

Commentary

Open Access

Initiation of the Primate Genome Project

A crucial step for understanding human evolution is to identify the genomic changes that occurred during primate evolution, thus allowing investigators to reconstruct the ancestral states preceding the human condition. In the past several decades, the primate clade has been a research focus in genome sequencing due to its unique phylogenetic position and key importance. Comparative genomic analyses of several primate lineages have radically expanded our knowledge on the tempo and mode of different features in primate genome evolution, revealing many genomic innovations contributing to the development and evolution of human phenotypes. However, with less than 10% of primate species currently sequenced, a considerable gap remains regarding the evolutionary history of every base pair in human and non-human primate (NHP) genomes. To fill this gap, we propose to organize and establish the Primate Genome Project (PGP) to scale up the number of high-quality reference genome assemblies for primate species using cutting-edge sequencing technologies. We outline here the possible paths going forward and some of the major questions to be addressed within this ambitious project. We anticipate that genomic comparisons, including broader taxon sampling of extant primate species, will significantly contribute to our understanding of the evolution of human phenotypes and diseases and the genomic mechanisms of primate speciation and adaptation, which will ultimately assist in primate conservation efforts.

Worldwide, there are currently more than 500 primate species from 80 genera and 16 families, with new primate species still being discovered in recent years (Fan et al., 2017; Nater et al., 2017; Roos et al., 2020). As our closest biological relatives, NHPs hold many clues for understanding the origin and evolution of human complex traits, behaviors, and diseases. NHPs are widely used as biomedical research models for studying the genetic basis of human diseases (e.g., neurodegeneration, diabetes, cancer, and cardiovascular disease), as well as models for reproduction, transplantation, and pharmacology (Rogers & Gibbs, 2014). Furthermore, over

the last two years, NHPs have played critical roles in the study of SAR-CoV-2 infection mechanisms and vaccine development (Chandrashekar et al., 2020; Mercado et al., 2020). Therefore, current primate genome sequencing efforts have prioritized the great ape lineage, i.e., taxa closest to humans, and other biomedically relevant species (Carbone et al., 2014; Gibbs et al., 2007; Gordon et al., 2016; Locke et al., 2011; Scally et al., 2012; The Chimpanzee Sequencing and Analysis Consortium, 2005; The Marmoset Genome Sequencing and Analysis Consortium, 2014; Warren et al., 2020), with 90% of primate species yet to be characterized. Due to anthropogenic interference, climate change, and other factors, some 60% of the world's primate species are threatened with extinction and 75% are experiencing population declines (Estrada et al., 2017), thus highlighting an impending extinction crisis. Therefore, the phylogenomic study of primates is critical for broadening our knowledge on the evolutionary and adaptive history of each species, and for providing relevant information for conservation decisions going forward.

Currently, the genomes of a dozen representative NHP species from 22 genera have been published (Ensembl v103), with 72% of genera not yet sequenced. The PGP aims to close this gap and generate long-read-based high-quality reference genomes for at least one representative species in each primate genus in the next few years. Ultimately, we wish to determine the genome sequences for all primate species. These high-quality genomes will provide rapid information for comparisons in the context of primate phylogeny to reconstruct the primate tree of life and clarify the genomic changes underlying the speciation and adaptation processes of major primate lineages. We anticipate that a detailed evolutionary landscape will be disclosed for all genomic variations across primate lineages from chromosomal rearrangements to single base-pair substitutions. This landscape will inform the evolutionary patterns of structural variations, segmental duplications, protein-coding genes, and non-coding regulatory elements. By integrating the meta-records of trait data, genomic information will enable us to

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2022 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 04 January 2022; Accepted: 06 January 2022; Online: 06 January 2022

Foundation items: This work was supported by the National Natural Science Foundation of China (31822048) and Strategic Priority Research Program of the Chinese Academy of Sciences (XDPB17)

investigate phylogeny, incomplete lineage sorting, and chromosomal evolution, and to map genetic changes associated with human-specific phenotypes (such as brain and cognitive ability), and phenotypic diversity among primate lineages that may contribute to their adaptive evolution.

Due to the difficulties in collecting samples of all primate species, it may not be feasible to sequence all individual primate species within a short time. Nevertheless, depending on sample availability, we will select certain genera as primary sequencing targets and generate reference genomes and population sequences for all recorded species in those genera. This will allow us to investigate the genetic basis of speciation and characterize the history and genomic outcomes of hybridization, as observed in major primate lineages (Ayoola et al., 2021; de Manuel et al., 2016; Svardal et al., 2017). Conservation genomic analyses of population data will be highly informative regarding genomic diversity and recent demographic history, which are crucial parameters for evaluating the conservation status of species (Orkin et al., 2021).

We expect that the PGP will generate a rich resource for primate lineage genomics over the next few years. Given the advances in sequencing technologies, it is important to initiate interdisciplinary cooperation and joint data analysis by integrating expertise in primatology, conservation, biomedicine, genomics, and evolutionary biology. Therefore, we welcome researchers to join the consortium and contribute to this necessary and timely research.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

D.D.W. and G.Z. conceived the review and prepared the draft. All authors contributed to the discussions. All authors read and approved the final version of the manuscript.

Dong-Dong Wu^{1,2,3,4,*}, Xiao-Guang Qi⁵, Li Yu⁶, Ming Li^{4,7}, Zhi-Jin Liu⁷, Anne D. Yoder⁸, Christian Roos⁹, Takashi Hayakawa^{10,11}, Jeffrey Rogers¹², Tomas Marques-Bonet^{13,14,15,16}, Bing Su^{1,4}, Yong-Gang Yao^{2,17,18}, Ya-Ping Zhang^{1,4,18}, Guojie Zhang^{1,4,19,20,*}

¹ State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China

² National Resource Center for Non-Human Primates, Kunming Primate Research Center, and National Research Facility for Phenotypic & Genetic Analysis of Model Animals (Primate Facility), Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650107, China

³ Kunming Natural History Museum of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China

⁴ Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China

⁵ Shaanxi Key Laboratory for Animal Conservation, College of Life Sciences, Northwest University, Xi'an, Shaanxi 710069, China

⁶ State Key Laboratory for Conservation and Utilization of Bio-Resource in Yunnan, Yunnan University, Kunming, Yunnan 650091, China

⁷ CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

⁸ Department of Biology, Duke University, Durham, NC 27708, USA

⁹ Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen 37077, Germany

¹⁰ Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan

¹¹ Japan Monkey Centre, Inuyama, Aichi 484-0081, Japan

¹² Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

¹³ Institute of Evolutionary Biology (UPF-CSIC), PRBB, Dr. Aiguader 88, Barcelona 08003, Spain

¹⁴ Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, Barcelona 08010, Spain

¹⁵ CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, Barcelona 08028, Spain

¹⁶ Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Edifici ICTA-ICP, c/ Columnes s/n, Cerdanyola del Vallès, Barcelona 08193, Spain

¹⁷ Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, and KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650201, China

¹⁸ Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650204, China

¹⁹ Villum Centre for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen DK-2100, Denmark

²⁰ China National GeneBank, BGI-Shenzhen, Shenzhen, Guangdong 518083, China

*Corresponding authors, E-mail: wudongdong@mail.kiz.ac.cn; guojie.zhang@bio.ku.dk

REFERENCES

- Ayoola AO, Zhang BL, Meisel RP, Nneji LM, Shao Y, Morenikeji OB, et al. 2021. Population genomics reveals incipient speciation, introgression, and adaptation in the African Mona Monkey (*Cercopithecus mona*). *Molecular Biology and Evolution*, **38**(3): 876–890.
- Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, et al. 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature*, **513**(7517): 195–201.
- Chandrashekar A, Liu JY, Martinot AJ, McMahan K, Mercado NB, Peter L,

- et al. 2020. SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science*, **369**(6505): 812–817.
- De Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, et al. 2016. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science*, **354**(6311): 477–481.
- Estrada A, Garber PA, Rylands AB, Roos C, Fernandez-Duque E, Di Fiore A, et al. 2017. Impending extinction crisis of the world's primates: why primates matter. *Science Advances*, **3**(1): e1600946.
- Fan PF, He K, Chen X, Ortiz A, Zhang B, Zhao C, et al. 2017. Description of a new species of *Hoolock* gibbon (Primates: Hylobatidae) based on integrative taxonomy. *American Journal of Primatology*, **79**(5): e22631.
- Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, **316**(5822): 222–234.
- Gordon D, Huddleston J, Chaisson MJP, Hill CM, Kronenberg ZN, Munson KM, et al. 2016. Long-read sequence assembly of the gorilla genome. *Science*, **352**(6281): aae0344.
- Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, et al. 2011. Comparative and demographic analysis of orang-utan genomes. *Nature*, **469**(7331): 529–533.
- Mercado NB, Zahn R, Wegmann F, Loos C, Chandrashekar A, Yu JY, et al. 2020. Single-shot Ad26 vaccine protects against SARS-CoV-2 in rhesus macaques. *Nature*, **586**(7830): 583–588.
- Nater A, Mattle-Greminger MP, Nurcahyo A, Nowak MG, De Manuel M, Desai T, et al. 2017. Morphometric, behavioral, and genomic evidence for a new orangutan species. *Current Biology*, **27**(22): 3487–3498.e10.
- Orkin JD, Kuderna LFK, Marques-Bonet T. 2021. The diversity of primates: from biomedicine to conservation genomics. *Annual Review of Animal Biosciences*, **9**: 103–124.
- Rogers J, Gibbs RA. 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nature Reviews Genetics*, **15**(5): 347–359.
- Roos C, Helgen KM, Miguez RP, Thant NML, Lwin N, Lin AK, et al. 2020. Mitogenomic phylogeny of the Asian colobine genus *Trachypithecus* with special focus on *Trachypithecus phayrei* (Blyth, 1847) and description of a new species. *Zoological Research*, **41**(6): 656–669.
- Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature*, **483**(7388): 169–175.
- Svardal H, Jasinska AJ, Apetrei C, Coppola G, Huang Y, Schmitt CA, et al. 2017. Ancient hybridization and strong adaptation to viruses across African vervet monkey populations. *Nature Genetics*, **49**(12): 1705–1713.
- The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, **437**(7055): 69–87.
- The Marmoset Genome Sequencing and Analysis Consortium. 2014. The common marmoset genome provides insight into primate biology and evolution. *Nature Genetics*, **46**(8): 850–857.
- Warren WC, Harris RA, Haukness M, Fiddes IT, Murali SC, Fernandes J, et al. 2020. Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. *Science*, **370**(6523): eabc6617.