



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The ninth life of the cat reference genome, *Felis_catus*

Citation for published version:

Zhang, W & Schoenebeck, JJ 2020, 'The ninth life of the cat reference genome, *Felis_catus*', *PLoS Genetics*, vol. 16, no. 10, pp. e1009045. <https://doi.org/10.1371/journal.pgen.1009045>

Digital Object Identifier (DOI):

[10.1371/journal.pgen.1009045](https://doi.org/10.1371/journal.pgen.1009045)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

PLoS Genetics

Publisher Rights Statement:

Copyright: © 2020 Zhang, Schoenebeck. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



PERSPECTIVE

The ninth life of the cat reference genome, *Felis_catus*Wengang Zhang*, Jeffrey J. Schoenebeck¹

The Roslin Institute and Royal (Dick) School for Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, United Kingdom

* wengang.zhang@ed.ac.uk (WZ); jeff.schoenebeck@roslin.ed.ac.uk (JJS)

Few animal species are as storied and intertwined with human history as domestic cats, *Felis catus*. With an estimated 600 million cats living with humans, cats' popularity as pets is indisputable. The earliest hint of our relationship with felines come from Neolithic skeletal remains found in Cyprus, where a human and wildcat were co-interred some 9,500 years ago [1]. Five thousand years later, the human–cat bond would be proclaimed in ancient Egyptian iconography and burials (Fig 1A), as well as through the mitochondrial DNA of many contemporary cats whose mitotypes were traced back to Northern Africa. The dispersal of cats from Anatolia, the Levante, and Northern Africa coincided with human trade and agriculture [2,3]. Though companionship was probably welcome, it is predation of rodent pests that likely precipitated the union between man and cat [4].

Buckley and colleagues present new insight into the domestic cat genome and genetic variation in this issue of *PLOS Genetics* with the release of *Felis_catus_9.0*, the newest version of the cat reference genome [5]. With its dramatic improvements in both assembly contiguity and annotation, this reference represents a significant step forward for scientists and cat lovers interested in demography, evolution, domestication, and genomic medicine in companion animals.

Contemporary household cats share both our environment and our exposures to infectious diseases. As they grow old, many will succumb to age-related diseases whose names ring familiar: diabetes, lymphoma, kidney disease, cardiomyopathy, and dementia to name a few. For many cats, these morbidities are managed with varying degrees of success through access to state-of-the-art veterinary care. As companion animals, dogs also share many of these same attributes; however, the 2 species' genetic architectures are distinct: unlike most dogs that can be categorized to various degrees as belonging to a breed, the vast majority of cats are the products of random mating. As a result, the genetic composition of "random breeds," by far the most common type of household cat worldwide, is comparatively diverse and not uncommonly admixed with local subspecies of wildcats [1,6,7]. From this perspective, the genomic architecture of the cat is more akin to our own in terms of diversity and population stratification. For these reasons, cats have the potential to be a comparative model of human medicine and disease etiology.

In 2006, the International Cat Genome Sequencing Consortium (ICGSC) released ASM18133v3, an assembly produced from a female Abyssinian named "Cinnamon." This draft assembly was produced from Sanger-sequenced plasmid and fosmid libraries. At just 2x read coverage, the assembly was highly fragmented: only 50% of the genome was covered by contigs of lengths greater than 2.7 kb (a statistic also known as contig N50). It required 174,000 contigs to cover half the cat genome [8], and genome assembly and annotation depended heavily on radiation hybrid and comparative mapping to human and dog genomes.

OPEN ACCESS

Citation: Zhang W, Schoenebeck JJ (2020) The ninth life of the cat reference genome, *Felis_catus*. *PLoS Genet* 16(10): e1009045. <https://doi.org/10.1371/journal.pgen.1009045>

Editor: Gregory S. Barsh, HudsonAlpha Institute for Biotechnology, UNITED STATES

Published: October 22, 2020

Copyright: © 2020 Zhang, Schoenebeck. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: WZ receives funding from the China Scholarship Council. JJS receives funding from the Biotechnology and Biological Sciences Research Council (BBS/E/D/20211553, BBS/E/D/30002275). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

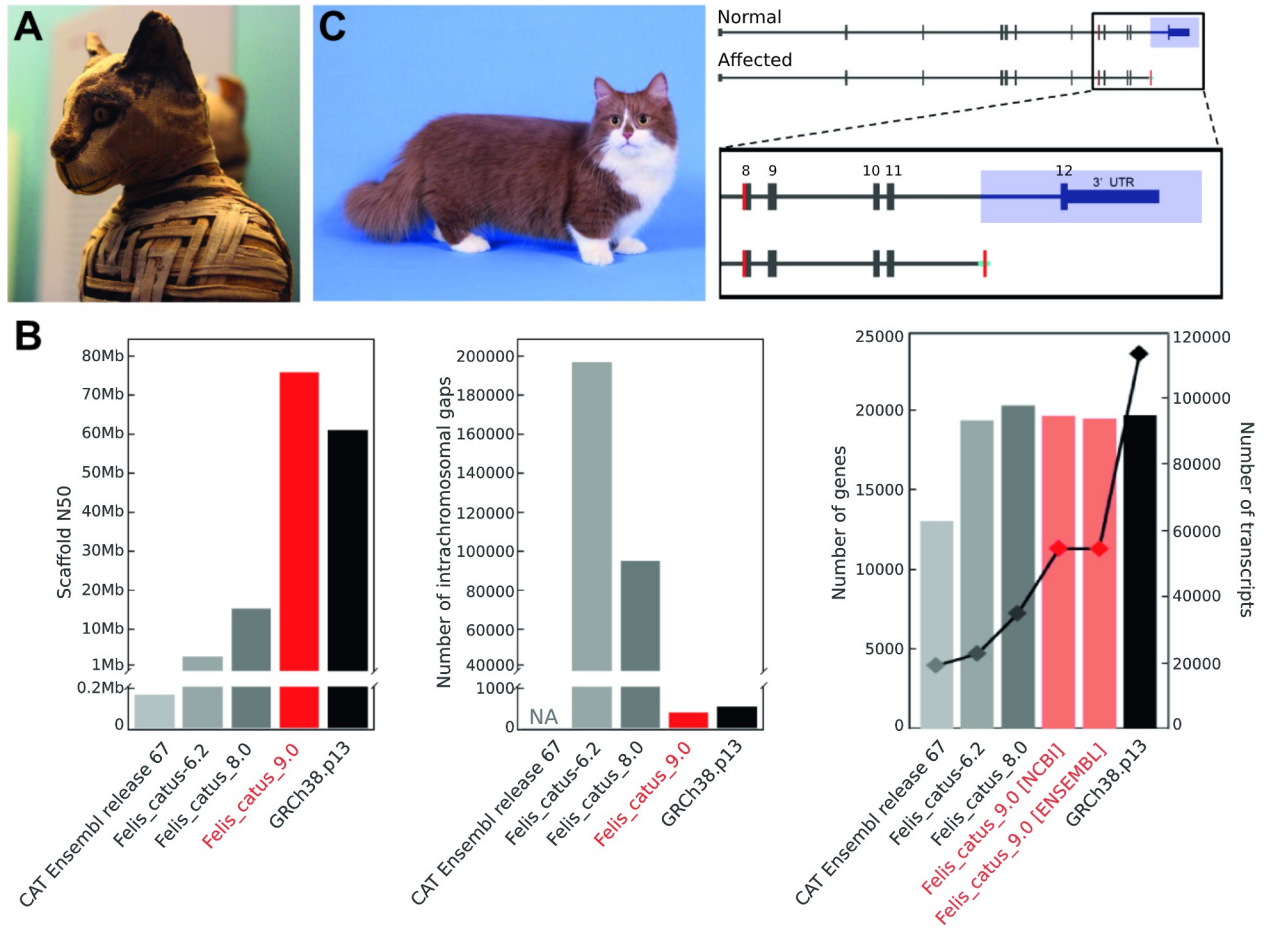


Fig 1. A summary of *Felis_catus_9.0*. (A) Egyptian mummified cat from the Roman period, housed at The British Museum. (B) Comparison of assemblies. GRCh38.p13 is the reference genome of *Homo sapiens*. (C) Asymmetric dwarfism observed in Munchkin breed cats is likely due to a 3.3 kb deletion in the *UGDH* gene [5]. *UGDH*, UDP-glucose 6-dehydrogenase. *Image credit: Justin Ennis, Flickr.*

<https://doi.org/10.1371/journal.pgen.1009045.g001>

The ICGSC has incrementally improved Cinnamon’s reference genome through releases rebranded as *Felis_catus*. *Felis_catus_5.0* and *Felis_catus_6.2* incorporated read data from bacterial artificial chromosome (BAC) end sequencing (2× coverage), 454 Titanium GS_FLX (12× coverage), and enhanced scaffolding, which was underpinned by improved radiation hybrid maps [9,10], together with some variant and methylation data [11]. In 2014, *Felis_catus_6.2* was supplanted by *Felis_catus_8.0*, which added 20× pooled Illumina short-read sequencing from wildcats and pedigree cats to vastly increase variant information. Scaffolding was also improved through yet another high-density radiation hybrid map [12]. However, the *Felis_catus* assembly remained stubbornly fragmented, more than 80× as much as its fellow carnivore, the dog, whose assembly benefitted from deeper Sanger sequencing (and capital) at the outset.

With the advent of long-read sequencing and optical mapping technologies, the ICGSC were among the first to apply them to mammalian genomes. Their efforts, embodied by *Felis_catus_9.0*, were released to the wider research community in November 2017. Low-passage fibroblasts grown from Cinnamon provided high molecular weight DNA that was key to maximizing the benefits of Pacific Biosystem’s long-read sequencing technology and Bionano Saphyr’s single-molecule optical mapping technology. This quality of this chromosome level

assembly rivaled other popular species including human (Fig 1B), mouse, rat, pig, cattle, and goat.

As described by Buckley and colleagues, *Felis_catus_9.0* is remarkably contiguous, with 4,909 contigs and an N50 of 42 Mb and exceptionally long gap-free segments (Fig 1B). The improved assembly facilitated production of gene models, a process whose ab initio predictions were refined by RNA-sequencing, which was used to profile numerous tissues. Improvements in genomic features also include the definition of noncoding genes, pseudogenes, and novel genes that were absent in previous versions of *Felis_catus*.

Buckley and colleagues also describe approximately 40,000,000 single-nucleotide variants (SNVs) and approximately 13,000,000 indels based on resequencing data from 74 animals. Beyond using annotation tools to predict the functional impact of these variants, the authors binned these variants by functional constraint: significantly fewer loss of function mutations were observed in genes with essential functions. Rather, constrained genes were differentially enriched for presumably benign synonymous variants. This examination, novel in its application to nonhuman species, provides additional granularity to uncovering disease-causing alleles. The authors also generated a comprehensive structure variation (SV) atlas. Over 200,000 SVs (insertions, deletions, inversions, and duplications) were identified, a 300-fold increase from *Felis_catus_6.2*. One of the SVs was discovered in resequenced Munchkins. It occurs on chromosome B1, within a 5.2 Mb region that was previously associated with this breed's disproportionate dwarfism [13]. The putatively causal allele removes the final exon of UDP-glucose 6-dehydrogenase (*UGDH*), a gene whose product is postulated to participate in proteoglycan synthesis within the articular cartilages of long bones (Fig 1C).

The improvements made to the cat reference genome are likely to yield many more biological insights that will impact medicine across many mammalian species. However, *Felis_catus_9.0* is but one of many steps that are needed to make genomic veterinary medicine common practice. Significant hurdles remain in the quest to deliver genomic medicine to cats and other companion animals: annotations that depict RNA isoforms are sparse; National Center for Biotechnology Information (NCBI) and ENSEMBL gene models often conflict; noncoding epigenetic features are lacking; Y chromosome assemblies are incomplete [14,15]; and the sequence contents of nucleolar organizer regions, telomeres, and centromeres are undefined. As more conspecific assemblies are produced [16], how will we integrate their information?

Companion animal cell lines and cell-based assays are needed to functionally validate variants suspected of causing disease. Clearly, future research investments and technological innovations are needed to create the repertoire of resources that are required to ensure that clinical DNA samples are interrogated exhaustively and interpreted correctly. Not to be overlooked, the costs of clinical sequencing are currently too high for most owners, especially in the absence of health insurance.

The dream of bringing genomic medicine to veterinary clinics must overcome these and other formidable challenges; scaling these challenges will require research team initiatives, project coordination, sustainable funding, and most importantly, communicating to funders and science policy makers the unmet need that our community can fill. To the latter point, we must unapologetically embrace and communicate that cats (and dogs) are more than pets; they are also unique clinical animal models and sentinels of human health.

References

1. Vigne JD, Guilaine J, Debue K, Haye L, Gerard P. Early taming of the cat in Cyprus. *Science*. 2004; 304:259. <https://doi.org/10.1126/science.1095335> PMID: 15073370

2. Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, et al. The Near Eastern origin of cat domestication. *Science*. 2007; 317:519–523. <https://doi.org/10.1126/science.1139518> PMID: 17600185
3. Ottoni C, Van Neer W, De Cupere B, Daligault J, Guimaraes S, Peters J, et al. The palaeogenetics of cat dispersal in the ancient world. *Nat Ecol Evol*. 2017; 1:103–107. <https://doi.org/10.1038/s41559-017-0103-0> PMID: 28812674
4. Clutton-Brock J. A natural history of domesticated mammals. 2nd ed. Cambridge; 1999.
5. Buckley RM, Davis BW, Brashear WA, Farias FHG, Kuroki K, Graves T, et al. A new domestic cat genome assembly based on long sequence reads empowers feline genomic medicine and identifies a novel gene for dwarfism. *PLoS Genet*. 2020; 16:e1008926. <https://doi.org/10.1371/journal.pgen.1008926>
6. Alhaddad H, Khan R, Grahn RA, Gandolfi B, Mullikin JC, Cole SA, et al. Extent of linkage disequilibrium in the domestic cat, *Felis silvestris catus*, and its breeds. Ellegren H, editor. *PLoS ONE*. 2013; 8: e53537. <https://doi.org/10.1371/journal.pone.0053537> PMID: 23308248
7. Lipinski MJ, Froenicke L, Baysac KC, Billings NC, Leutenegger CM, Levy AM, et al. The ascent of cat breeds: genetic evaluations of breeds and worldwide random-bred populations. *Genomics*. 2008; 91:12–21. <https://doi.org/10.1016/j.ygeno.2007.10.009> PMID: 18060738
8. Pontius JU, Mullikin JC, Smith DR, Agencourt Sequencing Team, Lindblad-Toh K, Gnerre S, et al. Initial sequence and comparative analysis of the cat genome. *Genome Res*. 2007; 17:1675–1689. <https://doi.org/10.1101/gr.6380007> PMID: 17975172
9. Montague MJ, Li G, Gandolfi B, Khan R, Aken BL, Searle SMJ, et al. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci U S A*. 2014; 111:17230–17235. <https://doi.org/10.1073/pnas.1410083111> PMID: 25385592
10. Davis BW, Raudsepp T, Pearks Wilkerson AJ, Agarwala R, Schäffer AA, Houck M, et al. A high-resolution cat radiation hybrid and integrated FISH mapping resource for phylogenomic studies across Felidae. *Genomics*. 2009; 93:299–304. <https://doi.org/10.1016/j.ygeno.2008.09.010> PMID: 18951970
11. Tamazian G, Simonov S, Dobrynin P, Makunin A, Logachev A, Komissarov A, et al. Annotated features of domestic cat—*Felis catus* genome. *Gigascience*. 2014; 3:13. <https://doi.org/10.1186/2047-217X-3-13> PMID: 25143822
12. Li G, Hillier LW, Grahn RA, Zimin AV, David VA, Menotti-Raymond M, et al. A high-resolution SNP array-based linkage map anchors a new domestic cat draft genome assembly and provides detailed patterns of recombination. *G3 (Bethesda)*. 2016; 6:1607–1616. <https://doi.org/10.1534/g3.116.028746> PMID: 27172201
13. Lyons LA, Fox DB, Chesney KL, Britt LG, Buckley RM, Coates JR, et al. Localization of a feline autosomal dominant dwarfism locus: a novel model of chondrodysplasia. *bioRxiv*. 2019; 48:687210. <https://doi.org/10.1101/687210>
14. Li G, Davis B, Raudsepp T, Pearks Wilkerson A, Mason V, Ferguson-Smith M, et al. Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Res*. 2013; 23:1486–1495. <https://doi.org/10.1101/gr.154286.112> PMID: 23788650
15. Brashear WA, Raudsepp T, Murphy WJ. Evolutionary conservation of Y Chromosome ampliconic gene families despite extensive structural variation. *Genome Res*. 2018; 28:1841–1851. <https://doi.org/10.1101/gr.237586.118> PMID: 30381290
16. Isobe S, Matsumoto Y, Chung C, Sakamoto M, Chan T-F, Hirakawa H, et al. AnAms1.0: a high-quality chromosome-scale assembly of a domestic cat *Felis catus* of American Shorthair breed. *bioRxiv*. 2020;138:2020.05.19.103788. <https://doi.org/10.1101/2020.05.19.103788>