



Exploring the primate genome: Unraveling the mysteries of evolution and human disease

Yaoxi He^{1,2,*} and Yafei Mao^{3,4,*}

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

²National 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 650107, China

³Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Jiao Tong University, Shanghai 200230, China

⁴Center for Genomic Research, International Institutes of Medicine, The Fourth Affiliated Hospital, Zhejiang University, Yiwu 322000, China

*Correspondence: heyaoxi@mail.kiz.ac.cn (Y.H.); yafmao@sjtu.edu.cn (Y.M.)

Received: June 4, 2023; Accepted: June 21, 2023; Published Online: June 21, 2023; <https://doi.org/10.1016/j.xinn.2023.100467>

© 2023 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Citation: He Y. and Mao Y. (2023). Exploring the primate genome: Unraveling the mysteries of evolution and human disease. *The Innovation* 4(4), 100467.

Nonhuman primates (NHPs), our closest relatives, represent one of the most successful lineages of adaptive radiation in mammals. The primate order contains several hundred species living in varied ecological niches with phenotypic diversity. However, what we know about NHPs is still limited because of genome diversity.

To address this gap, the Primate Genome Project (PGP) consortium recently made a comprehensive sequencing effort.¹ They sequenced 703 individuals from 233 primate species (16 families, 68 genera) with short-read sequencing technology as well as 27 primate species with long-read sequencing technologies, covering nearly half of all primate species.² These efforts allow us to test long-standing evolutionary hypotheses, explore novel phenotypes, investigate social behavior changes, understand hybrid speciation, and refine variants associated with human disease risk. These efforts have significantly advanced our understanding of divergence, speciation, diversity, adaptation, and human disease during primate evolution, which paves the way for future breakthroughs in conservation strategies and precision medicine (Figure 1).

GENETIC DIVERSITY AND ADAPTATION

Genetic diversity and mutation rates are pivotal factors in understanding the complexities of biodiversity and the emergence of novel phenotypes. This comprehensive dataset challenges the prevailing hypothesis suggesting a positive correlation between genetic diversity and extinction risk. In addition, the data support the drift-barrier hypothesis, which posits that the mutation rate per generation decreases with the effective population size. Another intriguing finding is the widespread occurrence of recurrent single nuclear mutations in primates,

aligning with a recent study on structural variations in great apes. These findings shed light on the genetic forces driving evolutionary changes in primates.

Advancements in more accurate long-read primate genome assemblies enable identification of large-scale genomic rearrangements in chromosome evolution among primates. Additionally, comparison of large-scale orthologous genes/regions has contributed to our understanding of multiple phenotypic innovations involving amino acid replacements driven by natural selection or lineage-specific accelerated sequences in specific lineages. These phenotypic innovations encompass traits related to the brain, skeletal structure, and digestion (e.g., brain evolution and limb development).² These endeavors broaden the horizon, allowing us to explore phenotype novelty across various primate lineages beyond the previously studied great apes.

Furthermore, the authors address a long-standing question regarding the relationship between incomplete lineage sorting (ILS), neutral evolution, and natural selection.³ Gene tree discordance, a prevalent occurrence in primate evolution, has been extensively investigated in great ape lineages. However, a higher number of ILS loci is observed by sequencing additional monkey genomes. This observation supports the hypothesis that ILS loci are not randomly distributed in these additional monkey genomes, and these loci tend to have higher recombination rates because of selection. The genes affected by ILS also contribute to phenotypic novelty in primates.

The profound impact of genetic diversity, adaptation, and ILS deepens our understanding of phenotypic novelty and genetic changes in primate evolution. However, there is still a need to explore unresolved phylogenetic

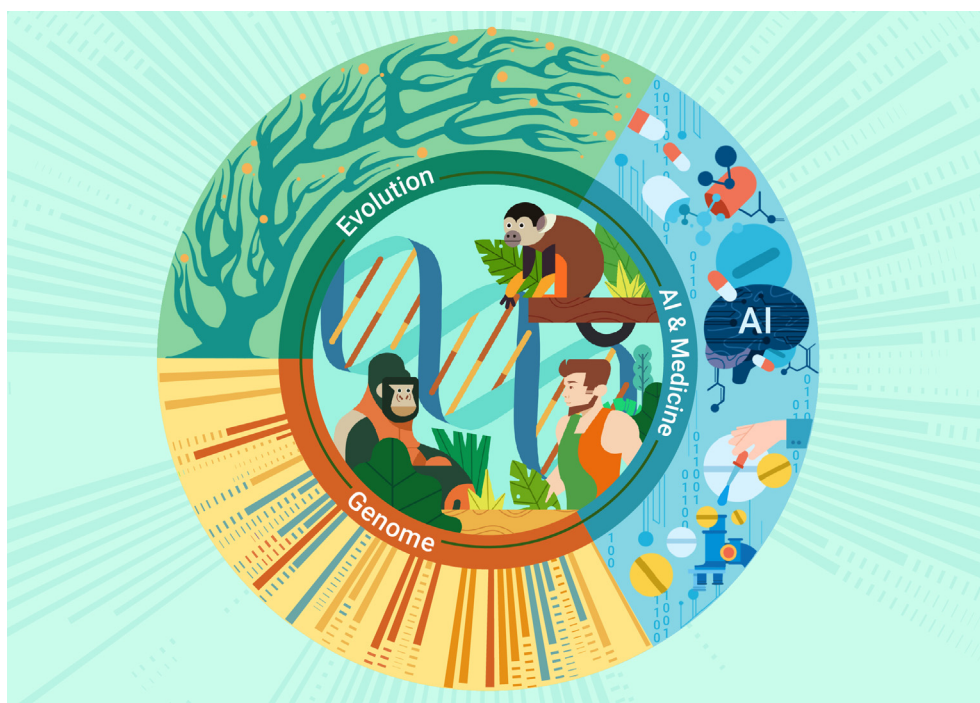


Figure 1. Evolution, genomics, and AI facilitate biodiversity, conservation, and precision medicine studies

relationships in certain clades, such as gibbons. Additionally, a precise understanding of divergence time and speciation time in primates remains challenging because of the high levels of ILS and introgression. While our understanding of orthologous genes and segments has significantly improved with these datasets and analyses, further exploration of divergent, recurrently mutated, or lineage-specific regions will likely provide additional insights into primate evolution.

DIVERGENCE AND SPECIATION

Speciation is a fundamental and crucial evolutionary process that contributes to our understanding of biodiversity and the origin of human beings. A simple speciation model is believed to occur through geographic isolation, leading to genetic and reproductive divergence and eventual formation of distinct species. However, recent studies utilizing long-read and short-read population-scale data from three representative clades challenge the simplicity of allopatric models of speciation. These studies provide a novel and more nuanced understanding of the speciation process in primates. The authors have made significant discoveries regarding the roles of hybridization in speciation and adaptive radiation, challenging the notion that hybrid speciation is limited to plants and a few animal species. The three monkey clades undergo rapid adaptive radiation and are susceptible to natural hybridization because of weak reproductive isolation. The authors also identify shared and species-specific genomic elements through comprehensive comparisons.

The integration of knowledge from macro- and micro-evolutionary scales opens new avenues for studying hybrid speciation. The authors explore the processes associated with morphological and social behavioral differences and innovations that drive speciation.⁴ For example, they have identified genes involved in pigmentogenesis that undergo natural selection, contributing to the distinct mosaic coat coloration observed in the gray snub-nosed monkey. Moreover, the genes associated with reproduction and neurofunction exhibit divergence in the other two monkey clades. These macro-micro adaptations in new species establish reproductive barriers in both parent species, facilitating colonization of new ecological niches or adaptive peaks.

These studies underscore the increasing significance of genomic approaches in speciation research. The rapid pace of ecological data collection and the compilation of primate genomes are poised to address long-standing and emerging questions about the genetics of speciation. However, several aspects remain unclear, such as identification of specific “genomic islands” in these species and the need for further development of theoretical models to understand hybridization.

EVOLUTIONARY MEDICINE AND ARTIFICIAL INTELLIGENCE (AI)

In the field of human medical genomics, numerous common and rare variants are identified through large cohorts, pedigree studies, and case-control studies, linking them to human disease risk. However, assuming that variants in NHPs have similar effects to those in humans, leveraging NHP variants can aid in predicting human disease risk.

In two groundbreaking studies, the authors refined millions of previously unknown risk variants in humans and utilized these data in conjunction with state-of-the-art AI technology to develop PrimateAI-3D, a powerful tool that elucidates the structural changes in proteins caused by variants. By applying PrimateAI-3D and employing polygenic risk score (PRS) models derived from PrimateAI-3D to analyze common and rare variants, they not only achieved improved variant effect prediction but also gained valuable insights into common and rare variants associated with phenotype variation and human disease risk.

This successful integration of evolutionary biology, medical genomics, and AI exemplifies the immense potential of NHP resources in advancing the field of evolutionary medicine. However, future AI models should consider more complex factors, such as non-coding variants, variant dosage effects, compensation effects, and phenotypic differences in NHPs. These refinements will contribute to a more comprehensive understanding of phenotype- and disease-associated variants and enable more accurate predictions in personalized medicine.

WHAT'S NEXT

The significant efforts in studying primate genomes greatly contribute to our understanding of primate evolution and human diseases. However, only a small

proportion (<10%) of primate genomes have been assembled to the chromosome level, hindering our comprehensive understanding of large-scale variants in primate evolution. Eventually, we expect all primate genomes to be sequenced at the telomere-to-telomere (T2T) level with extensive genome annotations.⁵

In addition to the genome sequences, the annotation of regulatory elements is of utmost importance. While the Encyclopedia of DNA Elements (ENCODE) project has made significant progress in annotating functional elements in the human genome, there is currently no such resource available for NHPs. Exploring non-coding variants and functional elements is crucial for a deeper understanding of genome evolution. We believe that the comprehensive annotation of functional elements in NHPs will undoubtedly provide new insights into understanding our primate lineage in the near future. Fortunately, the Primate Research Center of the Kunming Institute of Zoology (Chinese Academy of Sciences, CAS) has initiated the monkey ENCODE project, which aims to generate a comprehensive dataset of genomic annotations covering the entire life cycle of the rhesus macaque, from fertilized eggs to aged monkeys.

The accessibility of T2T genomes and comprehensive annotations holds immense potential for unraveling scientific questions related to primate evolution and human diseases. It provides a valuable opportunity to investigate the ecological and genetic factors that contribute to phenotypic diversity and adaptation, including the intriguing phenomenon of brain expansion in primates. Moreover, functional experiments in NHPs often require advanced experimental skills and lengthy assessments. In contrast to mouse models, as an alternative, NHP organoids can serve as a proxy system for conducting certain functional validations.

Ancient DNA (aDNA) analysis is also a powerful tool for gaining a better understanding of evolution and medical implications. While human aDNA has been extensively studied, aDNA from NHP fossils remains largely unexplored, even though extinct primates have been found worldwide, and primate fossils are abundant. Accessing aDNA from these samples would greatly advance our understanding of evolutionary novelties, biodiversity, and human diseases.

Data sharing and effective collaboration are essential for the success of consortium work. Ecological data collection, genome sequencing, and data analysis require significant efforts that cannot be accomplished by a single lab within a short period. Therefore, we call for close collaborations, whether through individual collaborations or funding agency initiatives. Collaboration and competition are not mutually exclusive, and working together while focusing on different biological aspects can lead to an exemplary win-win strategy. This cooperative approach has been successful in this primate consortium and many other scientific endeavors, such as the Human Pangenome Reference Consortium (HPRC) and the Chinese Pangenome Consortium (CPC).

REFERENCES

1. Wu, D.D., Qi, X.G., Yu, L., et al. (2022). Initiation of the Primate Genome Project. *Zool. Res.* **43**, 147–149.
2. Shao, Y., Zhou, L., Li, F., et al. (2023). Phylogenomic analyses provide insights into primate evolution. *Science* **380**, 913–924.
3. Rivas-González, I., Rousselle, M., Li, F., et al. (2023). Pervasive incomplete lineage sorting illuminates speciation and selection in primates. *Science* **380**, eabn4409.
4. Qi, X.G., Wu, J., Zhao, L., et al. (2023). Adaptations to a cold climate promoted social evolution in Asian colobine primates. *Science* **380**, eabl8621.
5. Mao, Y., and Zhang, G. (2022). A complete, telomere-to-telomere human genome sequence presents new opportunities for evolutionary genomics. *Nat. Methods* **19**, 635–638.

ACKNOWLEDGMENTS

We would like to express our sincere apologies for the limitations in discussing all the advancements made by the PGP consortium and for the inability to provide an extensive list of references because of space constraints. We acknowledge Dr. Bing Su (KIZ, CAS) for helpful comments on the manuscript draft and thank Rick Lee for proofreading. This study was funded by grants from the National Natural Science Foundation of China (NSFC) (3217040584 and 32000390 to Y.H.), the Youth Innovation Promotion Association of CAS (to Y.H.), the Science and Technology General Program of Yunnan Province (202301AW070010 and 202001AT070110 to Y.H.), and the Shanghai Pujiang Program (22PJ1407300) and Shanghai Jiao Tong University 2030 Program (WH510363001-7) (to Y.M.).

DECLARATION OF INTERESTS

The authors declare no competing interests.