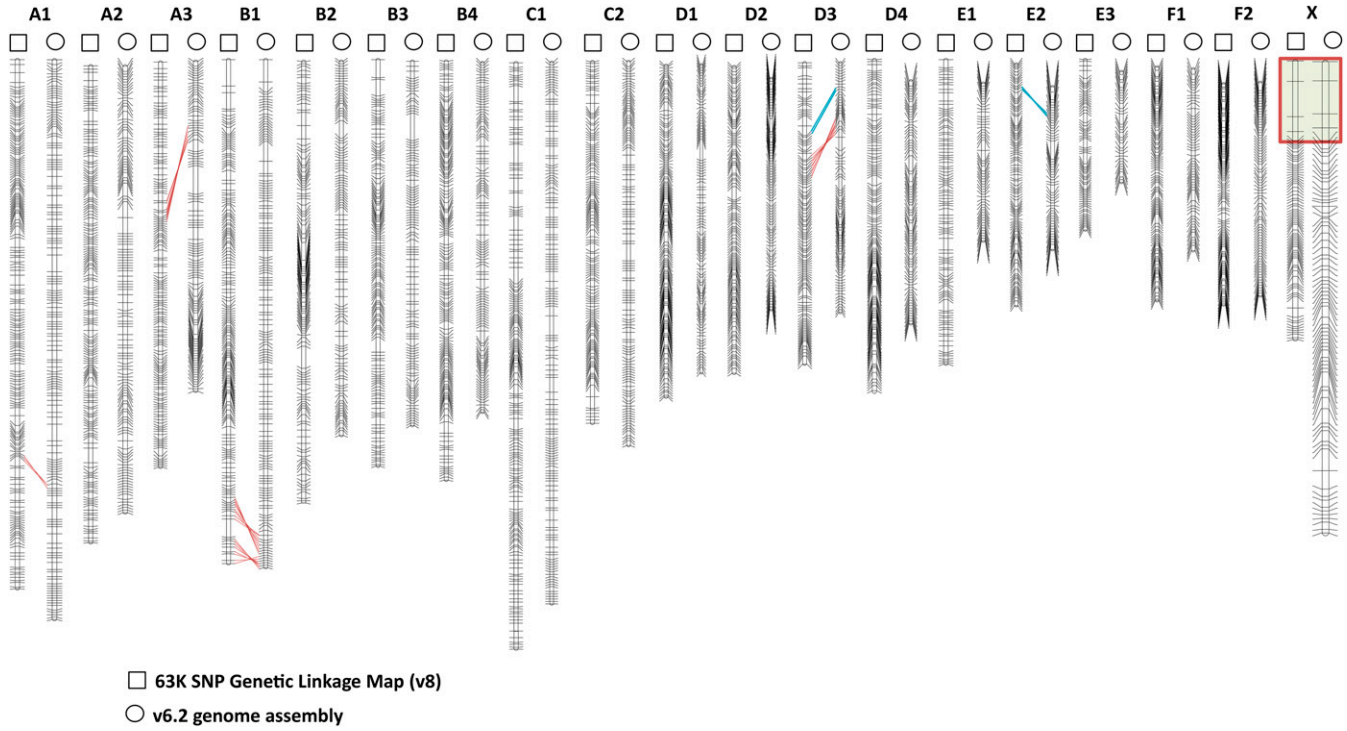

Figure 1 Domestic cat SNP linkage map (indicated above by squares) comparison with the v6.2 genome assembly (indicated above by circles). Black horizontal lines indicate framework marker (LOD 3.0) positions on both maps. Discordant marker orders are displayed between the maps with red (inversions) and blue (translocations) lines. The red box on the X chromosome maps indicate the pseudoautosomal region markers. LOD, logarithm of the odds; SNP, single nucleotide polymorphism.

and X chromosome (Table 2). An additional 146 Mb of scaffolds have probable chromosome assignments (chr*_random), and ~76 Mb of unplaced scaffolds have no known assignment (chr_UN). The size of the UN is likely an underestimate as these scaffolds are generally short, repetitive sequences that have been collapsed and have high sequence read depth. In total, this constitutes a 12% increase in assembly length, although the total gap length remained nearly the same (41 *vs.* 40 Mb) because more sequence was incorporated into the assembly. The v8.0 assembly contig and scaffold N50 lengths were 45 kb and 18 Mb, respectively, an increase of nearly 2.1 × and 3.8 × over the v6.2 assembly (Table 2).

We examined the integrity of the v8/linkage map chromosome marker orders by aligning the SNP-based linkage map with published RH and microsatellite-based linkage maps, which were previously used to guide the assembly of the domestic genome sequence (v6.2) (Table 1). The new high-density linkage map confirmed the larger scale order of markers in the v6.2 assembly (Montague *et al.* 2014) but also identified several structural differences with the v6.2 assembly (Figure 1). These include five inversions (chr. A1, chr. A3, two on chr. B1, and chr. D3) and two translocations (chr. D3 and chr. E2).

DISCUSSION

Production of "draft" quality mammalian genome assemblies has grown tremendously over the last few years with reduced costs of next-generation sequence data production. Despite the increase in throughput, these initial "blueprints" are often insufficient for detailed genetic investigations of the molecular signatures of selection or disease-causing alleles of particularly complex traits (Alkan *et al.* 2011; Zhang *et al.* 2009; Kirby *et al.* 2013). Given the substantial progress to date in feline trait mapping, coupled with the availability of genomic resources, there is continued motivation to improve the quality of the cat reference genome to further facilitate mapping of simple and complex traits of biomedical, aesthetic, and evolutionary interest.

We constructed a high-resolution genetic linkage map of the domestic cat genome, based upon Illumina 63K SNP genotypes from 453 domestic cats in three multi-generational pedigrees. The 63K array was designed with 62,897 SNPs ascertained from six domestic cats and an African wild cat (*F. lybica*) (Mullikin *et al.* 2010). The average marker spacing on the array is ~1 SNP marker every 50 kb. The new SNP-based linkage map provides a > 20-fold increase in improvement with regard to marker density compared to previous linkage (Menotti-Raymond *et al.* 2009; Schmidt-Küntzel *et al.* 2005) and RH (Murphy *et al.* 2007; Davis *et al.* 2009) maps. In addition, the contiguity and accuracy of the current reference genome sequence has improved markedly over previous assembly versions used to estimate chromosome wide recombination rates (Menotti-Raymond *et al.* 2009), hence facilitating the resolution of finer-scale domestic cat recombination rates across the genome.

The new reference genome presented here incorporates 280 Mb of new sequence data into the assembly, and shows a substantial improvement of contiguity as a result of new sequence data and use of the high-resolution linkage map. Compared to v6.2, the v8.0 assembly showed 2.2 and 3.9-fold increases in contig and scaffold N50 lengths, respectively, and corrected ~7 larger misassemblies that may have impacted GWAS interpretation. These discrepancies may have resulted from parts of the v6.2 assembly that relied heavily on published lower-density RH and linkage maps (Figure 1). Furthermore, a majority of scaffolds in the v6.2 assembly were ordered and oriented based on extrapolations from conserved human–dog gene order because they were not tagged by a genetic marker from the existing lower resolution linkage and RH maps.

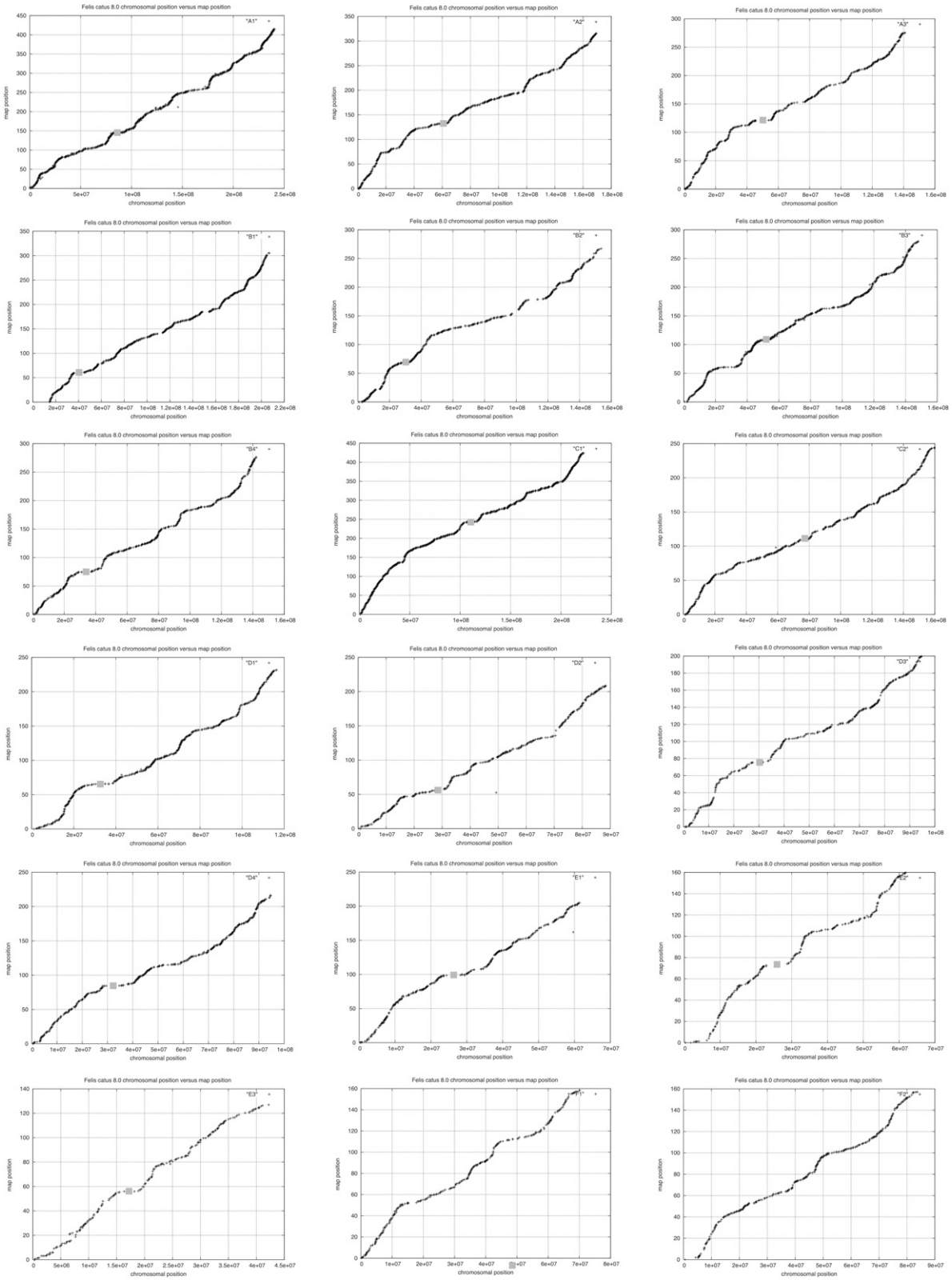

Figure 2 Plots of recombination distance (*y* axis, in cM) *vs.* physical distance (*x* axis, v8.0 assembly, in bp) for each of the 18 feline autosomes (identified in the upper right corner of each plot). Centromeres are indicated by a gray box.

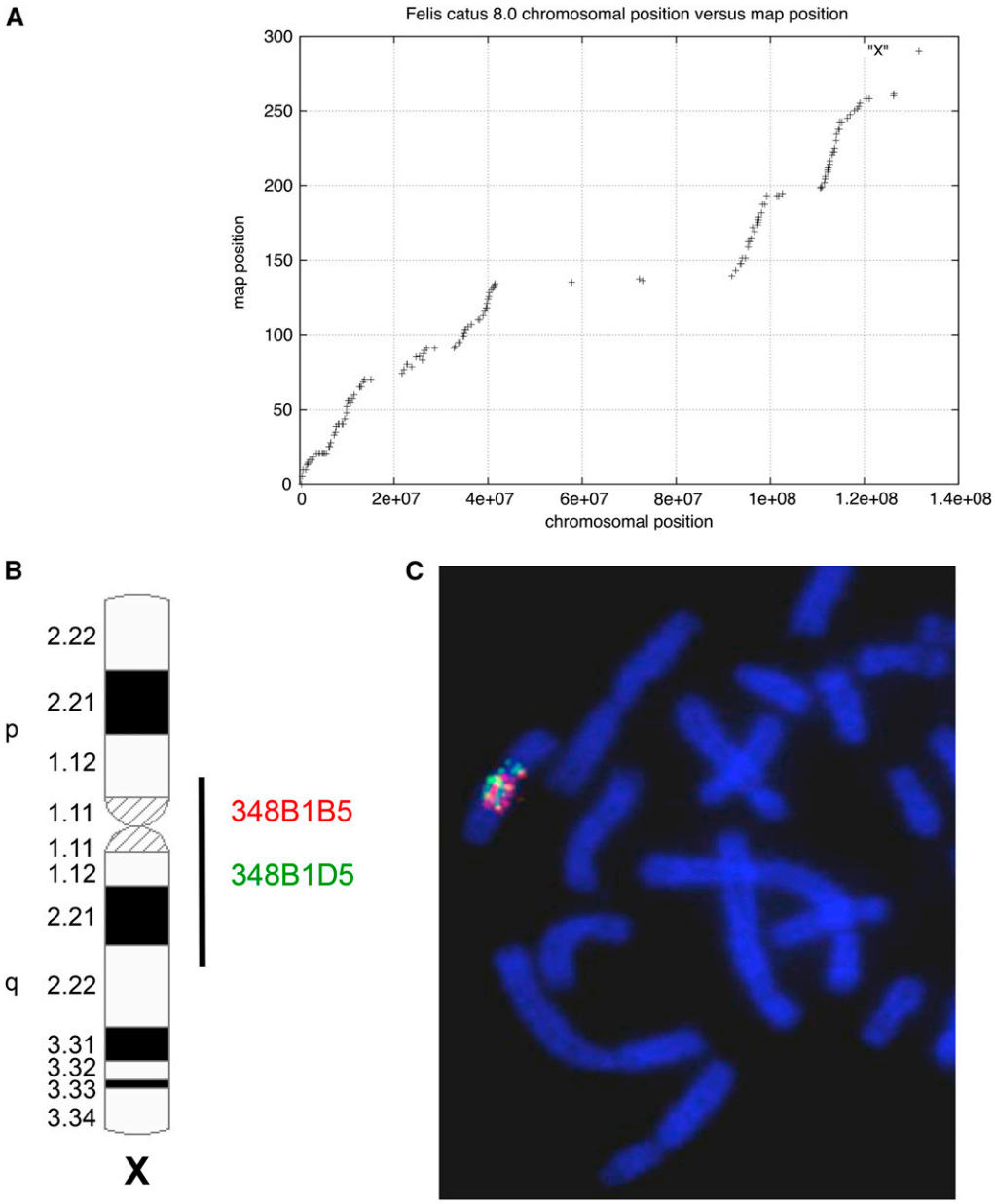

Figure 3 (A) Plot of recombination distance (*y* axis, in cM) *vs.* physical distance (*x* axis, v8.0 assembly, in bp) for the X chromosome. (B) Ideogram showing the distribution of hybridization signals of two randomly selected domestic cat BAC (bacterial artificial chromosome) clones (from Davis *et al.* 2009) derived from the centrally located X chromosome recombination desert. (C) Both clones paint the entire chromosomal region corresponding with the approximate boundaries of the recombination desert.

Recombination landscape

The detailed linkage map allowed for the first precise estimates of recombination rates across the genome. At 4464 cM, the domestic cat possesses the longest sex-averaged autosomal recombination map length of any eutherian mammal studied to date. This result is corroborated by independent cytogenetic estimates of meiotic rates of recombination in felids, which showed domestic cat and tiger possess the highest rates of any other analyzed eutherian mammals (Segura *et al.* 2013). Our new map length is also consistent with an independent and equally long estimate of 4370 cM observed by Menotti-Raymond *et al.* (2009), based on genotyping of 483 autosomal microsatellites on a smaller cohort of domestic cats. In addition to these maps, an earlier, less contiguous genetic linkage map calculated from genotyping 328 microsatellite loci in an interspecies hybrid backcross hybrid pedigree between the domestic cat and the Asian leopard cat (*Prionailurus bengalensis*) produced a map of 3300 cM, also relatively longer than linkage maps of most other eutherian species (Menotti-Raymond *et al.* 2003). The genome-wide sex-averaged recombination rate for the autosomes was 1.9 cM/Mb, also higher than all other reported mammalian genetic maps, and two to three times the rate of genetic maps for pig (0.85 cM/Mb, Tortereau *et al.* 2012), dog (0.97 cM/Mb, Wong *et al.* 2010), and mouse (0.63 cM/Mb, Shifman *et al.* 2006).

The distribution of recombination rates was quite variable along each chromosome. Recombination rates were reduced around most, but not all, centromeres, and increased in subtelomeric/telomeric regions (Figure 2). The most notable feature is an extremely long (~50 Mb) recombination desert (with virtually no recombination) on the center of chromosome X (Figure 3), in addition to two smaller recombination coldspots on Xp and distal Xq. These cold regions are flanked by recombination hotspots. A similarly large recombination coldspot, also flanked by high recombination rate regions, was described from the domestic pig X chromosome in a colinear region of conserved synteny (Ma *et al.* 2010; Ai *et al.* 2015). Furthermore, this region in cat and pig is colinear with the human X chromosome (Murphy *et al.* 1999; Davis *et al.* 2009; Ma *et al.* 2010); a

smaller but overlapping region surrounding the human X centromere also shows reduced recombination (Weeks *et al.* 1995).

Recently, Ai *et al.* (2015) and Li *et al.* (2016) observed similar signatures of interspecies genetic introgression within the large X chromosome recombination coldspot between high and low-altitude populations of pigs, and big cats of the genus *Panthera*, respectively. Both sets of authors posited that positive selection was a driving force behind the introgression patterns enriched on this region of the X chromosome, as natural selection is most effective in low-recombination regions (Smith and Haigh 1974). Selection is particularly effective on the X chromosome, where both beneficial and deleterious recessive alleles are exposed in hemizygous males (Charlesworth *et al.* 1987). Ai *et al.* (2015) reported enrichment of a thymidine-rich repeat within this region, and similarly, Davis *et al.* (2009) noted that BAC clones isolated from this region of the X chromosome also produced paint-like FISH signals on domestic cat chromosome spreads, implying a conserved X chromosome landmark across eutherian mammal orders (Figure 3).

Previous studies have hypothesized that selection during domestication events could potentially result in elevated recombination rates (Burt and Bell 1987; Ross-Ibarra 2004; Coop and Przeworski 2007), although recent studies have questioned this association (Muñoz-Fuentes *et al.* 2015). Indeed, in their analysis of selection signatures associated with cat domestication, Montague *et al.* (2014) observed a significant enrichment in genes involved in homologous recombination (*e.g.*, *BRCA2*, *RAD51B*, and *ZFYVE26*) within regions of high differentiation between wild and domestic cats. The authors also identified a putative case of adaptive linkage (*sensu*, Otto and Barton 1997) between two regulators of homologous recombination (*RAD51B* and *ZFYVE26*) and a candidate domestication gene (*PLEKHH1*). Considering these facts, together with the evidence presented here that the domestic cat recombination rate is higher than any other studied eutherian mammal, our data lend support to the hypothesis that the process of domestication may have driven an increase in the domestic cat's rate of recombination. Future analyses of chiasma frequencies in wild cat predecessors of the domestic cat would test this hypothesis.

Future prospects

The domestic cat genome assembly is unique among many other mammalian genome assemblies, being based on a large variety of sequencing read lengths, including a combination of legacy and NGS technology, and is anchored by a variety of physical mapping resources, including 2 × Sanger sequence data, multiple genetic linkage and RH maps, and fosmid and BAC-end sequences. This combination of mapping technologies has resulted in an extremely well-validated chromosome level assembly. These improvements to the assembly, coupled with whole genome SNP arrays, have led to a surge in trait and disease mapping efforts in cats, similar to advances in canine trait and disease mapping following the release of a high coverage draft assembly (Karlsson and Lindblad-Toh 2008).

The majority of the v8.0 draft genome sequence coverage is, nonetheless, primarily based on short NGS reads (454 and Illumina), which are known to inhibit the construction of highly contiguous *de novo* genome assemblies (Alkan *et al.* 2011; Bradnam *et al.* 2013, Sudmant *et al.* 2015, Chaisson *et al.* 2015a,b). As a result, we estimated that approximately 10–15% of the genome sequence (~480 million bp) remains in gaps or cannot be assigned to unique chromosomal positions at this time due to mapping ambiguity arising from repetitive structures, particularly segmental duplications. These regions are therefore not accessible to GWAS using the 63K array.

A growing body of evidence suggests that redundant genomic elements (*i.e.*, duplications) are important contributors to health-related phenotypes, possibly due to overlapping, interchangeable, or even novel tissue-specific functions of slightly differentiated, physically separated sequences that regulate gene expression (Zhang *et al.* 2009; Girirajan *et al.* 2012; Sudmant *et al.* 2010, 2015). Many phenotypes that are of interest to cat breeders, such as disease and disease-resistance, pattern and coloration, and reproduction, are regulated by genes that are members of gene families that are segmentally duplicated and copy number variable (Li *et al.* 2013, Montague *et al.* 2014). It is well known that these rapidly evolving, dynamic regions are often collapsed or even absent from draft genome assemblies (Church *et al.* 2009; Alkan *et al.* 2011; Chaisson *et al.* 2015a). These observations imply that a significant amount of feline phenotypic variation is likely to be modulated by genes or gene families imbedded within repetitive structures (*e.g.*, David *et al.* 2014, Davis *et al.* 2015). While the current genome assembly is a major step forward toward generating a more complete reference, the future incorporation of data from long single molecule sequencing (*e.g.*, Pacific Biosciences, 10 × Genomics) and other emerging scaffolding technologies (*e.g.*, BioNano Genomics and Dovetail Genomics maps) will lead to additional improvements in the quality and completeness of the domestic cat genome (English *et al.* 2012; Chaisson *et al.* 2015b, Steinberg *et al.* 2014).

ACKNOWLEDGMENTS

We thank the animal husbandry staff at the University of California–Davis and Paul Samollow (Texas A&M) for helpful discussions. We acknowledge funding from the Morris Animal Foundation/Cat Health Network (grants D12FE019 and D12FE502 to W.J.M., D12FE-501 to S.J.O. and M.M.R., and D12FE507 and D12FE560 to L.A.L.); the Winn Feline Foundation (W11-041 to L.A.L.); the University of California–Davis, School of Veterinary Medicine, Center for Companion Animal Health (2010-09-F to L.A.L.); the National Institutes of Health (NIH) National Center for Research Resources (L.A.L.); the NIH National Human Genome Research Institute (5U54HG00307907 to R.K.W.); the US Department of Agriculture, National Institute of Food and Agriculture, National Robotics Initiative (2009-35205-05209 and 2008-04049 to A.V.Z.); the NIH (R01-HG002945 and R01-HG006677 to A.V.Z.); and the Russian Ministry of Science (Megagrant no. 11.G34.31.0068 to S.J.O.). This research was supported (in part) by the Intramural Research Program of the NIH, Center for Cancer Research, National Cancer Institute-Frederick.

LITERATURE CITED

Ai, H., X. Fang, B. Yang, Z. Huang, H. Chen *et al.*, 2015 Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. Nat. Genet. 47: 217–225.

Alhaddad, H., B. Gandolfi, R. A. Grahn, H. C. Rah, C. B. Peterson *et al.*, 2014 Genome-wide association and linkage analyses localize a progressive retinal atrophy locus in Persian cats. Mamm. Genome 25(7–8): 354–362.

Alkan, C., S. Sajjadian, and E. E. Eichler, 2011 Limitations of next-generation genome sequence assembly. Nat. Methods 8(1): 61–65.

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

Arias, J. A., M. Keehan, P. Fisher, W. Coppieters, and R. Spelman, 2009 A high density linkage map of the bovine genome. BMC Genet. 10: 18.

Bach, L. H., B. Gandolfi, J. C. Grahn, L. V. Millon, M. S. Kent *et al.*, 2012 A high-resolution 15,000(Rad) radiation hybrid panel for the domestic cat. Cytogenet. Genome Res. 137(1): 7–14.

Bradnam, K. R., J. N. Fass, A. Alexandrov, P. Baranay, M. Bechner *et al.*, 2013 Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. Gigascience 2(1): 10.

Burt, A., and G. Bell, 1987 Mammalian chiasma frequencies as a test of two theories of recombination. Nature 326(6115): 803–805.

Chaisson, M. J., J. Huddleston, M. Y. Dennis, P. H. Sudmant, M. Malig *et al.*, 2015a Resolving the complexity of the human genome using single-molecule sequencing. Nature 517(7536): 608–611.

Chaisson, M. J., R. K. Wilson, and E. E. Eichler, 2015b Genetic variation and the *de novo* assembly of human genomes. Nat. Rev. Genet. 16: 627–640.

Charlesworth, B., J. A. Coyne, and N. H. Barton, 1987 The relative rates of evolution of sex-chromosomes and autosomes. Am. Nat. 130: 113–146.

Church, D. M., L. Goodstadt, L. W. Hillier, M. C. Zody, S. Goldstein *et al.*, 2009 Lineage-specific biology revealed by a finished genome assembly of the mouse. PLoS Biol. 7: e1000112.

Coop, G., and M. Przeworski, 2007 An evolutionary view of human recombination. Nat. Rev. Genet. 8(1): 23–34.

Cox, A., C. L. Ackert-Bicknell, B. L. Dumont, Y. Ding, J. T. Bell *et al.*, 2009 A new standard genetic map for the laboratory mouse. Genetics 182(4): 1335–1344.

David, V. A., M. Menotti-Raymond, A. C. Wallace, M. Roelke, J. Kehler *et al.*, 2014 Endogenous retrovirus insertion in the KIT oncogene determines white and white spotting in domestic cats. G3 (Bethesda) 4(10): 1881–1891.

Davis, B. W., T. Raudsepp, A. J. P. Wilkerson, R. Agarwala, A. A. Schaffer *et al.*, 2009 A high-resolution cat radiation hybrid and integrated FISH mapping resource for phylogenomic studies across Felidae. Genomics 93 (4): 299–304.

Davis, B. W., C. M. Seabury, W. Brashear, G. Li, M. Roelke Parker *et al.*, 2015 Mechanisms underlying mammalian hybrid sterility in two feline interspecies models. Mol. Biol. Evol. 32: 2534–2546.

English, A. C., S. Richards, Y. Han, M. Wang, V. Vee *et al.*, 2012 Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One 7(11): e47768.

Gandolfi, B., and H. Alhaddad, 2015 Investigation of inherited diseases in cats: genetic and genomic strategies over three decades. J. Feline Med. Surg. 17(5): 405–415.

Gandolfi, B., T. J. Gruffydd-Jones, R. Malik, A. Cortes, B. R. Jones *et al.*, 2012 First *WNK4*-hypokalemia animal model identified by genome-wide association in Burmese cats. PLoS One 7(12): e53173.

Gandolfi, B., H. Alhaddad, V. K. Affolter, J. Brockman, J. Haggstrom *et al.*, 2013a To the root of the curl: a signature of a recent selective sweep identifies a mutation that defines the Cornish Rex cat breed. PLoS One 8 (6): e67105.

Gandolfi,, B., H. Alhaddad, S. E. K. Joslin, R. Khan, S. Filler *et al.*, 2013b A splice variant in KRT71 is associated with curly coat phenotype of Selkirk Rex cats. Sci. Rep. DOI: .10.1038/srep02000

Girirajan, S., J. A. Rosenfeld, B. P. Coe, S. Parikh, N. Friedman *et al.*, 2012 Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N. Engl. J. Med. 367(14): 1321–1331.

Grahn, R. A., B. M. Lemesch, L. V. Millon, T. Matise, Q. R. Rogers *et al.*, 2005 Localizing the X-linked orange colour phenotype using feline resource families. Anim. Genet. 36(1): 67–70.

Green, P., K. Falls, and S. Crooks, 1990 *Documentation for CRI-MAP, Version 2.4*. Washington University School of Medicine, St. Louis.

Ishida, Y., V. A. David, E. Eizirik, A. A. Schaffer, B. A. Neelam *et al.*, 2006 A homozygous single-base deletion in MLPH causes the dilute coat color phenotype in the domestic cat. Genomics 88(6): 698–705.

Karlsson, E. K., and K. Lindblad-Toh, 2008 Leader of the pack: gene mapping in dogs and other model organisms. Nat. Rev. Genet. 9(9): 713–725.

Kent, W. J., 2002 BLAT—the BLAST-like alignment tool. Genome Res. 12 (4): 656–664.

Kirby, A., A. Gnirke, D. B. Jaffe, V. Barešová, N. Pochet *et al.*, 2013 Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in *MUC1* missed by massively parallel sequencing. Nat. Genet. 45(3): 299–303.

Kong, A., D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson *et al.*, 2002 A high-resolution recombination map of the human genome. Nat. Genet. 31(3): 241–247.

Li, G., B. W. Davis, T. Raudsepp, A. J. Pearks Wilkerson, V. C. Mason *et al.*, 2013 Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. Genome Res. 23: 1486–1495.

Li, G., B. W. Davis, E. Eizirik, and W. J. Murphy, 2016 Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). Genome Res. 26(1): 1–11.

Lyons, L. A., 2015 DNA mutations of the cat. The good, the bad and the ugly. J. Feline Med. Surg. 17(3): 203–219.

Lyons, L. A., D. L. Imes, H. C. Rah, and R. A. Grahn, 2005 Tyrosinase mutations associated with Siamese and Burmese patterns in the domestic cat (*Felis catus*). Anim. Genet. 36(2): 119–126.

Ma, J., N. Iannuccelli, Y. Duan, W. Huang, B. Guo *et al.*, 2010 Recombinational landscape of porcine X chromosome and individual variation in female meiotic recombination associated with haplotypes of Chinese pigs. BMC Genomics 11: 159.

Menotti-Raymond, M., V. A. David, L. A. Lyons, A. A. Schaffer, J. F. Tomlin *et al.*, 1999 A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). Genomics 57(1): 9–23.

Menotti-Raymond, M., V. A. David, M. E. Roelke, Z. Q. Chen, K. A. Menotti *et al.*, 2003 Second-generation integrated genetic linkage/radiation hybrid maps of the domestic cat (*Felis catus*). J. Hered. 94(1): 95–106.

Menotti-Raymond, M., V. A. David, A. A. Schaffer, R. Stephens, D. Wells *et al.*, 2007 Mutation in CEP290 discovered for cat, model of human retinal degeneration. J. Hered. 98(3): 211–220.

Menotti-Raymond, M., V. A. David, A. A. Schaffer, J. F. Tomlin, E. Eizirik *et al.*, 2009 An autosomal genetic linkage map of the domestic cat, *Felis silvestris catus*. Genomics 93(4): 305–313.

Meurs, K. M., X. Sanchez, R. M. David, N. E. Bowles, J. A. Towbin *et al.*, 2005 A cardiac myosin binding protein C mutation in the Maine Coon cat with familial hypertrophic cardiomyopathy. Hum. Mol. Genet. 14(23): 3587–3593.

Montague, M. J., G. Li, B. Gandolfi, R. Khan, B. L. Aken *et al.*, 2014 Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. Proc. Natl. Acad. Sci. USA 111(48): 17230–17235.

Mullikin, J. C., N. F. Hansen, L. Shen, H. Ebling, W. F. Donahue *et al.*, 2010 Light whole genome sequence for SNP discovery across domestic cat breeds. BMC Genomics 11: 406.

Muñoz-Fuentes, V., M. Marcet-Ortega, G. Alkorta-Aranburu, C. Linde Forsberg, J. M. Morrell *et al.*, 2015 Strong artificial selection in domestic mammals did not result in an increased recombination rate. Mol. Biol. Evol. 32(2): 510–523.

Murphy, W. J., S. Sun, J. Pecon-Slattery, Z. Q. Chen, and S. J. O'Brien, 1999 Extensive conservation of sex chromosome organization between cat and human revealed by parallel radiation hybrid mapping. Genome Res. 9: 1223–1230.

Murphy, W. J., B. Davis, V. A. David, R. Agarwala, A. A. Schaffer *et al.*, 2007 A 1.5-Mb-resolution radiation hybrid map of the cat genome and comparative analysis with the canine and human genomes. Genomics 89 (2): 189–196.

Otto, S. P., and N. H. Barton, 1997 The evolution of recombination: removing the limits to natural selection. Genetics 147(2): 879–906.

Paigen, K., J. P. Szatkiewicz, K. Sawyer, N. Leahy, E. D. Parvanov *et al.*, 2008 The recombinational anatomy of a mouse chromosome. PLoS Genet. 4(7): e1000119.

Pontius, J. U., J. C. Mullikin, D. R. Smith, K. Lindblad-Toh, S. Gnerre *et al.*, 2007 Initial sequence and comparative analysis of the cat genome. Genome Res. 17(11): 1675–1689.

Ross-Ibarra, J., 2004 The evolution of recombination under domestication: a test of two hypotheses. Am. Nat. 163(1): 105–112.

Schmidt-Küntzel, A., E. Eizirik, S. J. O'Brien, and M. Menotti-Raymond, 2005 Tyrosinase and tyrosinase related protein 1 alleles specify domestic cat coat color phenotypes of the albino and brown loci. J. Hered. 96(4): 289–301.

Segura, J., L. Ferretti, S. Ramos-Onsins, L. Capilla, M. Farré *et al.*, 2013 Evolution of recombination in eutherian mammals: insights into the mechanisms that affect recombination rates and crossover interference. Proc. Biol. Sci. 280: 20131945.

Shifman, S., J. T. Bell, R. R. Copley, M. S. Taylor, R. W. Williams *et al.*, 2006 A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol. 4(12): e395.

Smith, J. M., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 89: 391–403.

Steinberg, K. M., V. A. Schneider, T. A. Graves-Lindsay, R. S. Fulton, R. Agarwala *et al.*, 2014 Single haplotype assembly of the human genome from a hydatidiform mole. Genome Res. 24(12): 2066–2076.

Sudmant, P. H., J. O. Kitzman, F. Antonacci, C. Alkan, M. Malig *et al.*, 2010 Diversity of human copy number variation and multicopy genes. Science 330(6004): 641–646.

Sudmant, P. H., T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov *et al.*, 2015 An integrated map of structural variation in 2,504 human genomes. Nature 526(7571): 75–81.

Tortereau, F., B. Servin, L. Frantz, H. J. Megens, D. Milan *et al.*, 2012 A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. BMC Genomics 13: 586.

Weeks, D. E., T. G. Nygaard, M. Neystat, and L. D. Harby, 1995 A high-resolution genetic linkage map of the pericentromeric region of the human X chromosome. Genomics 26: 39–46.

Wong, A. K., A. L. Ruhe, B. L. Dumont, K. R. Robertson, G. Guerrero *et al.*, 2010 A comprehensive linkage map of the dog genome. Genetics 184 (2): 595–605.

Yao, G. H., L. Ye, H. Y. Gao, P. Minx, W. C. Warren *et al.*, 2012 Graph accordance of next-generation sequence assemblies. Bioinformatics 28(1): 13–16.

Zhang, F., W. L. Gu, M. E. Hurles, and J. R. Lupski, 2009 Copy number variation in human health, disease, and evolution. Annu. Rev. Genomics Hum. Genet. 10: 451–481.

Zimin, A. V., G. Marcais, D. Puiu, M. Roberts, S. L. Salzberg *et al.*, 2013 The MaSuRCA genome assembler. Bioinformatics 29(21): 2669–2677.

Communicating editor: D. J. de Koning