

RESEARCH ARTICLE SUMMARY

NONHUMAN GENOMICS

Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility

Wesley C. Warren* *et al.*

INTRODUCTION: The rhesus macaque (*Macaca mulatta*) is one of the most widely used non-human primate (NHP) models for studying human biology and disease. As a representative of the Old World monkey lineage, its genetic sequence is also critical for studies of primate evolution.

RATIONALE: Because of the central role of rhesus macaques in both biomedical research and primate adaptation, we sought to generate a new reference genome for this NHP in which most gaps were closed and most protein-coding genes were annotated. A more comprehensively annotated macaque genome and extensive sequencing of individual macaques from existing research populations enables the characterization of standing genetic variation. Understanding the extent of genetic variation among research populations under phenotypic surveillance will identify new models of human genetic disease and allow for the further development of NHP models for investigating aspects of genome function such as gene regulation.

RESULTS: We sequenced and assembled the genome of a female rhesus macaque of Indian origin using a multiplatform genomics approach

that included long-read sequencing, extensive manual curation, and experimental validation. With the exception of humans, the resulting assembly is one of the most complete primate references to date, with 99.7% of the gaps now closed and >99% of the genes represented. We generated 6.5 million full-length transcripts and used these to create a comprehensive set of protein-encoding and noncoding gene models, including the identification of new macaque isoforms and gene candidates.

The more complete macaque genome overcomes many of the limitations of the previous assemblies. Segmental duplications are improved threefold, leading to the characterization of lineage-specific genes and gene families (e.g., *ZNF669*) that have expanded recently during evolution. Most full-length, active mobile elements have been resolved at the sequence level and are now integrated into the genome assembly instead of being fragmented and unassigned. In the case of LINEs, this has led to a reclassification of the order of appearance of active elements during Old World monkey evolution. Human-macaque gene comparisons identify a limited number of lineage-specific exon changes of potential functional effect,

including the formation of isoforms that distinguish the two species.

We generated whole-genome sequence data for 850 rhesus macaques from captive U.S. research colonies and three wild-caught Chinese samples, including 133 previously published samples. We used these data to identify 85.7 million single-nucleotide variants (SNVs; 21.3 million singletons) in addition to 10.5 million indels, generating the most extensive collection of segregating genetic variants for any NHP species. We can now confirm that research rhesus macaques are more than twice as diverse per individual as humans, with the average macaque carrying 9.7 million SNVs, and used this variation to understand the genetic diversity of existing research populations. We also identified potentially deleterious mutations in macaque genes that are intolerant to mutation in humans. Such mutations segregating in rhesus macaque research centers offer the opportunity to develop new genetic models of disease.

CONCLUSION: This new macaque reference genome and the genetic characterization of research populations will substantially advance biomedical research and studies of primate genome evolution by providing an improved framework for more complete studies of genetic variation and its phenotypic consequence. ■

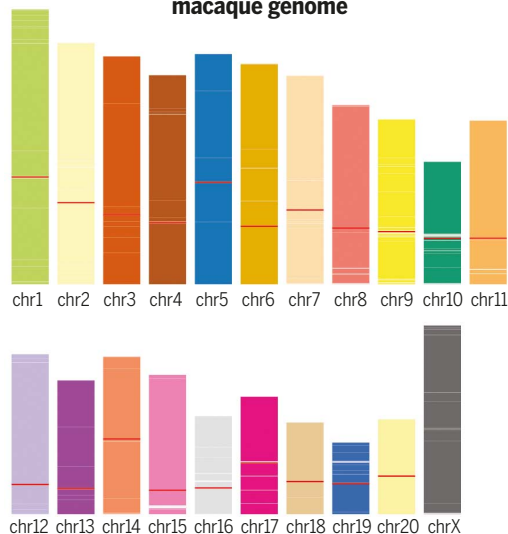
The list of author affiliations is available in the full article online.

*Corresponding author. Email: warrenwc@missouri.edu; jr13@bcm.edu (Jeffrey Rogers); eee@gs.washington.edu (Evan E. Eichler)

Cite this article as: W. C. Warren *et al.*, *Science* **370**, eabc6617 (2020). DOI: 10.1126/science.abc6617

READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abc6617>

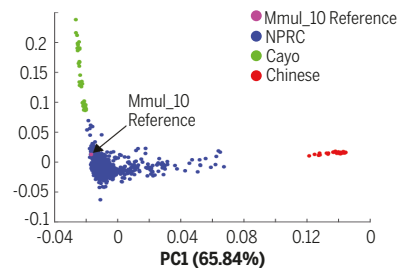
More complete and annotated Rhesus macaque genome



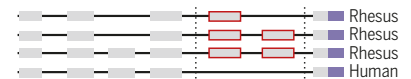
Sequencing of Rhesus macaque research populations



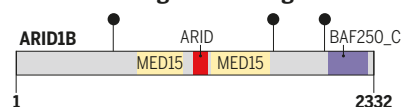
Diversity



Gene differences



Variant catalog in disease genes



Genetic diversity in the rhesus macaque. A more completely assembled and annotated macaque reference genome (left panel) coupled to sequencing of research populations (middle panel) provides a deep understanding of diversity, functional changes in gene models, and rare variants that may be used to develop better genetic models of disease (right panels). Photo credit: Kathy West.