

DATA NOTE

Open Access

# Annotated features of domestic cat – *Felis catus* genome

Gaik Tamazian<sup>1</sup>, Serguei Simonov<sup>1</sup>, Pavel Dobrynin<sup>1</sup>, Alexey Makunin<sup>1</sup>, Anton Logachev<sup>1</sup>, Aleksey Komissarov<sup>1</sup>, Andrey Shevchenko<sup>1</sup>, Vladimir Brukhin<sup>1</sup>, Nikolay Cherkasov<sup>1</sup>, Anton Svitin<sup>1</sup>, Klaus-Peter Koepfli<sup>1</sup>, Joan Pontius<sup>1</sup>, Carlos A Driscoll<sup>2</sup>, Kevin Blackstone<sup>2</sup>, Cristina Barr<sup>2</sup>, David Goldman<sup>2</sup>, Agostinho Antunes<sup>3</sup>, Javier Quilez<sup>4</sup>, Belen Lorente-Galdos<sup>5</sup>, Can Alkan<sup>6</sup>, Tomas Marques-Bonet<sup>5</sup>, Marylin Menotti-Raymond<sup>7</sup>, Victor A David<sup>7</sup>, Kristina Narfström<sup>8</sup> and Stephen J O'Brien<sup>1,9\*</sup>

## Abstract

**Background:** Domestic cats enjoy an extensive veterinary medical surveillance which has described nearly 250 genetic diseases analogous to human disorders. Feline infectious agents offer powerful natural models of deadly human diseases, which include feline immunodeficiency virus, feline sarcoma virus and feline leukemia virus. A rich veterinary literature of feline disease pathogenesis and the demonstration of a highly conserved ancestral mammal genome organization make the cat genome annotation a highly informative resource that facilitates multifaceted research endeavors.

**Findings:** Here we report a preliminary annotation of the whole genome sequence of Cinnamon, a domestic cat living in Columbia (MO, USA), bisulfite sequencing of Boris, a male cat from St. Petersburg (Russia), and light 30× sequencing of Sylvester, a European wildcat progenitor of cat domestication. The annotation includes 21,865 protein-coding genes identified by a comparative approach, 217 loci of endogenous retrovirus-like elements, repetitive elements which comprise about 55.7% of the whole genome, 99,494 new SNVs, 8,355 new indels, 743,326 evolutionary constrained elements, and 3,182 microRNA homologues. The methylation sites study shows that 10.5% of cat genome cytosines are methylated. An assisted assembly of a European wildcat, *Felis silvestris silvestris*, was performed; variants between *F. silvestris* and *F. catus* genomes were derived and compared to *F. catus*.

**Conclusions:** The presented genome annotation extends beyond earlier ones by closing gaps of sequence that were unavoidable with previous low-coverage shotgun genome sequencing. The assembly and its annotation offer an important resource for connecting the rich veterinary and natural history of cats to genome discovery.

**Keywords:** *Felis catus*, Domestic cat, *Felis silvestris silvestris*, European wildcat, Genome sequence, Annotation, Assembly

## Data description

The genome of a female Abyssinian cat (“Cinnamon” who resides at the University of Missouri-Columbia, USA) was sequenced at 1.8 × and 3.0 × whole genome shotgun (WGS) coverage at Agencourt Inc. Fca-6.2, an additional 12 × coverage of 454 reads and BAC ends was

sequenced, assembled with CABOG [1] and analysed at Washington University, St. Louis (USA) [2]. Fca-6.2 is anchored to chromosome coordinates with two physical framework maps, a radiation hybrid map [3] and a short tandem repeat (STR) linkage map [4]. Further, 1943 distinct sites identified in a recently built linkage map using a single nucleotide polymorphism (SNP) genotyping array including ≈ 60,000 SNPs from an Illumina custom cat genotyping array are also mapped to the assembly.

Here we present a genome browser, Genome Annotation Resource Fields — GARfield [5], which displays the Fca-6.2 assembly and included annotated genome

\*Correspondence: lgdchief@gmail.com

<sup>1</sup>Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, 199004, St. Petersburg, Russia

<sup>9</sup>Oceanographic Center, Nova Southeastern University, 33004 Ft Lauderdale, FL, USA

Full list of author information is available at the end of the article

**Table 1 Annotated cat genome features available as genome browser tracks for GARfield and UCSC genome browsers**

Feature	Additional file 1
I. Assembly of <i>Felis catus</i> genome Fca-6.2	
II. Gene annotation	Tables S1–S7
III. Domestic cat DNA variants	Tables S8, S9; Figures S2, S3
IV. Repeats content	Tables S10–S16; Figures S4–S13
V. Nuclear mitochondrial (Numt) pseudo gene fragments	Figure S14
VI. Evolutionary constrained elements (ECE)	Tables S17, S18
VII. Feline endogenous retrovirus-like elements	Table S19; Figure S18
VIII. Methylation sites	Table S20
IX. MicroRNA	Table S21
X. Variants between <i>F. silvestris</i> and <i>F. catus</i> .	

features. In Table 1 we list the features of GARfield annotated in the cat genome assembly which are described and illustrated in the Additional file 1 of this Data Note. The genome features detected in Fca-6.2 include a merged list of 21,865 genes derived from a comparative gene identification strategy using BLAST alignments between gene exons of reference genome from eight reference mammalian gene maps (human, chimpanzee, macaque, dog, cow, horse, rat, and mouse) obtained from the Ensembl Gene 75 database [6]. In addition, the whole genome methylation sites and a methylome bisulfite sequence pattern of cat whole blood cells is presented, previewing epigenetic profiling in important complex disease associations, including diseases with viral and neoplastic etiology.

Approximately 55.7% of the cat genome is composed of repetitive elements of familiar classes (LINEs, SINEs, satellite DNA, LTRs and others). We report more than 25 novel families of complex tandem repeat elements in the cat genome uncovered by multiple repeat detection algorithms. We searched for STR-microsatellite loci useful in population and forensic applications. Putative PCR primers for 53,710 STR loci are annotated. We also mapped known feline endogenous retroviral loci (full length RD114, FeLV, FERV) and detected 125 kb of partial retroviral genome sequences dispersed across the cat genome. Nuclear mitochondrial (Numt) DNA pseudogenes derived from ancient transposition from cytoplasmic mitochondrial chromosomes to nuclear chromosomal positions comprise 176 kb in addition to the Lopez-Numt, a 7.8 kb element tandem-repeated 38–76 times on Chromosome D2 previously described in the 1.8× analysis of Cinnamon's genome [7].

The earlier 3,078,438 feline single nucleotide variants (SNVs) [7,8] from largely non-repetitive regions of the cat genome are supplemented with a new group of 99,494 newly annotated SNPs plus 8,355 detected indels. In addition, we performed an assisted assembly with a 40× Illumina SOLID DNA sequence coverage of Sylvester, a European wildcat, *F. silvestris silvestris*, a wild representative of the species from which cats were domesticated approximately 10,000 years ago [9]. Genome variations (SNVs and indels) between *F. silvestris* and *F. catus* SNPs are reported here and both species' genomes and their associated data have been uploaded to the GARfield genome browser (see Availability of supporting data section).

Our annotation resolved cat homologues of 743,362 evolutionarily constrained elements (ECEs) recently identified in the human genome by alignment to 29 different mammalian genomes [10] and these were compared to the conserved sequence blocks obtained by the reciprocal best match (RBM) screen for cat genes with seven mammalian genomes (human, chimp, macaque, dog, cow, rat and mouse). A conservative alignment approach implicated 54% of the human ECE sequence comprising  $\approx 3\%$  of the cat genome. A total of 3,182 feline microRNA (miRNA) homologues were detected and mapped based upon homology to miRNA sequences from 36 species with miRNA sequence described in the miRBase database [11]. Finally we screened the genome sequence for copy number variation and segmental duplications. All annotated features listed in Table 1 are described in detail in Additional file 1 and tracked in the GARfield genome browser.

#### Availability of supporting data

The assembly sequences are available at NCBI RefSeq database (accession numbers #PRJNA175699 and #PRJNA253950). The annotated features are available in the Genome Association Resource Fields (GARfield) genome browser <http://garfield.dobzhanskycenter.org> and the UCSC Genome Browser (<http://genome.ucsc.edu>), which links to a Dobzhansky Center Hub (<http://public.dobzhanskycenter.ru/Hub/hub.txt>) (See Section 2 of Additional file 1 for instructions). Supplementary tables and figures that refer to GARfield features are given in Additional file 1 and listed in Table 1.

Sequence and variation data is available in NCBI (SAMN02795853 for Boris the cat and SAMN02898152 for wildcat) and supporting data is also available in the GigaDB repository [12].

#### Additional file

Additional file 1: Supplementary materials.

### Abbreviations

ECE: Evolutionary constrained element; SNP: Single nucleotide polymorphism; SNV: Single nucleotide variant; STR: Short tandem repeat.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

SJO developed the overall project design. CAD, KB, CB and DG provided the sequence and analyses of *F. silvestris*. MMR and VAD provided the tissue and cat selection tissue for transcriptome and genome variation discovery. KN provided the tissues and sample for Cinnamon, the individual cat whose genome was assembled. SJO, AS and GT wrote the paper. GT performed gene annotation. SS designed the GARfield genome browser. GT prepared tracks of the annotated genome features. PD, AA and GT described DNA variants (SNVs and indels). AL, AK and GT annotated repetitive elements. JP designed STR primers. AM and GT detected evolutionary constrained elements. AS, AS, KPK and GT annotated endogenous retrovirus-like elements. VB, GT and NC analyzed genome methylation. PD and GT detected microRNA homologues. PD and AA searched for Numts. TMB, CA, BLG, and JQ provided the sequence and analyses of copy number variation for *F. catus*. PD assembled *F. silvestris* genome, derived SNVs from it and compared them to *F. catus* SNVs. All authors read and approved the final manuscript.

### Acknowledgments

The authors are grateful to Elena Savelyeva (Clinical Biochemistry Laboratory of St. Petersburg Academy of Veterinary Medicine) for preparing samples of Boris the cat. This work was supported, in part, by Russian Ministry of Science Mega-grant no.11.G34.31.0068; Stephen J. O'Brien, Principal Investigator and ERC Starting Grant (260372) and MICINN (Spain) BFU2011-28549 grants to Tomas Marques-Bonet.

### Author details

<sup>1</sup>Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, 199004, St. Petersburg, Russia. <sup>2</sup>Laboratory of Neurogenetics, NIAAA, 5625 Fishers Lane, 20852 Rockville, MD, USA. <sup>3</sup>CIIMAR — Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Rua dos Bragas, n. 289, 4050–123 Porto, Portugal. <sup>4</sup>Department of Animal and Food Science, Veterinary Molecular Genetics Service, Universitat Autònoma de Barcelona, 08003 Barcelona, Catalonia, Spain. <sup>5</sup>IBE, Institute of Evolutionary Biology, Universitat Pompeu Fabra-CSIC, PRBB (The Barcelona Biomedical Research Park), 08003, Barcelona, Catalonia, Spain. <sup>6</sup>Department of Computer Engineering, Bilkent University, 06800 Ankara, Turkey. <sup>7</sup>Laboratory of Genomic Diversity, Frederick National Laboratory for Cancer Research, 21702 Frederick, MD, USA. <sup>8</sup>Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, University of Missouri, 08028 Columbia, MO, USA. <sup>9</sup>Oceanographic Center, Nova Southeastern University, 33004 Ft Lauderdale, FL, USA.

Received: 1 December 2013 Accepted: 23 July 2014

Published: 5 August 2014

### References

1. Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A, Johnson J, Li K, Mobarry C, Sutton G: **Aggressive assembly of pyrosequencing reads with mates.** *Bioinformatics* 2008, **24**(24):2818–2824.
2. Hillier L, Warren W, O'Brien S, Wilson R, International Cat Genome Sequencing Consortium: **NCBI.** <http://www.ncbi.nlm.nih.gov/nuccore/AANG00000000>.
3. Davis BW, Raudsepp T, Pearks Wilkerson AJ, Agarwala R, Schäffer AA, Houck M, Chowdhary BP, Murphy WJ: **A high-resolution cat radiation hybrid and integrated FISH mapping resource for phylogenomic studies across Felidae.** *Genomics* 2009, **93**(4):299–304.
4. Menotti-Raymond M, David VA, Schäffer AA, Tomlin JF, Eizirik E, Phillip C, Wells D, Pontius JU, Hannah SS, O'Brien SJ: **An autosomal genetic linkage map of the domestic cat, *Felis silvestris catus*.** *Genomics* 2009, **93**(4):305–313.
5. Theodosius Dobzhansky Center for Genome Bioinformatics: **GARfield genome browser.** <http://garfield.dobzhanskycenter.org>.
6. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminecki

- L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, et al.: **The Ensembl genome database project.** *Nucleic Acids Res* 2002, **30**:38–41.
7. Pontius JU, Mullikin JC, Smith DR, Agencourt Sequencing Team, Lindblad-Toh K, Gnerre S, Clamp M, Chang J, Stephens R, Neelam B, Volfovsky N, Schäffer AA, Agarwala R, Narfström K, Murphy WJ, Giger U, Roca AL, Antunes A, Menotti-Raymond M, Yuhki N, Pecon-Slatery J, Johnson WE, Bourque G, Tesler G, NISC Comparative Sequencing Program, O'Brien SJ: **Initial sequence and comparative analysis of the cat genome.** *Genome Res* 2007, **17**(11):1675–1689.
8. Mullikin J, Hansen N, Shen L, Ebling H, Donahue W, Tao W, Saranga D, Brand A, Rubenfield M, Young A, Cruz P, Program NCS, Driscoll C, David V, Al-Murrani S, Locniskar M, Abrahamson M, O'Brien S, Smith D, Brockman J: **Light whole genome sequence for SNP discovery across domestic cat breeds.** *BMC Genomics* 2010, **11**:406.
9. Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, Harley EH, Delibes M, Pontier D, Kitchener AC, Yamaguchi N, O'Brien SJ, Macdonald DW: **The near eastern origin of cat domestication.** *Science* 2007, **317**(5837):519–523.
10. Lindblad-Toh K, Garber M, Zuk O, Lin M, Parker B, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, Ward L, Lowe C, Holloway A, Clamp M, Gnerre S, Alföldi J, Beal K, Chang J, Clawson H, Cuff J, Di Palma F, Fitzgerald S, Flicek P, Guttman M, Hubisz M, Jaffe D, Jungreis I, Kent W, Kostka D, Lara M, et al.: **A high-resolution map of human evolutionary constraint using 29 mammals.** *Nature* 2011, **478**(7370):476–482.
11. Griffiths-Jones S, Grocock RJ, Van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res* 2006, **34**(suppl 1):D140–D144.
12. Tamazian G, Simonov S, Dobrynin P, Makunin A, Logachev A, Komissarov A, Shevchenko A, Brukhin V, Cherkasov N, Svitin A, Koepfli K, Pontius J, Driscoll CA, Blackstone K, Barr C, Goldman D, Antines A, Quilez J, Lorente-Galdos B, Alkan C, Marques-Bonet T, Menotti-Raymond M, David V, Narfström K, O'Brien SJ: **Genomic data of the domestic cat (*Felis catus*).** *GigaSci Database* 2014. <http://dx.doi.org/10.5524/100098>.

doi:10.1186/2047-217X-3-13

Cite this article as: Tamazian *et al.*: Annotated features of domestic cat – *Felis catus* genome. *GigaScience* 2014 **3**:13.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

