Primate genomics

NELSON TING
University of Oregon, USA

Genomics is the branch of biology that is concerned with the content, structure, and function of genomes. A genome, in turn, is the complete set of an organism's DNA (see GENOMES), including both genes and intergenic regions. Genomics thus differs in scope from more traditional conceptions of the field of "genetics" by focusing on an organism's entire DNA instead of a single gene or a small set of genetic loci. A genomic approach has many advantages. Only through the sequencing and assembly of all of an organism's DNA can we understand how the content, structure, and function of the genome has changed throughout evolutionary history, which provides critical insight into the evolution of species-specific traits. Meanwhile, genome sequencing from multiple individuals within a species allows for more accurate inference of population history. Applied to primates, genomic approaches have led to a new and improved understanding of how and why primate variation exists both between and within taxa.

The genomic era was ushered in with the Human Genome Project (HGP) and the first published assembly of the human genome in 2001, which marked the first extensively sequenced vertebrate genome. This initial version was a draft assembly, meaning that it still contained thousands of gaps, that the order and/or orientation of its component pieces were not known with complete confidence, and that these pieces were only sequenced with relatively low redundancy (four to five times coverage of each nucleotide site). Since then, revisions to this draft human genome have been released, and the human genome is now considered "complete," although it is continually improved as new complex variation within our species is uncovered. The human genome is approximately 3 billion base pairs and includes approximately 20,000 genes (see GENOMES), suggesting that the majority of our DNA does not encode for proteins. Interestingly, approximately half of the human genome is comprised of repetitive elements (e.g., see TRANSPOSABLE ELEMENTS (*ALU*S, LTRs, LINEs, SINEs)).

Sequencing of the human genome helped pave the way for the sequencing of other primate genomes. The first nonhuman primate genome to be sequenced and published was that of the common chimpanzee (*Pan troglodytes*) in 2005, and publication of the rhesus macaque (*Macaca mulatta*) genome soon followed in 2007. Since then, most primate genome sequencing efforts have either focused on the apes (due to their close relationship to humans), on species from representative branches of the primate tree, and/or on taxa that serve as biomedical models for human disease. As a result, within-species genomic variation has been relatively well characterized for all of the great apes, and at least one representative genome has been published from each of the gibbon, cercopithecine, colobine, New World monkey, and lemur groups. Most of these nonhuman primate genomes are considered draft assemblies, although current genome sequencing projects are using newer technologies that make the sequencing and complete assembly of large and complex genomes much easier.

Comparative analysis of primate genomes has allowed us to better estimate the degree of genetic divergence between primate taxa. When comparing sequences that can be easily aligned across species, humans and chimpanzees differ by 1.1–1.4 percent, while humans and macaques differ by approximately 6.5 percent. However, while human and nonhuman primate genomes are similar in size and repetitive complexity, the proliferation of repetitive or duplicative events has been shown to vary among primate lineages. Thus, when insertions and deletions (see INDELS), repetitive elements, duplications, copy number

The International Encyclopedia of Biological Anthropology. Edited by Wenda Trevathan.

DOI: 10.1002/9781118584538.ieba0401
Based in part on Ting, N. (2017). Primate Genomics. In Fuentes, A. (ed.) *The International Encyclopedia of Primatology*, John Wiley & Sons, Inc., with permission.

variants (see COPY NUMBER VARIANTS (CNVs)), and gene family differences are included, genome divergence values increase. For example, when small insertion/deletion events are included, the estimated divergence between the human and macaque genomes is closer to 10 percent (Gibbs et al. 2007).

Analysis of genomic data also provides insights into how natural selection (see GENETIC ADAPTATION) may have shaped regions of the genome along specific primate lineages, and it is thus an important first step in developing hypotheses for the genetic basis of phenotypic differences among primate species. For example, the genome of *Rhinopithecus*, an Asian leaf-eating monkey, contains evidence of functional evolution of genes related to a folivorous diet (Zhou et al. 2014), and the genome of *Callithrix*, a small-bodied New World monkey that routinely gives birth to twin offspring, shows evidence of positive selection on genes involved with growth, metabolism, and reproduction (The Marmoset Genome Sequencing and Analysis Consortium 2014). From genomic studies we are also able to simultaneously analyze many different loci from across the genome, which provides more accurate pictures of primate population divergence, population size change through time, and genetic diversity. As a result, we know that post-divergence gene flow between primate lineages was sometimes common as migration and mating occurred between species after initial divergence. We also know that different primate species often show different patterns of past population size change, even when faced with similar environmental pressures, and that these patterns from the distant past can have a larger effect on modern genetic diversity than current census population size. For example, the Sumatran orangutan currently has more genetic diversity than the Bornean orangutan despite having a more restricted range and smaller census population size, which is likely the result of differences in past population size change since their divergence (Locke et al. 2011). Primatologists in turn can use this valuable information provided by genomic data in attempts to protect endangered species from extinction threats (see PRIMATE CONSERVATION GENETICS).

Primate genomics will continue to grow at a rapid pace as technology advances. This will allow for complete genome assemblies at lower cost, as well as breakthroughs in accessibility such as whole-genome sequencing while in the field. As this happens, sequenced genomes from a wider variety of primate species will reveal novel insights into primate evolution, and population-level sequencing will provide new information on evolutionary and demographic processes within species. Lastly, other types of "-omics" fields housed under the umbrella of "genomics" are still nascent as applied to nonhuman primates but will continue to advance. These typically attempt to further understand genome function through the analysis of gene products (e.g., proteomics, transcriptomics, see TRANSCRIPTOME), of alterations made to the genome within the lifetime of an individual (e.g., epigenomics, see EPIGENETICS), or of interactions between primate genes and environmental genes (e.g., "metagenomics," see MICROBIOME). Such action will be crucial in providing a deeper understanding of primate genome function to address how and why variation exists among our closest relatives.

SEE ALSO: Exome; Genetic markers (SNPs/Satellites/STRs/RFLPs); Massively parallel sequencing; Molecular clock; Regulatory elements/regulatory genes

REFERENCES

Gibbs, Richard A., Jeffrey Rogers, Michael G. Katze, Roger Bumgarner, George M. Weinstock, Elaine R. Mardis, Karin A. Remington, et al. 2007. "Evolutionary and Biomedical Insights from the Rhesus Macaque Genome." *Science* 316 (5822): 222–34.

Locke, Devin P., LaDeana W. Hillier, Wesley C. Warren, Kim C. Worley, Lynne V. Nazareth, Donna M. Muzny, Shiaw-Pyng Yang, et al. 2011. "Comparative and Demographic Analysis of Orang-Utan Genomes." *Nature* 469 (7331): 529–33.

The Marmoset Genome Sequencing and Analysis Consortium. 2014. "The Common Marmoset Genome Provides Insight into Primate Biology and Evolution." *Nature Genetics* 46 (8): 850–57.

Zhou, Xuming, Boshi Wang, Qi Pan, Jinbo Zhang, Sudhir Kumar, Xiaoqing Sun, Zhijin Liu, et al. 2014. "Whole-Genome Sequencing of the Snub-Nosed Monkey Provides Insights into Folivory and Evolutionary History." *Nature Genetics* 46 (12): 1303–10.

FURTHER READING

Rogers, Jeffrey, and Richard A. Gibbs. 2014. "Comparative Primate Genomics: Emerging Patterns of Genome Content and Dynamics." *Nature Reviews. Genetics* 15 (5): 347–59.

Ting, Nelson, and Kirstin N. Sterner. 2013. "Primate Molecular Phylogenetics in a Genomic Era." *Molecular Phylogenetics and Evolution* 66 (2): 565–68.