A Chromosome-Level Genome of the Agile Gracile Mouse Opossum (*Gracilinanus agilis*)

Ran Tian[1,2,†], Kai Han[3,†], Yuepan Geng[1], Chen Yang[1], Han Guo[1], Chengcheng Shi[3], Shixia Xu[2], Guang Yang[2], Xuming Zhou[4], Vadim N. Gladyshev[5], Xin Liu[3], Lisa K. Chopin[6,7], Diana O. Fisher[8], Andrew M. Baker[9,10], Natália O. Leiner[11], Guangyi Fan[3,12,13,*], and Inge Seim [1,6,7,9,*]

[1]Integrative Biology Laboratory, College of Life Sciences, Nanjing Normal University, Jiangsu, China

[2]Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Jiangsu, China

[3]BGI-Qingdao, BGI-Shenzhen, Qingdao, Shandong, China

[4]Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

[5]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

[6]Ghrelin Research Group, Translational Research Institute, School of Biomedical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia

[7]Australian Prostate Cancer Research Centre-Queensland, Translational Research Institute, Queensland University of Technology, Brisbane, Queensland, Australia

[8]School of Biological Sciences, University of Queensland, Brisbane, Queensland, Australia

[9]School of Biology and Environmental Science, Queensland University of Technology, Brisbane, Queensland, Australia

[10]Natural Environments Program, Queensland Museum, South Brisbane, Queensland, Australia

[11]Laboratório de Ecologia de Mamíferos, Instituto de Biologia, Universidade Federal de Uberlândia, MG, Brazil

[12]State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, China

[13]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen, Guangdong, China

*Corresponding authors: E-mails: fanguangyi@genomics.cn; inge@seimlab.org.

Accepted: 5 July 2021

[†]These authors contributed equally to this work.

Abstract

There are more than 100 species of American didelphid marsupials (opossums and mouse opossums). Limited genomic resources for didelphids exists, with only two publicly available genome assemblies compared with dozens in the case of their Australasian counterparts. This discrepancy impedes evolutionary and ecological research. To address this gap, we assembled a high-quality chromosome-level genome of the agile gracile mouse opossum (*Gracilinanus agilis*) using a combination of stLFR sequencing, polishing with mate-pair data, and anchoring onto pseudochromosomes using Hi-C. This species employs a rare life-history strategy, semelparity, and all *G. agilis* males and most females die at the end of their first breeding season after succumbing to stress and exhaustion. The 3.7-Gb chromosome-level assembly, with 92.6% anchored onto pseudochromosomes, has a scaffold N50 of 683.5 Mb and a contig N50 of 56.9 kb. The genome assembly shows high completeness, with a mammalian BUSCO score of 88.1%. Around 49.7% of the genome contains repetitive elements. Gene annotation yielded 24,425 genes, of which 83.9% were functionally annotated. The *G. agilis* genome is an important resource for future studies of marsupial biology, evolution, and conservation.

Key words: genome, chromosome-level, mouse opossum, South America, *Gracilinanus*.

Significance

There is currently a distinct lack of genome assemblies of more than 100 species of American marsupials, with only two assemblies available, compared with dozens for Australasian marsupials. Here, we present a chromosome-level assembly of the agile gracile mouse opossum (*Gracilinanus agilis*). This species exhibits semelparity—a rare life-history strategy, where all *G. agilis* males and most females die at the end of their first breeding season after succumbing to stress and exhaustion. This genome will contribute to research on marsupial biology, evolution, and conservation.

Introduction

Living marsupials fall within seven orders spread unevenly across the Americas and Australasia (Nilsson et al. 2010; Kumar et al. 2017). Dozens of Australasian marsupials have been genetically sequenced or are forthcoming (reviewed in Deakin and O'Neill [2020]), but few American marsupials have been sequenced. Indeed, of the ~100 species of didelphids (opossums and mouse opossums) (Astúa 2015; Faurby et al. 2018) (supplementary fig. 1, Supplementary Material online), only two genomes are publicly available: the gray short-tailed opossum (*Monodelphis domestica*) (Mikkelsen et al. 2007) and the Virginia opossum (*Didelphis virginiana*) (Dudchenko et al. 2018).

At least three Australian marsupial genera are characterized by a semelparous reproductive strategy where all males of certain species or populations die at the end of their first breeding season after succumbing to stress and exhaustion (Baker and Dickman 2018; Collett et al. 2018; Mutton et al. 2019). In a small number of didelphids, both sexes are reported to be semelparous in four species of two tribes (three in Thylamyini and one in Marmosini) in the subfamily Didelphinae (Leiner et al. 2008; de Andreazzi et al. 2011; Baladrón et al. 2012; Lopes and Leiner 2015; Puida and Paglia 2015; Hernandez et al. 2018; Zangrandi 2018; Albanese et al. 2021).

The genetic basis of marsupial semelparity remains largely unknown, and to understand it we must catalogue various genome sequences across the life-history continuum. Here, we contribute to this effort by presenting a chromosome-level genome assembly of the agile gracile mouse opossum (*Gracilinanus agilis*)—a small, nocturnal, insectivore–omnivore species inhabiting the tropical savannas of central South America (Gardner 2008).

Results and Discussion

The *G. agilis* genome was obtained by stLFR sequencing, polishing with mate-pair data, and anchoring onto seven pseudochromosomes ($2n = 14$) using Hi-C (fig. 1*a* and supplementary table 1 and fig. 2, Supplementary Material online). The final assembly size (3.7 Gb; including unanchored scaffolds) and GC content (37.87%) (table 1 and supplementary fig. 3, Supplementary Material online) are similar to the two other sequenced South American marsupial species—3.61 Gb and 37.82% for *M. domestica* and 3.42 Gb and 37.36% for the *D. virginiana*, respectively. The *G. agilis* assembly has a contig N50 of 56.9 kb, a scaffold N50 of 683.5 Mb (table 1), and the assigned chromosomes are highly homologous to *M. domestica* assembly MonDom5 (contig N50 108.0 kb; scaffold N50 Mb 528.0 Mb) (fig. 1*b*). The *G. agilis* genome is composed of 49.7% repeat elements, including 42.3% LINEs, 12.1% LTRs, and 12.0% SINEs (table 1 and supplementary tables 2 and 3, Supplementary Material online). Out of 9,226 mammalian BUSCO genes, we recovered 7,889 (88.10%) (table 1). We also obtained a complete 16,336 bp mitochondrial genome from the mate-pair data (supplementary fig. 4, Supplementary Material online).

Conclusions

In this work, we report the genome assembly of *G. agilis*—the third from the more than 100 species of South American marsupials. We hope that our current efforts, which employed the relatively inexpensive (~USD 1,000 per sample) stLFR sequencing technology (Stiller and Zhang 2019), will provide an impetus for a wave of genomics research in South America. Indeed, a consortium to facilitate such work will be delineated in an upcoming manuscript (Fisher *et al.*, in preparation). *Gracilinanus agilis* is also one of a handful of South American marsupials that exhibits semelparity and, together with recent genomes of semelparous Australian relatives of genus *Antechinus* (Brandies et al. 2020; Tian et al. 2021), provides the first of many required to unravel a complex life-history strategy. Taken together, the chromosome-level *G. agilis* genome presented in this report should provide a valuable resource for a wide range of research on marsupials.

Materials and Methods

DNA Sequencing

An adult male agile gracile mouse opossum (*Gracilinanus agilis*; LMUSP501) was sampled in Estação Ecológica do Panga, Uberlândia, MG, Brazil (19° 9′ S, 48° 23′ W) in May 2019. Kidney and liver tissues were sequenced by single-tube Long Fragment Read (stLFR) (Fan et al. 2019; Wang et al. 2019) and short-insert library whole-genome sequencing on the BGISEQ-500 platform (2 × 100 bp reads), respectively. A

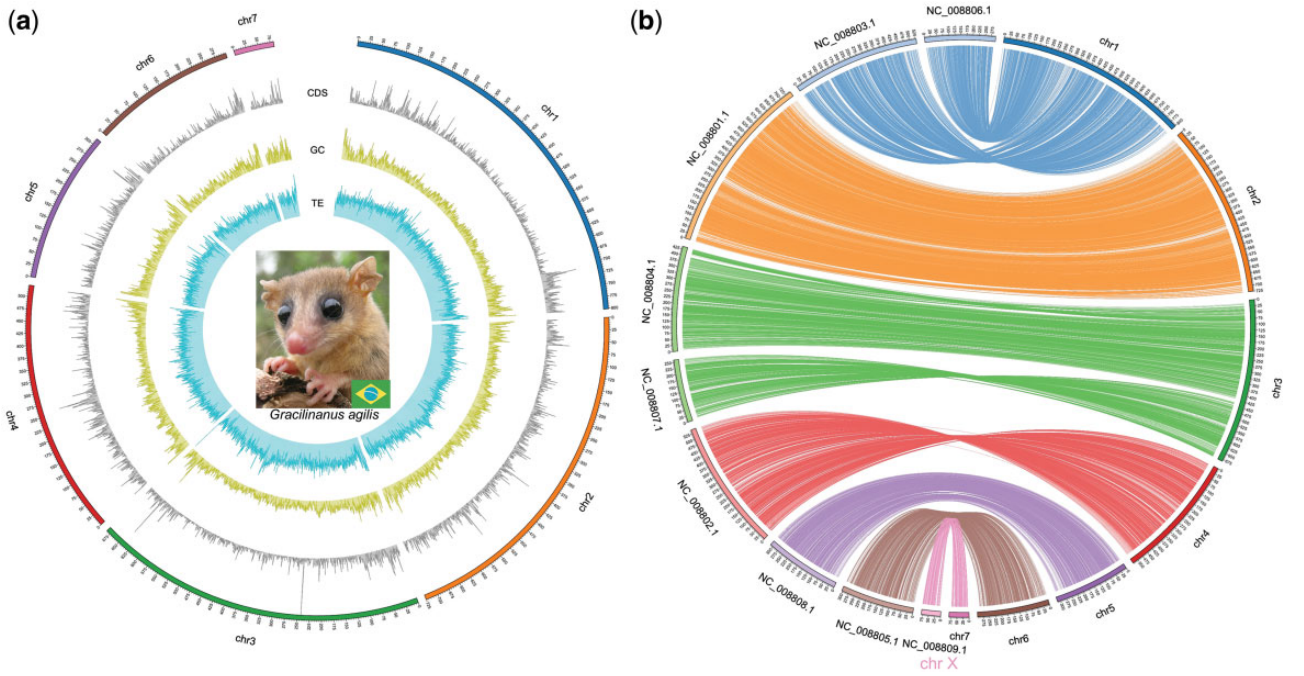

Fig. 1.—Overview of the *Gracilinanus agilis* genome assembly. (*a*) Assembly circos plot. The outermost segment represents chromosome sequences, with the numbers on the external surface indicating genome size (Mb). Line plots, from outside to inside, respectively, represent the distribution of CDS density (from 0 to 0.15), GC content (from 0.30 to 0.65) and TE ratio (from 0.2 to 1.0). Frequencies were calculated in 500 kb sliding windows. Photography courtesy of Noé U. de la Sancha (Chicago State University and Field Museum of Natural History, Chicago). (*b*) Circos plot showing shared synteny of *G. agilis* (chr1–chr7) and the gray short-tailed opossum (*Monodelphis domestica*) (NC_008801.1-NC_008809.1). Aligned using LASTZ. The synteny blocks are linked using lines colored in accordance with the *G. agilis* chromosomes. Aligned blocks with length shorter than 10 kb are not shown. Chr7 in *G. agilis* corresponds to the X chromosome of *M. domestica*.

total of ~358 Gb (~100×) stLFR reads were generated. SOAPnuke v1.5 (Chen et al. 2018) was used to filter out low-quality reads, PCR duplicates, and adaptors. Next, ~264 Gb filtered (clean) data were assembled, using Supernova v2.1.1 (Weisenfeld et al. 2017) and the SOAPdenovo2 module Gapcloser v1.10 (Luo et al. 2012), and short-insert library WGS data (~50×) were used to close gaps. Genome size was estimated by *k*-mer analysis of 100 bp paired-end WGS reads by GCE (Genomic Charactor Estimator) v1.0.0 (Marcais and Kingsford 2011) (supplementary fig. 5, Supplementary Material online). Liver Hi-C libraries were sequenced on the BGISEQ-500 platform and quality controlled using HiC-Pro v2.8.0_devel (Servant et al. 2015), resulting in ~29 Gb uniquely aligned read pairs. Reads validated by HiC-Pro were used to scaffold contigs into seven chromosome clusters using the 3D-DNA v1.12 (Dudchenko et al. 2017). The assembly was further improved by interactive correction using Juicebox v1.11.08 (Durand et al. 2016; Dudchenko et al. 2018). Assembly quality was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs) v5.0.0_cv1 (Seppey et al. 2019) (mammalia_odb10 gene set).

We also generated the complete mitochondrial genome of *G. agilis* from 100 bp WGS reads (see supplementary methods, Supplementary Material online).

Genome Annotation

We identified repetitive elements by integrating homology and de novo prediction data. Protein-coding genes were annotated using homology-based prediction, de novo prediction, and RNA-seq-assisted (generated from kidney, skeletal muscle, and liver from two male individuals) prediction methods. For details, see supplementary methods, Supplementary Material online.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) (Grant Nos. 31950410545 [to I.S.] and 31900310 [to R.T.]), the Jiangsu Science and Technology Agency (to I.S.), and a Brazilian Council of Research and Scientific Development (CNPQ/PELD) (Grant No. 441225/2016-0) (to N.O.L.). Sample collection was approved by the Brazilian Government's Chico Mendes Institute for Biodiversity and Conservation (22629-1) and the Federal

Table 1
Summary of *Gracilinanus agilis* Genome Assembly and Annotation

Genome assembly	Estimated genome size	3.40 Gb
	Assembly size (scaffold)	3.70 Gb
	Assembly size (contig)	3.40 Gb
	Hi-C anchored rate	92.57%
	Contig number	146,614
	Contig N50	56.91 kb
	Longest contig	649.78 kb
	Scaffold number	61,400
	Scaffold N50	683.52 Mb
	Longest scaffold	801.37 Mb
	GC content	37.87%
	Gaps (N)	8.25%
Transposable elements	Annotation	Percent
	DNA	2.18
	LINE	42.30
	SINE	11.98
	LTR	12.05
	Other	0.000094
	Unknown	1.64
	Total	49.71
Protein-coding genes	Predicted genes	24,425
	Average transcript length	64,360 bp
	Average coding sequence length	1,510 bp
	Average exon length	179 bp
	Average intron length	8,448 bp
	Functionally annotated genes	20,492
BUSCO	Complete BUSCOs (C)	8,128 (88.10%)
	Complete and single-copy BUSCOs (S)	7,889
	Complete and duplicated BUSCOs (D)	239
	Fragmented BUSCOs (F)	290
	Missing BUSCOs (M)	808

Note.—Hi-C anchored rate refers the proportion of scaffolded bases assembled onto seven pseudochromosomes. Assembly quality was assessed using BUSCO 5.0.0_cv1 with the 9,226-gene mammalian odb10 data set.

University of Uberlândia Ethics Committee on Animal Use (152/13).

Author Contributions

I.S. and G.F. initiated and coordinated the project. N.O.L. collected tissue samples. C.S. led the sequencing and assembly efforts with G.F. R.T., K.H., and C.S. contributed to genome assembly and annotation. I.S. wrote the paper with input from all authors.

Data Availability

StLFR reads are available at China National GenBank (CNGB) Project ID CNP0001147. Short-insert library whole-genome sequencing and RNA sequencing data (BGISEQ-500) are available at NCBI BioProject PRJNA565840. The nuclear genome assembly is available at NCBI Genomes (JADWME000000000) and the mitochondrial assembly at GenBank (MT219489).

Literature Cited

Albanese MS, Ojeda RA, Astié AA. 2021. Delayed mortality of males in *Thylamys bruchi*, a semelparous marsupial from the Monte Desert, Argentina. J Mamm. 102(1):258–269.

Astúa D. 2015. Family Didelphidae (Opossums). Handb Mamm World. 5:70–186.

Baker A, Dickman C. 2018. Secret lives of carnivorous marsupials. CSIRO Publishing, Melbourne, Australia.

Baladrón AV, Malizia AI, Bó MS, Liébana MS, Bechard MJ. 2012. Population dynamics of the southern short-tailed opossum (*Monodelphis dimidiata*) in the Pampas of Argentina. Aust J Zool. 60(4):238–245.

Brandies PA, Tang S, Johnson RSP, Hogg CJ, Belov K. 2020. The first *Antechinus* reference genome provides a resource for investigating the genetic basis of semelparity and age-related neuropathologies. Gigabyte 2020:1–22.

Chen Y, et al. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience 7(1):1–6.

Collett RA, Baker AM, Fisher DO. 2018. Prey productivity and predictability drive different axes of life-history variation in carnivorous marsupials. Proc Biol Sci. 285:20181291.

Deakin JE, O'Neill RJ. 2020. Evolution of marsupial genomes. Annu Rev Anim Biosci. 8:25–45.

de Andreazzi A, et al. 2011. Population ecology of small rodents and marsupials in a semi-deciduous tropical forest of the southeast Pantanal, Brazil. Zoologia (Curitiba) 28:762–770.

Dudchenko O, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356(6333):92–95.

Dudchenko O, et al. 2018. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. Preprint at http://biorxiv.org/content/10.1101/254797v1.

Durand NC, et al. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3(1):99–101.

Fan H, et al. 2019. Chromosome-level genome assembly for giant panda provides novel insights into Carnivora chromosome evolution. Genome Biol. 20(1):267.

Faurby S, et al. 2018. PHYLACINE 1.2: the phylogenetic atlas of mammal macroecology. Ecology 99(11):2626.

Gardner AL. 2008. Mammals of South America, volume 1: marsupials, xenarthrans, shrews, and bats. University of Chicago Press, Chicago. USA.

Hernandez S, Strona A, Leiner N, Suzán G, Romano M. 2018. Seasonal changes of faecal cortisol metabolite levels in *Gracilinanus agilis* (Didelphimorphia: Didelphidae) and its association to life histories variables and parasite loads. Conserv Physiol. 6(1):coy021.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 34(7):1812–1819.

Leiner NO, Setz EZ, Silva WR. 2008. Semelparity and factors affecting the reproductive activity of the Brazilian slender opossum (*Marmosops paulensis*) in southeastern Brazil. J Mammal. 89(1):153–158.

Lopes GP, Leiner NO. 2015. Semelparity in a population of *Gracilinanus agilis* (Didelphimorphia: Didelphidae) inhabiting the Brazilian cerrado. Mamm Biol. 80(1):1–6.

Luo R, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1(1):18.

Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics 27(6):764–770.

Mikkelsen TS, et al. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. Nature 447(7141):167–177.

Mutton TY, Phillips MJ, Fuller SJ, Bryant LM, Baker AM. 2019. Systematics, biogeography and ancestral state of the Australian marsupial genus *Antechinus* (Dasyuromorphia: Dasyuridae). Zool J Linn Soc. 186(2):553–568.

Nilsson MA, et al. 2010. Tracking marsupial evolution using archaic genomic retroposon insertions. PLoS Biol. 8(7):e1000436.

Puida DBC, Paglia AP. 2015. Primary productivity and the demography of Gracilinanus agilis, a small semelparous marsupial. J Mamm. 96(1):221–229.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol. 1962:227–245.

Servant N, et al. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16:259.

Stiller J, Zhang G. 2019. Comparative phylogenomics, a stepping stone for bird biodiversity studies. Diversity 11(7):115.

Tian R, et al. 2021. A chromosome-level genome of *Antechinus flavipes* provides a reference for an Australasian marsupial genus with suicidal reproduction. Preprint at https://authorea.com/doi/full/10.22541/au.162324962.27977404. doi: 10.22541/au.162324962.27977404/v1.

Wang O, et al. 2019. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. Genome Res. 29(5):798–808.

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. 2017. Direct determination of diploid genome sequences. Genome Res. 27(5):757–767.

Zangrandi PL. 2018. Semelparity in Gracilinanus agilis (Didelphimorphia, Didelphidae): demographic and physiological variation in remaining patches of Cerrado in the Federal District – Brazil. University of Brasília, Brasília, Brazil.

Associate editor: Bonnie Fraser