

Development and annotation of shotgun sequence libraries from New World monkeys

NATALIE M. JAMESON,* KE XU,† SOOJIN V. YI† and DEREK E. WILDMAN*

*Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, 540 East Canfield, Detroit, MI 48021, USA, †School of Biology, Georgia Institute of Technology, 310 Ferst Drive, Atlanta, GA 30332, USA

Abstract

The draft genome sequences of several primates are available, providing insights into evolutionary and anthropological research. However, genomic resources from New World monkeys are conspicuously lacking. To date, the genomes of only two platyrrhine species, the common marmoset and the Bolivian squirrel monkey, have been fully sequenced. This is especially limiting for comparative genomics research, considering that New World monkeys are the most speciose primate group, and platyrrhine genetic diversity is comparable to that of the catarrhines (i.e. apes and Old World monkeys). Here, we present the generation and annotation of numerous sequence reads from the genomes of Spider monkey (*Ateles belzebuth*), Owl monkey (*Aotus lemurinus*) and Uakari (*Cacajao calvus*), representing the three platyrrhine families, Atelidae, Cebidae and Pitheciidae, respectively. These sequencing reads were developed from gDNA shotgun libraries containing over 3000 individual sequences with an average length of 726 bps. Of these sequences, 1220 contain <20% repeats, and thus are potentially highly useful phylogenetic markers for other platyrrhine species. Among them, a large number of sequencing reads were found to match unique regions within the human (2462 sequences) and the marmoset (2829 sequences) genomes. In particular, the majority of these sequencing reads are from putatively neutrally evolving intergenic regions. Thus, they are likely to be highly informative for inferring neutral evolutionary patterns and genomic evolution for other New World monkeys.

Keywords: *Aotus lemurinus*, *Ateles belzebuth*, *Cacajao calvus*, neotropical primate, platyrrhine, shotgun library

Received 7 February 2012; revision received 26 April 2012; accepted 4 May 2012

Introduction

New World monkeys (Primates: Platyrrhini) are the clade of primates that inhabit South and Central America (Groves 2005). The ancestral platyrrhines dispersed to South America between 40 and 26 million years ago thus splitting crown anthropoid primates into two distinct groups: (i) Platyrrhini and (ii) the Catarrhini, comprised of Old World monkeys, and apes including humans (Ciochon 1980; Fleagle 1999; Houle 1999). Upon dispersal, New World monkeys began a period of rapid diversification and divergence filling new environmental niches. This period of rapid divergence has resulted in the most speciose extant primate group, including three families, 15 genera and 126 species (Groves 2001, 2005; Wildman & Goodman 2004; Goodman *et al.* 2005; Opazo *et al.* 2006). The pronounced variation among species can be seen in the form of both morphological and molecular

differences. For example, there is an approximately 100-fold weight difference between the 110 g pygmy marmoset (*Callithrix pygmaea*) and the 11.4 kg Mexican howler monkey (*Alouatta palliata*). Moreover, platyrrhines include the only nocturnal anthropoid primate genus, *Aotus*; and platyrrhines range in chromosome number from $2n = 16$ in the black titi monkey (*Callicebus lugens*) to $2n = 58$ in Hershkovitz's Owl monkey (*Aotus hershkovitzi*) (Torres *et al.* 1998; Fleagle 1999; Stanyon *et al.* 2003).

The morphology, behaviour and ecology of New World monkeys have been widely studied and is well documented (i.e. Hershkovitz 1977; Ford 1986). This knowledge base has informed studies in the fields of phylogenetics, immunology, anthropology, taxonomy and population dynamics (Holmes *et al.* 1966; Rosenberger 1984; Ford 1986; Kay 1990; Genain & Hauser 2001; Marroig & Cheverud 2001; Opazo *et al.* 2006; Organ *et al.* 2009; Papper *et al.* 2009; Wildman *et al.* 2009; Babb *et al.* 2010; Hartwig *et al.* 2011; Terhune 2011). However, the availability of genetic data for New World monkeys remains limited. Although New World monkeys account for approximately one-third of all primate species, until

Correspondence: Derek E. Wildman, Fax: 313-577-5218;

E-mail: dwildman@wayne.edu

Soojin V. Yi, Fax: 404-894-2295;

E-mail: soojin.yi@biology.gatech.edu

very recently, only one of the 10 publicly available fully sequenced primate genomes was from a platyrrhine. In addition to the annotated genome of the common marmoset (*Callithrix jacchus*), the draft genome assembly of the Bolivian squirrel monkey (*Saimiri boliviensis*) has recently been released. Due to the limited amount of molecular data available, many New World monkey studies have instead relied on molecular markers designed based on human sequence data (Witte & Rogers 1999; Clisson *et al.* 2000). The majority of primers designed using human genome sequences fail when employed on New World monkeys because of the long evolutionary distance between them and humans. Thus, it is imperative to develop molecular markers from New World monkey genomes.

To establish a database of high-quality genetic data, we have developed genomic DNA shotgun libraries from representatives of each of the three extant New World monkey families; Atelidae, Cebidae and Pitheciidae (Fig. 1). The white-fronted spider monkey (*Ateles belzebuth*), a member of the Atelidae, inhabits the north-western Amazon and possesses the unique feature of a prehensile tail, a trait shared only by the Atelids and the capuchin monkey (Fleagle 1999; Boubli *et al.* 2008). The grey-bellied night monkey (*Aotus lemurinus*) inhabits the northern tropical Andes, one of only 11 nocturnal anthropoid primates species, and is a member of the Cebidae (Morales-Jiménez & de la Torre *et al.* 2008). The bald uakari (*Cacajao calvus*) inhabits the western Amazon and represents one of the four extant genera that comprise the Pitheciidae (Veiga *et al.* 2008). The inclusion of *Ateles belzebuth*, *Aotus lemurinus*, and *Cacajao calvus* in our shotgun libraries allowed us to capture a wide breadth of genetic diversity that exists through the New World monkey families.

Here, we present a molecular resource that includes 3154 individual genomic sequencing reads containing over 2.3 million nucleotides and is, to date, the largest single molecular resource for these species. A detailed annotation of all sequencing reads and assembled reads as well as alignments of both individual and assembled reads to the human and marmoset genomes are presented in the supplemental data of this manuscript (Supplemental Tables S2 and S3, and Supplemental Data S1 through four). Individual sequencing reads are all available through NCBI GenBank (JQ930030 – JQ933105 and JQ963345–JQ963421), and assembled sequence contigs are available in the supplemental data of this manuscript (Supplemental Data S5).

Materials and methods

Tissue collection and DNA isolation

Whole blood samples from *Cacajao calvus* and *Ateles belzebuth* were obtained from the Centro de Primatologia de

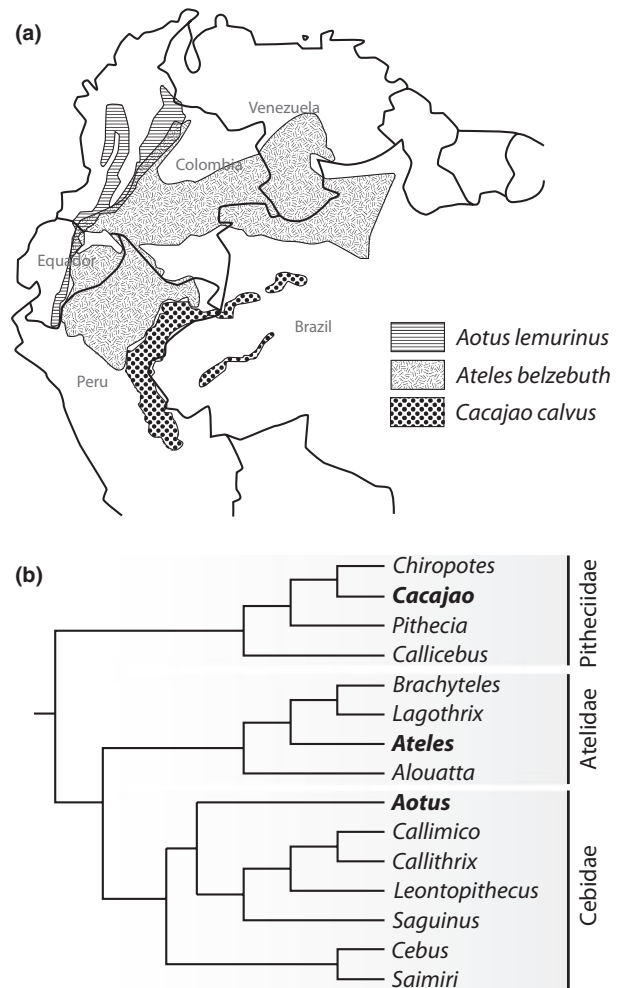


Fig. 1 (a) Partial map of South America showing the distribution of the ranges of the three species of New World monkey sequenced in this study (distribution based on those of the IUCN Red List, Boubli *et al.* 2008; Morales-Jiménez & de la Torre *et al.* 2008; Veiga *et al.* 2008). (b) Phylogenetic tree adapted from Wildman *et al.* 2009 showing the relationships of all New World monkey genera and the separation of the three New World monkey families.

Rio de Janeiro and the Brazil Centro Nacional de Primates, respectively. DNA was isolated from the blood samples using the gDNA blood mini kit from Eppendorf (Hamburg, Germany). DNA was also isolated from an *Aotus lemurinus* liver sample using the Qiagen Inc. (Valencia, CA, USA) DNeasy kit. See Fig. 2 for an overview of the molecular and computational methodology.

Generation of genomic shotgun libraries

DNA was sheared for 90 s at 10 psi using a nebulizer. Randomly cut DNA was first blunt-end repaired and dephosphorylated, then ligated into pCR[®] 4Blunt-TOPO[®] vectors (part of the TOPO[®] Shotgun Subcloning

Kit; Invitrogen, Grand Island, NY, USA). Ligation products were transformed in chemically competent *E. coli* cells. Transformants were placed on selective plates and incubated overnight at 37 °C. Individual colonies were picked and grown overnight in LB media with ampicillin resistance. Vectors with insert were isolated using the Qiaprep Miniprep procedure (Qiagen, Valencia, CA, USA). These inserts were then sequenced in both directions with Big Dye v.3.1 (Applied Biosystems, Foster City, CA, USA), using T3 and T7 universal primers on the ABI 3730xl genetic analyzer. This resulted in a total of 3286 individual sequences.

Sequence read analysis

Vector and primer sequences were trimmed from the resulting sequencing reads, and failed cycle sequencing reactions (sequences containing >300 N's) were discarded from further analysis. This resulted in 3154 sequences remaining for analysis. These sequences have been deposited into GenBank (Accession IDs: JQ930030 - JQ933105 and JQ963345-JQ963421). Repetitive elements within sequences were detected, annotated and masked using RepeatMasker open-3.0 (<http://www.repeatmasker.org/faq.html#faq3>; Seattle, WA, USA) (Smit *et al.* 1996–2004). The shotgun sequences were separately queried in BLAST (Altschul *et al.* 1990) using the BLASTN algorithm against the human (GRCh37/hg19) and marmoset (WUGSC3.2/calJac3) genomes. For both searches, the criteria for selecting hits were a) E-value <1.0E⁻⁵⁰ and b) HSPs must match more than 50% of the query sequence. The human genome annotation was used to determine the location of the BLAST hits. To determine orthology between the human and marmoset genomes, we used the reciprocal BLAST approach. Pairwise alignments against both human and marmoset were generated using T-coffee (Version_8.93; Notredame *et al.* 2000) for those shotgun library sequences having significant blast hits and containing <30% repeats. The alignments of the sequencing reads to the human and marmoset genomes are available in Supplemental Data S1 (Sequencing_Reads_Human_Alignments.txt) and Supplemental Data S2 (Sequencing_Reads_Marmoset_Alignments.txt). Sequence reads containing <20% repeats were assembled into contigs using Phrap (Green & Ewing 2002) while maintaining the default parameters. These contigs were then queried in BLAST against the human and marmoset genomes as described earlier. The alignments of the contigs to the human and marmoset genomes are available in Supplemental Data S3 (Assembled_Contig_Human_Alignments.txt) and in Supplemental Data S4 (Assembled_Contig_Marmoset_Alignments.txt). The sequences of all assembled contigs are available in Supplemental Data S5 (Assembled_Contig_Sequences.txt).

Results

Sequence composition

A total of 3286 sequencing reads were obtained from the shotgun libraries. After removing low-quality sequences, the data set contained 3154 sequence reads, 1272 from the *Ateles* library, 1045 from the *Cacajao* library and 837 from the *Aotus* library. This data set consists of 2 282 823 bps, the shortest and longest sequences measure 52 and 961 bp, respectively, and the mean length of sequences in the data set is 726 bp (Table 1). The GC content of all sequences from the *Ateles* library was 41.8%, 39.5% from the *Cacajao* library sequences and 40.6% from the *Aotus* library sequences. Of the 3154 sequences, 813 were found to be free of repetitive elements, 1220 sequence reads contained <20% repeats and 42 sequence reads contained 100% repeats. Of the repetitive sequences, 36.12% were LINEs, 33.66% were SINEs and 15.32% were LTR elements (Supplemental Table S1, Repetitive_Elements.pdf) as identified by RepeatMasker open-3.0 (Smit *et al.* 1996–2004). The abundance and composition of different transposable elements are similar to what has been observed in other primate genomes (Lander *et al.* 2001; Consortium TCSaA 2005).

Sequence annotation

BLAST searches identified 2462 sequencing reads with at least one hit in the human genome, 1085 of the sequencing reads contained <20% repeats and 704 were free of repetitive elements (Fig. 2). In addition, 2829 of the sequence reads produced at least one hit against the marmoset genome, 1153 of which contained <20% repeats and 735 were free of repetitive elements. As anticipated, the BLAST query against the marmoset genome identified more regions of sequence identity than the query against the human genome. Sequencing reads that produced hits against the human genome covered all chromosomes. Human chromosome 3 contained the most sequencing reads and chromosome 21 contained the least (293 and 28, respectively); see Supplemental Fig. S1 (Human_Chromosome_Coverage.pdf) for more details. With the exception of a few outliers, the number of sequencing reads per chromosome is correlated with the length of the chromosomes ($R^2 = 0.74$, $P = 0.016$). Using the human genome annotation, 57% of the BLAST hits are in intergenic regions, 38% are in introns, 2.2% are within 2 kb upstream or downstream of transcriptional start sites, 0.2% are inside exons, and the remaining span across two or more of the regions specified earlier. These frequencies are in accord with the estimated portions of different genomic regions in primate genomes (Lander *et al.* 2001; Consortium TCSaA 2005).

Table 1 Shotgun library composition

Shotgun libraries	Sequences	Individual sequence lengths				Basepairs	GC %
		Minimum	Maximum	Mean	Median		
<i>Aotus lemurinus</i>	837	133	957	737	793	616 971	40.6
<i>Ateles belzebuth</i>	1272	52	836	697	717	887 261	41.8
<i>Cacajao calvus</i>	1045	52	961	745	754	778 591	39.5
Combined	3154	52	961	726	754.6	2 282 823	

Thus, the sequencing reads developed in this study appear to be well representative of different genomic regions. Full annotation details for each sequencing read are available in Supplemental Table S2 (Sequencing_Read_Annotation.xls).

Orthologous sequences

Among all sequencing reads, 692 included no repeats and had hits in both the human and the marmoset genomes. We also tested the presence of putative single-copy orthologs between human and marmoset genomes among these sequencing reads. We found that 92.3% of the sequencing reads (639 pairs) were single-copy orthologs between human and marmoset genomes, by reciprocal BLAST hit searches between the human and marmoset genomes. Thus, our method is highly efficient in identifying orthologous sequencing reads. Because genomic coverage is relatively low (i.e. 0.01% of the genome was sequenced in each species), there were no sequencing reads that were orthologous between the three shotgun libraries of *Aotus*, *Cacajao* and *Ateles*.

Sequence assembly

The 1220 sequencing reads containing <20% repeats were assembled into contigs using Phrap (Green & Ewing 2002). The assembly produced a total of 186 contigs that contain 66% of the sequencing reads; this brought the total number of unique genomic regions identified in this study to 658 (158 from *Aotus lemurinus*, 250 from *Ateles belzebuth* and 250 from *Cacajao calvus*). Full annotation details for each assembled contig are shown in Supplemental Table S3 (Assembled_Contig_Annotation.xls).

Discussion

The data set presented here is the largest single molecular resource for these species and should prove to be useful for a variety of downstream applications. None of the genomes included in the shotgun libraries (*Aotus lemurinus*, *Ateles belzebuth* or *Cacajao calvus*) are scheduled for whole genome sequencing in the near future, and currently

New World monkey draft genome data are limited to the common marmoset (*Callithrix jacchus*) and the Bolivian squirrel monkey (*Saimiri boliviensis*). Although some molecular genetic work has been performed on the species included in the present study (i.e. Opazo *et al.* 2006; Wildman *et al.* 2009; Menezes *et al.* 2010; Perelman *et al.* 2011), the resources are still rather limited. For example, there are only 402 Entrez nucleotide entries belonging to these species in NCBI (<http://www.ncbi.nlm.nih.gov>, as of April 2012). The addition of this large-scale molecular resource works to fill a gap in the molecular knowledge base available within New World monkeys.

Our annotations determined these sequences to exist in a variety of genic and intergenic regions. However, given the limited depth of our sequence coverage, we cannot rule out the possibility that those sequences determined to be located in genes are not in pseudogenes or another member of the gene family. Due to the diversity of genomic regions captured by shotgun libraries, they present a good source of variety of molecular markers. As an example, previous and ongoing research by our group has resulted in the development of nongenic markers, which were used for phylogenetic analysis, developed from similar shotgun libraries (Wildman *et al.* 2009). In particular, markers generated from putatively neutral regions were highly useful to resolve phylogenetic relations (Wildman *et al.* 2009) as well as to determine molecular rate variation (Peng *et al.* 2009). In addition to the nongenic sequences, these data provide a good resource for the development of various types of molecular markers that could be used in phylogenetic, population or conservation studies. As the majority of these sequences were also found in either the human or the marmoset genome assemblies, these markers have and will continue to be a great resource for comparative genomic study approaches.

We have made this data set available in anticipation of its use for various downstream applications such as microsatellite marker development and phylogenetic markers (coding, noncoding or nongenic). It provides a valuable contribution to the genetic tools existing for the study of all New World monkey species as well as for other primates.

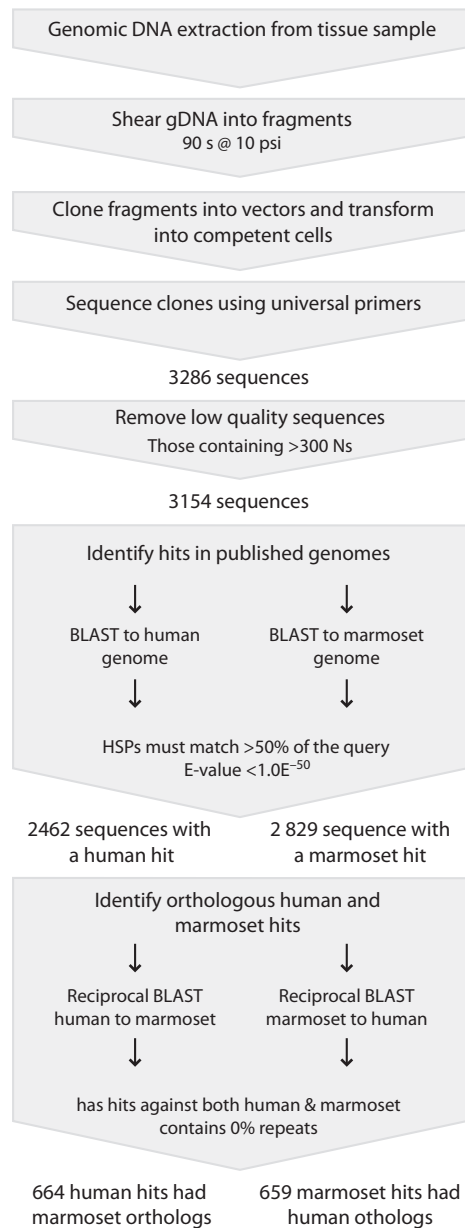


Fig. 2 Workflow showing the steps involved in the shotgun library construction, sequencing and annotation.

Acknowledgements

This research was funded by National Science Foundation Grants (BCS-0751508) to D.W. and (BCS-0751481) to S.Y. Natalie Jameson is funded by NSF Doctoral Dissertation Improvement Grant BCS-1061370. We thank The Centro de Primatologia de Rio de Janeiro, Centro Nacional de Primatas, Belém, Pará, Brazil, and Dr. Caro-Beth Stewart for providing samples.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

- Babb PL, Fernandez-Duque E, Schurr TG (2010) AVPR1A sequence variation in monogamous owl monkeys (*Aotus azarai*) and its implications for the evolution of platyrrhine social behavior. *Journal of Molecular Evolution*, **71**, 279–297.
- Boublil J-P, Di Fiore A, Stevenson P, Link A, Marsh L, Morales AL (2008) *Ateles belzebuth*. In: IUCN 2011. IUCN Red List of Threatened Species. Version 2011.2.
- Ciochon RCA (1980) *Evolutionary Biology of the New World Monkeys and Continental Drift*. Plenum Publishing Corporation, New York NY.
- Clisson I, Lathuilliere M, Crouau-Roy B (2000) Conservation and evolution of microsatellite loci in primate taxa. *American Journal of Primatology*, **50**, 205–214.
- Consortium TCSaA (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, **437**, 69–87.
- Fleagle JG (1999) *Primate Adaptation and Evolution*. Academic Press, San Diego CA.
- Ford SM (1986) Systematics of the New World monkeys. In: *Comparative Primate Biology* (eds Swindler DR & Erwin J), pp. 73–135. Alan Liss, New York.
- Genain CP, Hauser SL (2001) Experimental allergic encephalomyelitis in the New World monkey *Callithrix jacchus*. *Immunological Reviews*, **183**, 159–172.
- Goodman M, Grossman LI, Wildman DE (2005) Moving primate genomics beyond the chimpanzee genome. *Trends in Genetics*, **21**, 511–517.
- Green P, Ewing B (2002). Phred, version 0.020425c. Available from <http://phrap.org>.
- Groves C (2001) *Primate Taxonomy*. Smithsonian Institution Press, Washington DC.
- Groves C (2005) Order Primates. In: *Mammal Species of the World* (eds Wilson DE & Reeder DE), pp. 111–184. John Hopkins University Press, Baltimore MD.
- Hartwig W, Rosenberger AL, Norconk MA, Owl MY (2011) Relative brain size, gut size, and evolution in New World monkeys. *Anat Rec (Hoboken)*, **294**, 2207–2221.
- Hershkovitz P (1977) *Living New World Monkeys (Platyrrhini), Volume 1: with an Introduction to Primates*. Chicago University Press, Chicago.
- Holmes AW, Devine JA, Nowakowski E, Deinhardt F (1966) The epidemiology of a herpes virus infection of New World monkeys. *Journal of Immunology*, **96**, 668–671.
- Houle A (1999) The origin of platyrrhines: an evaluation of the Antarctic scenario and the floating island model. *American Journal of Physical Anthropology*, **109**, 541–559.
- Kay R (1990) The phyletic relationships of extant and fossil Pitheciinae (Platyrrhini, Anthropoidea). *Journal of Human Evolution*, **19**, 175–208.
- Lander ES, Linton LM, Birren B et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
- Marroig G, Cheverud JM (2001) A Comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of New World monkeys. *Evolution*, **55**, 2576–2600.
- Menezes AN, Bonvicino CR, Seuanez HN (2010) Identification, classification and evolution of owl monkeys (*Aotus*, Illiger 1811). *BMC Evolutionary Biology*, **10**, 248.
- Morales-Jiménez AL, de la Torre S (2008) *Aotus lemurinus*. In: IUCN 2011. IUCN Red List of Threatened Species. Version 2011.2. Available from <http://www.iucnredlist.org>.
- Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, **302**, 205–217.
- Opazo JC, Wildman DE, Prychitko T, Johnson RM, Goodman M (2006) Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). *Molecular Phylogenetics and Evolution*, **40**, 274–280.
- Organ JM, Teaford MF, Taylor AB (2009) Functional correlates of fiber architecture of the lateral caudal musculature in prehensile and

- nonprehensile tails of the platyrrhini (primates) and procyonidae (carnivora). *Anat Rec (Hoboken)*, **292**, 827–841.
- Papper Z, Jameson NM, Romero R *et al.* (2009) Ancient origin of placental expression in the growth hormone genes of anthropoid primates. *Proc Natl Acad Sci USA*, **106**, 17083–17088.
- Peng Z, Elango N, Wildman DE, Yi SV (2009) Primate phylogenomics: developing numerous nuclear non-coding, non-repetitive markers for ecological and phylogenetic applications and analysis of evolutionary rate variation. *BMC Genomics*, **10**, 247.
- Perelman P, Johnson WE, Roos C *et al.* (2011) A molecular phylogeny of living primates. *PLoS Genetics*, **7**, e1001342.
- Rosenberger AL (1984) Fossil New World monkeys dispute the molecular clock. *Journal of Human Evolution*, **13**, 737–742.
- Smit A, Hunley R, Green P (1996–2004) Repeatmasker Open-3.0. Available from <http://www.repeatmasker.org>.
- Stanyon R, Bonvicino CR, Svartman M, Seuanez HN (2003) Chromosome painting in *Callicebus lugens*, the species with the lowest diploid number (2n=16) known in primates. *Chromosoma*, **112**, 201–206.
- Terhune CE (2011) Dietary correlates of temporomandibular joint morphology in New World primates. *Journal of Human Evolution*, **61**, 583–596.
- Torres OM, Enciso S, Ruiz F, Silva E, Yunis I (1998) Chromosome diversity of the genus *Aotus* from Colombia. *American Journal of Primatology*, **44**, 255–275.
- Veiga LM, Bowler M, Silva JS Jr, Queiroz HL, Boubli J-P, Rylands AB (2008) *Cacajao calvus*. In: IUCN 2011. *IUCN Red List of Threatened Species*. Version 2011.2.
- Wildman DE, Goodman M (2004) Humankind's place in a phylogenetic classification of living primates. in: *Evolutionary Theory and Processes: modern Horizons, Papers in Honor to Eviatar Nevo* (ed. Wasser S), pp. 293–311. Kluwer Academic Publishers, Netherlands.
- Wildman DE, Jameson NM, Opazo JC, Yi SV (2009) A fully resolved genus level phylogeny of neotropical primates (Platyrrhini). *Molecular Phylogenetics and Evolution*, **53**, 694–702.
- Witte SM, Rogers J (1999) Microsatellite polymorphisms in Bolivian squirrel monkeys (*Saimiri boliviensis*). *American Journal of Primatology*, **47**, 75–84.

D.E.W., and S.V.Y. designed the research, N.M.J., and K.X. performed research and analyzed data. N.M.J., D.E.W., S.V.Y. and K.X. wrote the paper.

Data Accessibility

All data files have been included as Supplemental Data. Annotation and description of each sequencing read is presented in Supplemental Table S2 (Sequencing_Read_Annotation.xls). Annotation and description of each assembled contig is presented in Supplemental Table S3 (Assembled_Contig_Annotation.xls). Sequencing read alignments with the human and marmoset genomes are presented in Supplemental Data files 1 (Sequencing_Reads_Human_Alignments.txt) and 2 (Sequencing_Reads_Marmoset_Alignments.txt), respectively.

Assembled contig alignments with the human and marmoset genomes are presented in Supplemental Data S3 (Assembled_Contig_Human_Alignments.txt) and 4 (Assembled_Contig_Marmoset_Alignments.txt), respectively. The assembled contig sequences in FASTA format are available in Supplementary Data S5 (Assembled_Contig_Sequences.txt). The sequencing read files generated from the shotgun libraries have been deposited into GenBank (Accession IDs: JQ930030 – JQ933105 and JQ963345–JQ963421).

Supporting Information

Additional supporting information may be found in the online version of this article.

Figure S1 Human_Chromosome_Coverage.pdf: Distribution of sequencing reads among human chromosomes based on BLAST results.

Table S1 Repetitive_Elements.pdf: Annotation of repetitive elements present among the shotgun library sequencing reads.

Table S2 Sequencing_Read_Annotation.xls: Detailed annotation of all sequencing reads including a Readme tab with header definitions and explanations.

Table S3 Assembled_Contig_Annotation.xls: Detailed annotation of all assembled contigs including a Readme tab with header definitions and explanations.

Data S1 Sequencing_Reads_Human_Alignments.txt: Sequencing read alignments to the human genome.

Data S2 Sequencing_Reads_Marmoset_Alignments.txt: Sequencing read alignments to the marmoset genome.

Data S3 Assembled_Contig_Human_Alignments.txt: Sequencing read contig alignments to the human genome.

Data S4 Assembled_Contig_Marmoset_Alignments.txt: Sequencing read contig alignments to the marmoset genome.

Data S5 Assembled_Contig_Sequences.txt: FASTA format sequences of all assembled contigs.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.