APPLICATIONS OF NEXT-GENERATION SEQUENCING

Comparative primate genomics: emerging patterns of genome content and dynamics

Jeffrey Rogers and Richard A. Gibbs

Abstract | Advances in genome sequencing technologies have created new opportunities for comparative primate genomics. Genome assemblies have been published for various primate species, and analyses of several others are underway. Whole-genome assemblies for the great apes provide remarkable new information about the evolutionary origins of the human genome and the processes involved. Genomic data for macaques and other non-human primates offer valuable insights into genetic similarities and differences among species that are used as models for disease-related research. This Review summarizes current knowledge regarding primate genome content and dynamics, and proposes a series of goals for the near future.

Hominins
Members of the evolutionary lineage leading to humans after divergence from the ancestors of chimpanzees. Hominins include species that are directly ancestral to modern humans and related species such as Neanderthals or older branches such as australopithecines.

Human Genome Sequencing Center and Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. Correspondence to J.R.
e-mail: jr13@bcm.tmc.edu
doi:10.1038/nrg3707
Published online 8 April 2014

Current technologies for large-scale DNA sequencing have opened new avenues for the study of non-human primate genomes. Although the major focus of genomic research is human genetics and its relationship to disease, investigators are also pursuing comparative primate genomics. Two basic motivations exist for the detailed study of non-human primate genomes: first, this information can be applied in studies using primates as models for the analysis of human disease; second, comparative evolutionary analyses can reconstruct the history and mechanisms of genomic change, with a particular focus on the origin of the human genome. One unexpected outcome from new genomic data for non-human primates (such as chimpanzees, bonobos, gorillas and orangutans) and humans is a new perspective on the process of speciation and genetic divergence among these evolutionary lineages.

The first non-human primate genome sequenced and published was that of the chimpanzee (*Pan troglodytes*)[1], followed soon by that of the rhesus macaque (*Macaca mulatta*)[2]. Both genomes were analysed using shotgun sequencing that used exclusively Sanger sequencing methods. As a result, these projects entailed considerable cost and effort. A legacy of further primate sequencing projects that were initiated when Sanger sequencing was the only option is now reaching its end (TABLE 1). The genomes of all extant great ape species have been sequenced to draft quality. Analysis of a gibbon genome — the only remaining group of extant hominoids to be sequenced — is underway, and other non-human primate genome assemblies are at various stages of completion (see Supplementary information S1 (table)). Remarkably, researchers have obtained extensive sequence information for two extinct hominins — the Neanderthals[3] and the Denisovans[4] (BOX 1) — as well as substantial data on primate transcriptomes and genetic variation within species.

The widespread availability of next-generation sequencing technology promises even more rapid progress in this area. The amount of genomic information available for non-human primates is certain to grow at an accelerating pace. Our understanding of comparative primate genome content, diversity and evolution will necessarily change as new data appear. Conclusions based on current information may therefore be amended soon. Nevertheless, substantial progress has been made in the past several years, which justifies an assessment of the insights gained so far.

In this Review, we begin by summarizing available information about the content of and differences among primate genomes. Next, we present new insights regarding genomic differentiation and speciation with a particular reference to human evolution. Finally, we illustrate some of the ways that genomic data are expanding and improving the use of non-human primates in studies of human health and disease.

Old World monkeys
Members of the branch of primates that includes extant anthropoid primates (monkeys) native to Asia and Africa; they belong to the superfamily Cercopithecoidea.

New World monkeys
Members of the branch of primates that includes extant anthropoid primates (monkeys) native to South and Central America; they belong to the parvorder Platyrrhini.

Genomic differences among primates

Comparisons of annotated genome sequences across species allow researchers to directly identify genomic elements that are shared and that are species specific. We have known for many years that protein-coding sequences show greater similarity between primate species than intronic and intergenic sequences. Past studies also indicated that a large proportion of human protein-coding genes are found in most or all primates. However, detailed comparisons of all components of the genome have been impossible until recently. Investigators can now address fundamental questions concerning the content and function of genomic features across multiple species, thus providing new insights into the genetic bases of phenotypic similarity and differences among humans and other primates.

Differences in single-copy alignable sequence. With time, genomes accumulate mutations (for example, single base-pair substitutions) that may, through genetic drift or selection, become fixed differences that distinguish one species from its close relatives. Divergence in single-copy sequences occurs steadily among primate genomes, but such divergence does not arise at a uniform rate in all branches of the primate tree. Alignment of sequences across species shows that pairwise differences between species correlate fairly well with evolutionary divergence times inferred from other information. The human–chimpanzee sequence divergence is estimated to be 1.1–1.4%[1,5]. The time of separation of the human lineage from the chimpanzee lineage remains controversial[6] but is generally dated to 9–5 million years ago (FIG. 1). The uncertainty concerning the date of the human–chimpanzee divergence results from several factors, including the lack of a reliable paleontological record for that event and ambiguity concerning the appropriate mutation rate for inferring the divergence time from DNA sequences alone. The difference in single-copy sequence between humans and rhesus macaques is ~6.5%[2], and the divergence of these two lineages is more confidently dated to 28–25 million years ago. Dates for the human–chimpanzee divergence calculated using estimates of mutation rate that are derived from other inter-species differences (for example, the human–macaque or human–orang-utan divergence) differ from dates estimated on the basis of mutation rates obtained through pedigree-based analyses of current human mutation.

Despite these uncertainties, various analyses suggest that single-copy DNA accumulates individual base-pair substitutions more slowly through time in gorillas, chimpanzees, bonobos and humans than in other primates such as Old World monkeys or New World monkeys. This is not entirely unexpected given differences in generation time[7,8]. One exception may be the aye-aye (*Daubentonia madagascarensis*), which is a Malagasy lemur with an extraordinarily unique morphology. Synonymous substitutions are reported to accumulate more slowly than expected in the aye-aye on the basis of comparisons with other species[9]. Additional sequencing projects may identify other primate lineages that do not fit current expectations.

Small insertions and deletions. Although the difference in genome sequences between humans and chimpanzees is recently estimated at 1.4%[5], this is correct only for nucleotide substitutions in regions where the two genomes can be directly aligned. As one study[10] first noted, small insertions and deletions (indels; <100 bp) account for more total nucleotide differences among closely related species than single base-pair changes in alignable sequences. The human and chimpanzee genomes each contain ~1.5% of unique sequences that are not found in the other primarily owing to small indels. The alignable rhesus macaque sequence is 93.5% identical to the human genome, but the two genomes are only 90.8% identical when small indels are included[2]. These indel differences among species are found more frequently in intronic and intergenic regions than in protein-coding exons, primarily because indels in protein-coding sequences will generally have negative consequences on protein function. For example, only

Table 1 | **Published primate genome sequences**

Common name	Species name	Bases in contigs	Contig N50*	Scaffold N50*	Refs
Draft genome assemblies					
Chimpanzee	*Pan troglodytes*	2.7 Gb	15.7 kb	8.6 Mb	1
Chimpanzee (updated)	*P. troglodytes*	2.9 Gb	50.7 kb	8.9 Mb	116
Bonobo	*Pan paniscus*	2.7 Gb	67 kb	9.6 Mb	65
Gorilla	*Gorilla gorilla*	2.8 Gb	11.8 kb	914 kb	5
Orang-utan	*Pongo abelii*	3.1 Gb	15.5 kb	739 kb	13
Indian rhesus macaque	*Macaca mulatta*	2.9 Gb	25.7 kb	24.3 Mb	2
Chinese rhesus macaque	*M. mulatta*	2.8 Gb	11.9 kb	891 kb	92
Vietnamese cynomolgus macaque	*Macaca fascicularis*	2.9 Gb	12.5 kb	652 kb	92
Aye-aye	*Daubentonia madagascarensis*	3.0 Gb	NA	13.6 kb	9
Sanger genome sequences at 2× whole-genome coverage					
Mouse lemur	*Microcebus murinus*	1.8 Gb	3.5 kb	141 kb	11
Bushbaby‡	*Otolemur garnettii*	2.4 Gb	27.1 kb	13.9 Mb	11,117
Tarsier‡	*Tarsius syrichta*	3.4 Gb	38.2 kb	401 Mb	11,118
Whole-genome resequencing studies without assembly					
Indian rhesus macaque	*M. mulatta*	NA	NA	NA	33
Chinese rhesus macaque	*M. mulatta*	NA	NA	NA	119
Mauritian cynomolgus macaque	*M. fascicularis*	NA	NA	NA	91
Malaysian cynomolgus macaque	*M. fascicularis*	NA	NA	NA	115

NA, not applicable. *N50 is a weighted median statistic such that 50% of the entire assembly is contained in contigs or scaffolds that are equal to or larger than this value. ‡The bushbaby and tarsier genomes were published as 2× Sanger genomes, but these statistics reflect subsequent unpublished upgrades of these genomes.

Box 1 | **Genomic analysis of ancient hominins**

The fossil record for recent human evolution (that is, the past several hundred thousand years) is substantial. A great deal is known about morphology, biogeography and the archaeological evidence for behaviour concerning several extinct hominin species. Remarkably, through marked advances in techniques for investigating ancient DNA, we now have access to extensive genome sequence data for Neanderthals — an extinct hominin population from Europe and western Asia that diverged at least 250,000 years ago from the lineage leading to modern humans[3]. This work has shown that 1–4% of DNA sequences carried by modern humans outside Africa are derived from Neanderthals as a result of interbreeding and gene flow[54]. Another extinct hominin population (the Denisovans) was only recently recognized using genome sequence produced by extracting DNA from a finger bone found in the Altai Mountains[4]. The Denisovans diverged from human ancestors 700,000–170,000 years ago. Gene flow from the Denisovans into the modern human population has so far been detected only among aboriginal Australians and populations in Melanesia and southeast Asia[4]. These findings indicate that ancestral human populations interbred to some biologically significant degree with other populations that were distinct in their genetics and, at least in the case of Neanderthals, distinct in morphology as well. There is also evidence that introgression from Neanderthals into modern humans introduced alleles that are now associated with disease among modern humans, and that negative selection after this hybridization may have been driven by adverse effects of that hybridization on male fertility[110].

indels involving a multiple of three nucleotides do not induce frameshifts that result in substantial changes in the encoded amino acid sequence. Available comparisons show that small indels are most common in non-coding regions, presumably because they are better tolerated in these regions. However, as more non-coding regions with functional importance are identified[11,12], some indels in flanking or intergenic segments will become increasingly interesting and may potentially gain importance for understanding changes that affect enhancers and other regulatory sequences that influence gene expression and phenotypic differences among species.

Alu and other repetitive elements. The insertion of Alu repeats and other retrotransposons is an ongoing process in primate genomes. Repetitive elements collectively make up ~50% of the total genome in humans, apes and monkeys, but the number of species-specific insertions differs substantially across species, from ~5,000 in humans to 2,300 in chimpanzees and only 250 in orang-utans[13]. It is not entirely clear why the rate of accumulation differs. Nevertheless, *de novo* Alu insertions constitute a major source of genomic change but have not affected all primate species equally[14]. Retrotransposons also facilitate duplication or deletion events, which affect much larger DNA segments[15] and can thus have broader effects on gene and genome content.

Copy-number differences and gene family changes. The majority of protein-coding genes have 1:1 homologues among humans, the great apes and Old World monkeys sequenced so far, but gene content is not identical among species. Particular gene families have expanded or contracted in individual lineages. For example, 1,358 genes were identified as new duplications in the rhesus macaque genome compared with the human genome[2]. The major histocompatibility complex (MHC) gene cluster, which is crucial for response to pathogens and for other immunological processes, is expanded in macaques relative to humans[16]. Other interesting cases are changes in genes encoding zinc-finger transcription factors, which show gains and losses that distinguish between the genomes of humans, chimpanzees and orang-utans[17], as well as the marked expansion of genes encoding proteins with DUF1220 domains in humans[18–20], which might be related to the expansion of brain size.

However, the draft quality of current non-human primate genome assemblies makes it difficult to define all copy-number variations accurately. One can compare gene lists from different assemblies, but gaps and other issues in these assemblies create ambiguity[21,22]. Available evidence suggests that humans and chimpanzees underwent more rapid changes in gene copy number than orang-utans and rhesus macaques[13]. Among the great apes, gorillas show more copy-number variants than others[5]. Complete analyses await additional data, including better genome assemblies and information concerning copy-number polymorphism in non-human primates.

Segmental duplications. Segmental duplications (that is, chromosomal regions >1 kb that are >90% identical to other segments in the same genome) are an important aspect of primate genome structure and dynamics. Duplication and deletion of these segments are active in the human genome. Some of these mutations are apparently neutral, but many lead to adverse consequences and disease[23]. Similarly to the way that segmental duplications create variation among humans, these duplications are 'drivers' of evolutionary change across primate genomes. About 5% of the human and chimpanzee genomes, and 3.8% of the orang-utan genome, consist of segmental duplications[13,24]. The human and great ape genomes are enriched with dispersed duplications, as they were subjected to an interval after their divergence from Old World monkeys when the production of new duplications was particularly active[25,26]. Many expansions of specific protein-coding gene families result from segmental duplications, which sometimes involve repeated expansions of a given sequence[24,27,28]. Some genes within segmental duplications show evidence of positive selection on both protein-coding sequence and copy number[18,20,29,30]. Among the great apes, some of these expansion events have occurred as independent parallel events in different lineages, which strengthens the interpretation that these genomic changes are often the result of positive selection on both gene copy number and nucleotide sequences[25,31].

Genetic variation within primate species

Individual primate genome projects have assessed intra-species genetic variation in different ways, and a broad range of sample sizes and population genetic parameters were used to quantify variation. Earlier studies that analysed small samples or only small portions of the genome had suggested that, for the most part, non-human primate species have higher levels of intra-species genetic variation than humans[32–35], and this pattern

Positive selection
Natural selection acting on phenotypes and the relevant DNA sequences that results in directional change towards a new sequence and phenotype. It is in contrast with negative selection, which eliminates deleterious traits and therefore acts against any new mutations that generate them.

Effective population sizes
A basic concept from population genetics that describes the number of individuals required in an ideal breeding population (that is, equal numbers of breeding males and females, with equal reproductive success among them) of constant size to sustain a given amount of intra-population genetic variation. As genetic variation in a given population is affected by current and past demographic factors, estimation of effective population size allows researchers to infer aspects of population history.

holds in the larger data sets published more recently. Great ape species have all been reduced to low total population census numbers, but studies using whole-genome data indicate that genetic diversity within great ape species is consistent with effective population sizes as large as, or even larger than, that of humans[13,24]. The Great Ape Genome Project investigated genome-wide variation within and between all six great ape species[36] and found that some subspecies and species show levels of intra-species diversity (FIG. 2) that are approximately equivalent to that of non-African humans. Chimpanzees from east and central Africa, the Nigeria–Cameroon

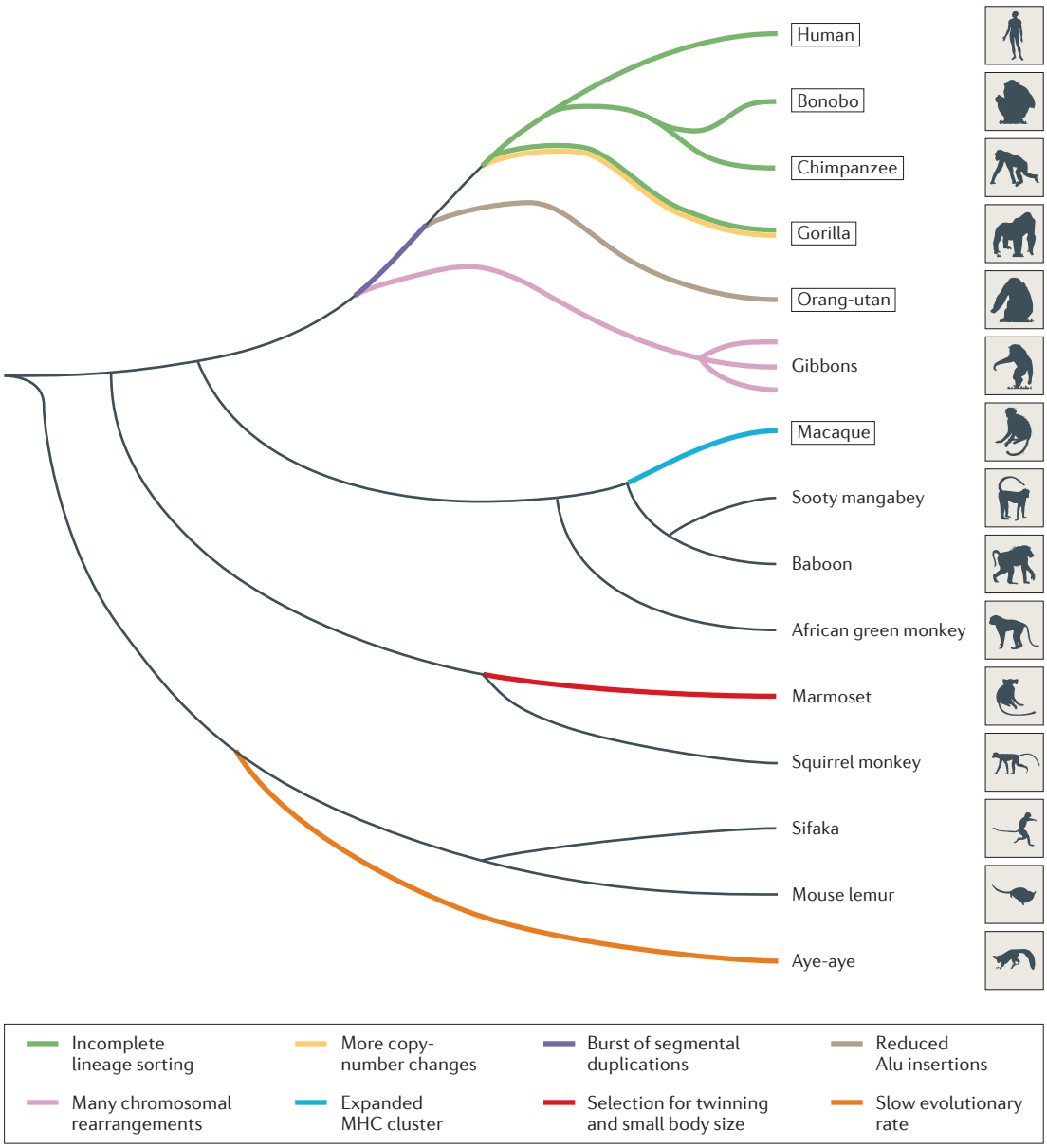


Figure 1 | **Primate phylogenetic tree.** The evolutionary relationships are shown among species for which genome sequences are published, available or in progress. The genomes for species enclosed in boxes are already published. Among the macaques, genomes of the rhesus and cynomolgus macaques are published[2,91,92], but sequencing of other macaque genomes is underway (see Supplementary information S1 (table)). Selected lineages are highlighted to indicate specific genomic features of interest or unexpected genomic traits, such as reduced rate of Alu insertion in the orang-utan genome[13] and the slower evolutionary rate in the aye-aye genome[9]. The time of separation of the human lineage from the rhesus macaque lineage is dated at 28–25 million years ago, whereas the human and chimpanzee lineages are believed to have diverged ~9–5 million years ago. Given that the dates of many branching points remain controversial — for example, owing to the lack of a reliable paleontological record for that event or to ambiguity concerning the appropriate mutation rate to use for inferring the time of divergence from DNA sequences alone — no dates have been added to the phylogenetic tree. MHC, major histocompatibility complex.

chimpanzee subspecies, western lowland gorillas and both orang-utan species exhibit significantly higher levels of intra-species variation than that found among humans. Application of coalescent models and incomplete lineage sorting (ILS) models enabled the researchers to re-estimate effective population sizes for different great ape species and subspecies. Each species has a unique history of population expansion and decline, and each chimpanzee subspecies has an independent history[36]. This observation of separate and unique demographic histories for regional populations or subspecies within a species is likely to be true for most or all primates[34].

Although the great apes have high intra-species diversity (despite low present-day census sizes), the rhesus macaque is much more widely distributed geographically and has larger extant populations (FIG. 2). As part of the rhesus genome sequencing project, DNA from Chinese and Indian rhesus macaques was sequenced for 150 kb from 5 genomic regions[34]. The density of single-nucleotide polymorphisms (SNPs) was significantly higher than that in human populations. Only about one-third of SNPs were shared between the two geographical populations, which indicates that most variation is region specific. Similar results were obtained in a study of 3′ untranslated regions of transcripts in a small number of macaques[37]. A study of whole-genome sequences of 3 Indian rhesus macaques found >3 million variants in at least 2 of the data sets examined[33]. About 14 million SNPs, including singletons, were found[33], and such density is substantially greater than that in humans (in which the average number of SNPs, including singletons, per individual was estimated at 3.6 million by the 1000 Genomes Project)[38].

Overall levels of variation are high in non-human primates, including additional non-hominoid species that have very small population sizes and that are in serious danger of extinction[39]. However, the amount of functionally important variation within particular species is not yet clear. Substantial numbers of nonsynonymous substitutions that are predicted to be possibly damaging have been identified in even small numbers of macaques[33,40], but a recent comparison of protein-coding variation between humans and rhesus macaques found little difference between these species[41]. It is possible that macaques and other non-human primates are segregating greater total intra-species variation than humans but equivalent levels of damaging or adverse variation.

Differences in gene expression

A. Wilson and his colleagues predicted years ago that much of the adaptively significant phenotypic change that distinguishes one species from others results from changes in gene expression rather than from mutations in protein-coding sequences[42,43]. Recent information on comparative primate gene expression is consistent with this prediction. It is likely that a large proportion of adaptive evolution involves changes in transcription factor binding, which possibly rivals adaptation through protein evolution[44]. Overall, the description of non-human primate transcriptomes lags behind the corresponding knowledge for humans and mice, but researchers are now developing larger information resources, such as the US National Institutes of Health Blueprint Non-Human Primate Atlas of gene expression for the rhesus macaque brain and the Non-human Primate Reference Transcriptome Resource[45].

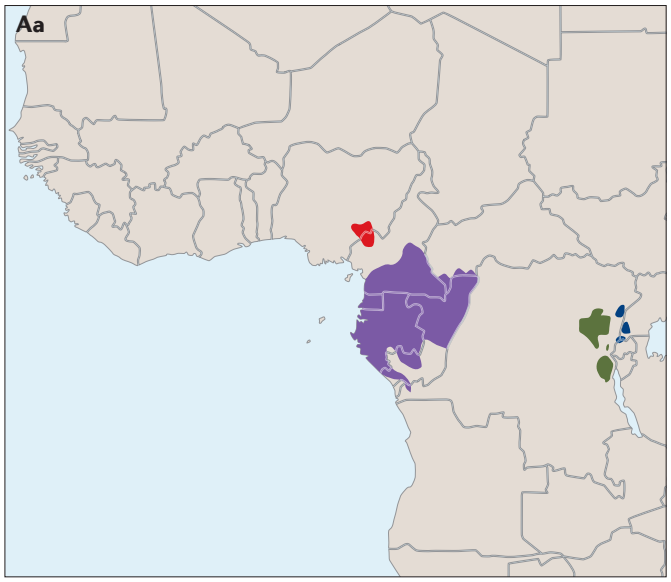
Nevertheless, comparisons of gene expression across primates have already proven valuable. Differences in gene expression among humans, chimpanzees and rhesus macaques are influenced by natural selection[46–48] and include substantial differences in alternative splicing[47,49]. RNA sequencing from the livers of humans and other mammals, including 11 primates, found strong evidence for positive selection in various expressed genes[39]. The results of this study show enrichment for changes that affect genes involved in peroxisome function, such as gamma-glutamyl hydrolase (*GGH*), peroxisomal biogenesis factor 7 (*PEX7*) and 2-hydroxyacyl-CoA lyase 1 (*HACL1*)[39]. Patterns of DNA methylation in the prefrontal cortex differ between humans and chimpanzees, and correlate with differences in gene expression[50]. Three-way comparisons find greater overall similarity in gene expression between chimpanzees and humans than that between either of these species and gorillas, which is in agreement with the overall phylogeny (FIG. 1). Notably, however, genes in specific regions of the genome (that is, chromosomal segments that show ILS (see below)) show a contrary pattern[5]. Studies of gene expression have recently been extended to wild populations of baboons[51] — an approach with exceptional potential for future discoveries.

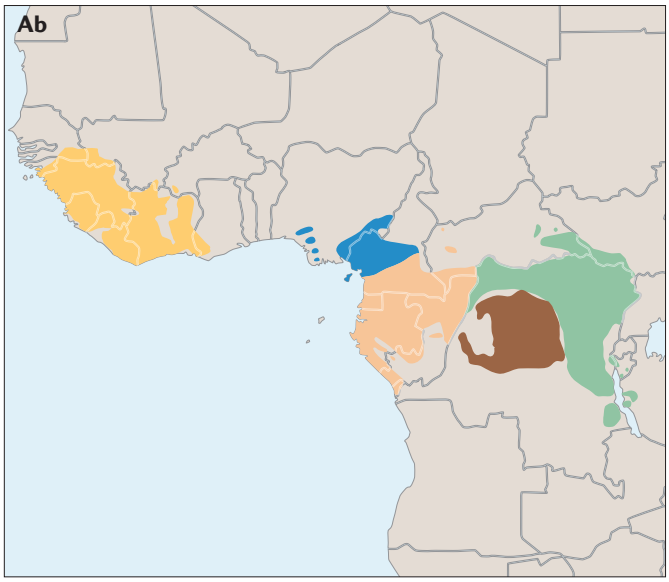
Primate evolutionary dynamics

One of the primary motivations for studying comparative primate genomics is the desire to understand the origin of the human genome. Whole-genome information that is now available for our closest relatives is altering and refining ideas about the processes of speciation, diversification and genome evolution for this clade. Although this new picture is still incomplete, it reveals previously unappreciated complexity in the processes that produced the modern human genome. The theory and modelling of speciation (BOX 2) are complex topics with a long history and an extensive amount of literature, and this is therefore outside the scope of this Review. However, the insights concerning gene exchange among the early human and chimpanzee ancestors[52], as well as among ancient hominins[3,4,53,54] (BOX 1), are remarkable indications that this history is of greater interest than previously recognized. In parallel, these inter-species comparisons are greatly increasing our ability to identify genes or genomic regions that have undergone positive selection during recent human evolution, thereby indicating genetic changes and phenotypes that have been important in both human and non-human primate adaptation. Comparisons also show that fundamental genetic processes, such as recombination, can undergo rapid changes, as local hot spots of recombination are not conserved in humans and chimpanzees despite the overall high sequence similarity and the general conservation of large-scale patterns of recombination[55].

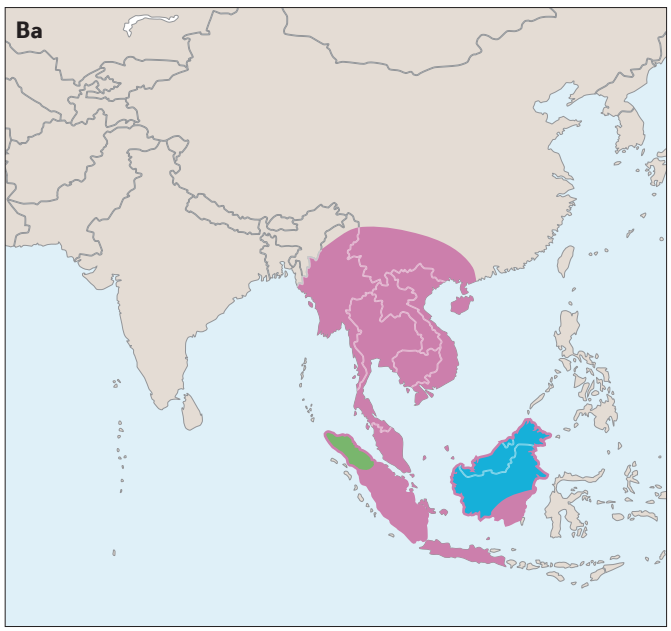
Coalescent models
Used in population genetics to investigate various aspects of population history and dynamics, these models are based on the genealogy or relationships within a gene tree among alleles of a specific DNA sequence. All alleles found in a population or a set of related populations can be traced back to a common ancestral sequence, and the statistical properties of those allelic relationships are exploited to investigate questions of population genetics and history.

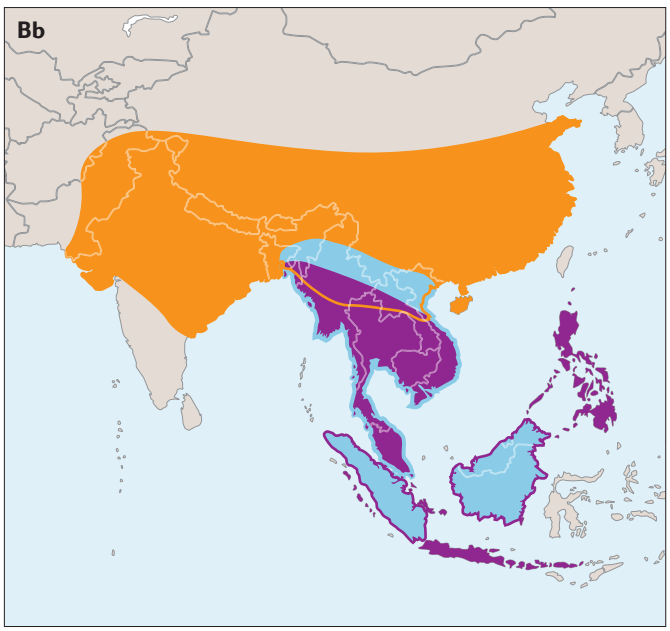
Incomplete lineage sorting
(ILS). The process by which, as a result of segregation of an ancestral polymorphism, the evolutionary relationships among a series of homologous DNA sequences in a set of distinct populations do not match the phylogenetic relationships among the overall populations; that is, the gene trees do not match the population trees.





			Species and subspecies	Genetic variation	Conservation status
Aa		1	Cross river gorilla	0.001 < Het < 0.0015	Critically endangered
		2	Western lowland gorilla	0.0015 < Het < 0.002	Critically endangered
		3	Eastern lowland gorilla	Het < 0.001	Endangered
		4	Mountain gorilla	NA	Critically endangered
Ab		5	Western chimpanzee	Het < 0.001	Endangered
		6	Nigeria–Cameroon chimpanzee	0.001 < Het < 0.0015	Endangered
		7	Central chimpanzee	0.0015 < Het < 0.002	Endangered
		8	Eastern chimpanzee	0.0015 < Het < 0.002	Endangered
		9	Bonobo	Het < 0.001	Endangered
Ba		10	Bornean orang-utan	0.0015 < Het < 0.002	Endangered
		11	Sumatran orang-utan	Het > 0.002	Critically endangered
		12	Gibbons and siamang	0.001 < Het < 0.002	Endangered and critically endangered
Bb		13	Rhesus macaque	Het > 0.002	Least concern
		14	Cynomolgus macaque	Het > 0.002	Least concern
		15	Pigtailed macaque	NA	Vulnerable





◀ Figure 2 | **Geographical distribution and genetic variation in selected primates.** Although most non-human primate species investigated so far have modest or small current population sizes and are, in most cases, either endangered or critically endangered (See the International Union for Conservation of Nature and Natural Resources (IUCN) Red List of Endangered Species), they have substantial levels of intra-species genetic variability. **A** | The approximate geographical distributions of African ape (gorilla, chimpanzee and bonobo) species and subspecies are shown, although actual population distributions are generally discontinuous and semi-isolated within these highlighted areas. **Ba** | Approximate geographical distributions for gibbons, siamangs and orang-utans are shown. As with the African apes, population distributions are not continuous in the areas indicated. **Bb** | Approximate geographical distributions for three macaque species (rhesus, cynomolgus and pigtailed macaques) are shown. The table presents the level of genetic variation for each species estimated either through the Great Ape Genome Project[36], or other studies of macaques[33,34] or gibbons[66]. Whole-genome data regarding genetic variation in mountain gorillas and pigtailed macaques are not available (NA). Conservation status is indicated following current IUCN Red List designations. Het, heterozygosity. Parts **Aa** and **Ab** are modified, with permission, from REF. 5 and REF. 114, respectively, © (2012) Macmillian Publishers Ltd. All rights reserved. Data in parts **Ba** and **Bb** from REF. 72.

Signatures of selection. There is substantial evidence for positive selection on protein-coding genes in various non-human primate species. Two classes of genes provide consistent evidence of positive selection: those involved in the immune system and pathogen resistance; and those involved in reproductive biology and gametogenesis[2,5,13,56]. These results are reasonable, as the constant pressure of infectious disease is a plausible driver of selection on the primate immune system, whereas reproductive competition within species is likely to account for the evidence of selection on those systems. Within individual species, positive selection has been detected for a wide range of phenotypes. The whole-genome comparisons among humans, chimpanzees and gorillas suggest that these three species have been subjected to approximately equal levels of positive selection[5]. This analysis also indicates shared positive selection in hominoids on traits that are related to neurodevelopment and brain morphology. The orang-utan genome provides evidence of selection on glycolipid metabolism and hearing acuity that is specific to that lineage[13]. Studies in marmosets and other callitrichine primates provide evidence for selection on multiple genes that are related to phyletic reduction in body size and the development of a unique form of dizygotic twinning, in which co-twins exchange haematopoietic stem cells early in gestation and consequently become life-long haematopoietic chimaeras[57].

Genes involved in the evolution of unique human traits have received much attention and have been reviewed elsewhere[58]. Detailed description of the genetic basis of human-specific traits has obvious interest to evolutionary biologists, anthropologists and the broader public. Studies have now attributed human-specific adaptations to deletions of gene regulatory elements (that is, enhancers)[59], rapid lineage-specific evolution of such elements[60], changes in gene copy number[19,61] and other types of genetic changes[58].

Initial genomic divergence and incipient speciation. Whole-genome sequence data from humans, chimpanzees and gorillas are not consistent with simple models of reproductive isolation, allopatric genetic divergence or the rapid development of species boundaries (BOX 2). Two processes that are known to shape genome evolution in other groups of animals — ILS and gene flow (FIG. 3) — are now crucial elements in discussions concerning the historical mechanisms that caused the differentiation between human and non-human primate genomes.

Incomplete lineage sorting. ILS occurs when a polymorphic ancestral species that has two or more alleles (that is, haplotypes) at a given locus divides into two lineages. Both alleles can be retained in the descendant branches, and when one of these lineages divides again, the phylogenetic tree for that locus (that is, the gene tree) may or may not match the branching order for the species-level evolutionary tree (FIG. 3a). The likelihood of discrepancy between the species-level phylogeny and any particular gene tree increases either as the time between the two successive branching events decreases or as effective population size increases[62]. Prior analyses of a few genes suggested that different regions in the human, chimpanzee and gorilla genomes show different evolutionary relationships (that is, different gene trees)[63,64]. Following assembly of the gorilla genome[5], researchers determined the evolutionary relationships for arbitrary segments across the human, chimpanzee and gorilla genomes. As expected, they found that for most of the genome, chimpanzees are more closely related to humans than to gorillas. However, for ~15% of the genome, chimpanzee DNA sequences share a more recent common ancestor with the homologous sequences in the gorilla genome than with those in the human genome. For another 15%, gorillas and humans are most closely related. ILS from a polymorphic common ancestor is a probable contributing factor, although gene flow among differentiating lineages (FIG. 3b) (see below) may also be implicated. This developing picture of evolutionary process complexity also applies in other cases. Bonobos and chimpanzees are undoubtedly sister taxa and are more closely related to each other than to any other species. Nevertheless, for 1.6% of the genome, sequences in bonobos are more similar to homologues in humans than to those in chimpanzees, whereas for 1.7% of the genome, chimpanzees are more closely related to humans than to bonobos[65]. ILS is the probable explanation and may also be common in other primates[66].

Gene flow among incipient lineages. Analyses of the human, chimpanzee and gorilla genomes indicate that genetic exchange between divergent lineages is not restricted to recent periods (BOX 1). The common ancestor of humans and chimpanzees began differentiating 12–5 million years ago, depending on the assumed mutation rate[1,52]. Using coalescent models, one study[52] estimated that those diverging lineages were subjected to reciprocal gene flow for ~3 million years; that is, the separation of the last common ancestor of humans and chimpanzees into two independent lineages was not a rapid event but included an extended period of progressive genetic divergence that was simultaneous with gene flow[52]. Such divergence with continuing gene

Allopatric
Pertaining to separate, non-overlapping geographical distributions.

Box 2 | **Initial genomic divergence and incipient speciation**

The theory and modelling of speciation are complex topics that have generated a large amount of discussion. Historically, the founders of the modern evolutionary synthesis (for example, E. Mayr and T. Dobzhansky) argued that genetic and reproductive isolation among populations precedes phenotypic and/or genetic differentiation that is substantial enough to justify recognizing those populations as distinct species[111]. Mayr's "biological species concept" and the allopatric speciation model long dominated discussion[111]. By contrast, the model of punctuated equilibrium[112] posited that most adaptively important genetic differentiation occurs during or immediately after initial divergence and isolation of incipient species, and a subsequent study suggested that punctuational episodes of evolution may have an important role in promoting evolutionary divergence in some cases[113]. More recently, other models and theories have addressed the greater complexity that is now known to be inherent in speciation and the genetic differentiation of many lineages[86–89]. For various types of species, the process of genetic divergence and incipient speciation seems to be more complex than that proposed by the traditional allopatric speciation model.

flow is also true for Bornean and Sumatran orang-utans (*Pongo pygmaeus* and *Pongo abelii*, respectively)[13,52].

Whole-genome data are not yet available for enough species of Old World or New World monkeys, or strepsirrhine primates, to support similar analyses of evolutionary divergence and exchange across partially isolated lineages. However, smaller data sets suggest that the complexity documented for humans and apes is common across primates. The number of documented hybrid zones between morphologically and/or behaviourally distinct primate populations is increasing[67]. Active hybrid zones facilitate the study of the process of genetic differentiation, including demographic and phenotypic correlates of hybridization. Various reviews of primate hybridization are available[67,68], but some examples will illustrate general principles.

Baboons (genus *Papio*) exhibit unusual phenotypic diversity and evolutionary complexity[69,70]. Baboon taxonomy has been controversial, but six morphologically distinct species with parapatric geographical ranges are now widely recognized[69,71–73]. Hybridization occurs at locations where distinct baboon 'types' meet[73–75], despite morphological differences and an increased frequency of developmental abnormalities in hybrids[76]. The evidence suggests a long history of gene flow[71,77,78]. Among baboon species the gene trees from different non-recombining genetic elements (such as mitochondrial DNA (mtDNA) and Y chromosomes) do not necessarily match observable phenotypic similarity among populations[69,71,77].

Rhesus macaques and cynomolgus macaques (*Macaca fascicularis*) are closely related but are universally regarded as separate species[72]. Across mainland Indochina, similarly to the African baboons, these macaque species form a hybrid zone (FIG. 2Bb) with apparently substantial gene flow. Y chromosomes from rhesus macaques are found in animals that are phenotypically cynomolgus macaques[79]. Autosomal gene flow occurs from rhesus macaque into cynomolgus macaque populations, and such gene flow affects only mainland populations of cynomolgus macaques but not other populations that are isolated on Indonesian islands and in the Philippines[80]. Thus, these two species present clear evidence for gene flow between well-differentiated species.

One recently discovered species of African monkey (*Rungwecebus kipunji*) shows evidence of ancient hybridization. Of the two geographically isolated populations of kipunji[81,82], one carries mtDNA sequences that are associated with *Cercocebus* sp. mangabeys, whereas the other carries mtDNA sequences that are more closely related to *Papio* sp. baboons. Despite this apparent genetic introgression (also known as introgressive hybridization) from baboons, the second population retains morphological and nuclear DNA features that are similar to those of its conspecific sister population. In another case of phylogenetic complexity, the genus *Cercopithecus* (commonly known as guenons) contains ~24 species[72,83], but their phylogeny has proved difficult to resolve. Novel Alu insertions generate a phylogeny[84] with multiple inconsistencies that suggest either ILS or ancient hybridization among differentiated species. Chromosome painting analyses also indicate interspecies hybridization[85], and field studies document active hybrid zones[68].

Therefore, simple allopatric speciation models and associated ideas that posit the rapid origin of species boundaries do not generally hold for humans or other primates. The bonobo–chimpanzee speciation may be one notable exception[52], which is possibly related to a rapid shift in the Congo River that may have created a robust barrier to gene flow. Newer models of speciation address these complexities[86–89] and provide frameworks for future studies. Unresolved questions regarding primate genome evolution are: what are the demographic circumstances associated with extended periods of progressive genetic differentiation despite continuing genetic exchange? What types of genes are able to transfer between lineages, and what genes or genetic pathways are the first to develop marked differences between diverging lineages? Finally, what changes correlate with the end of gene flow between differentiating lineages?

Biomedical relevance

The two most commonly used non-human primates in biomedical research are the rhesus macaque and the cynomolgus or long-tailed macaque. Their importance as models for studies of human health and disease justifies extensive analyses of these genomes (see Supplementary information S1 (table)). These two species are members of the genus *Macaca*, which is a successful radiation of Old World monkeys that contains 18 extant species[72] and is distributed across Asia from Afghanistan to Japan and the Philippines (FIG. 2), with relict populations in Morocco. Other important primate model organisms are now also receiving attention. The genomes of the marmoset (*Callithrix jacchus*), sooty mangabey (*Cercocebus atys*), African green monkey (*Chlorocebus aethiops*) and olive baboon (*Papio anubis*) have been sequenced and assembled. Genome assemblies for mouse lemur (*Microcebus murinus*) and pigtailed macaque (*Macaca nemestrina*) are in progress. These species are all used as animal models in disease-related research, and whole-genome assemblies, transcriptomic data and other information are therefore valuable. For example, both sooty mangabeys

Strepsirrhine primates
Members of the branch of primates that includes lemurs, lorises, galagos and cheirogaleids, and that belongs to the suborder Strepsirrhini.

Hybrid zones
Geographical areas that are often, but not always, elongated and narrow in shape, where two distinct species occur together, mate and produce hybrid offspring that are fertile.

Parapatric
Pertaining to geographical distributions that are adjoining but that do not overlap extensively.

Introgression
The transfer of alleles or genes by hybridization and gene flow from one species to another.

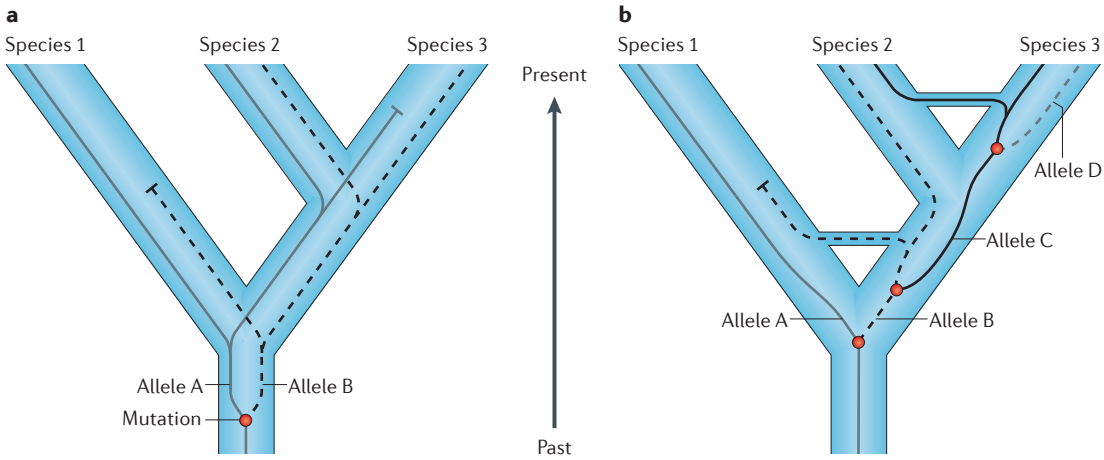


Figure 3 | **Incomplete lineage sorting and gene flow. a** | Incomplete lineage sorting can produce discrepancy between the phylogenetic tree for a specific gene or a genomic segment and the overall species-level phylogenetic tree. If an ancestral species is polymorphic (in this case, it is segregating Alleles A and B) and divides into two descendent lineages, then both alleles can be retained in the two daughter lineages. If one of these lineages divides again relatively soon, then all three species lineages may carry both alleles. Over time, each lineage will lose one or the other allele owing to genetic drift or selection. In this case, Species 1 retains Allele A and Species 3 retains Allele B. For this genomic segment, Species 2 will seem to be more closely related to either Species 1 or Species 3 depending on whether it retains Allele A or Allele B. Retention of Allele B would mean that this genomic segment matches the overall species-level phylogenetic tree, but retention of Allele A would lead to discrepancy. Analyses of whole-genome sequences for humans, chimpanzees and gorillas indicate that gene trees for a substantial portion of the genome do not match the overall species-level phylogeny, which places chimpanzees as closer relatives to humans than to gorillas[19]. **b** | Gene flow between diverging lineages is shown. Evidence from various primate species, especially for the ancestors of extant apes and humans, indicates that the process that produced the extant evolutionary lineages did not consist of rapid separation and immediate complete genetic isolation. Rather, recent analyses[52] suggest that in several cases, new evolutionary lineages can diverge and accumulate genetic differences despite maintaining some degree of genetic exchange (that is, gene flow). The situation depicted here shows three lineages arising from a common ancestor. In both cases of species-level divergence, there is gene flow (horizontal bars) between lineages. Allele B is transferred from one lineage to another through this process but is lost from the recipient lineage owing to either genetic drift or selection. Allele C is transferred between lineages and increases in frequency such that it is retained for an extended period, and it may therefore influence the evolutionary trajectory of the recipient lineage (Species 2). Rather than occurring as one discrete episode, the gene flow between diverging lineages probably decreases over time and eventually fades out.

and African green monkeys are important model species for Simian immunodeficiency virus (SIV) research because these animals are natural hosts that tolerate long-term infection with specific SIV viruses without developing disease[90]. Development of sequence data and related tools facilitates analyses of how these species tolerate SIV infection that is pathogenic in other primates.

Inter-species differences in disease-relevant variants. Comparisons of the cynomolgus macaque, rhesus macaque and human genomes are producing information that is directly relevant to specific biomedical questions. The genomes of rhesus and cynomolgus macaques are <1% different in single-copy sequences, but the two species carry specific differences in cytochrome P450 genes that are involved in drug metabolism[91]. Although most P450 genes are expressed at similar levels in humans, rhesus and cynomolgus, particular loci (for example, *CYP17A1*) are not. Knowledge of genomic differences should improve the interpretation of pharmacological studies using these species.

Humans and the two macaque species also show differences in other genetic components that are relevant to disease, such as pathways involving the melanocortin receptor, methyltransferases and the parathyroid hormone receptor 1 (REF. 92). Moreover, importantly, macaques have an expanded array of MHC class I genes that are central to their response to infectious agents and other immune system processes[16].

Several non-human primates carry sequences for protein-coding genes that are associated with increased risks of specific diseases in humans. Rhesus macaques carry variants in the ornithine carbamoyltransferase (*OTC*), phenylalanine hydroxylase (*PAH*) and *N*-acetylglucosaminidase alpha (*NAGLU*) genes that predispose some human individuals to disease (that is, OTC deficiency, which is a potentially severe disruption of the urea cycle; and phenylketonuria, which is a metabolic disorder that affects amino acid levels)[2]. Chimpanzees carry 'disease' alleles in genes that are related to cancer (mutL homologue 1 (*MLH1*)), diabetes mellitus (peroxisome proliferator-activated receptor gamma (*PPARG*)) and Alzheimer's disease (apolipoprotein E (*APOE*)) in

humans[1]. Gorillas have alleles at the granulin (*GRN*) gene and variants in titin-cap (*TCAP*) that are associated with dementia and hypertrophic cardiomyopathy in humans, respectively[5].

Polymorphism within species and disease phenotypes. Macaques and other primate species generally have higher levels of intra-species genetic variation than humans (see above). Thousands of nonsynonymous and splice-site variants have been identified in rhesus macaques[33,41], and such variability may influence the response of individual monkeys to experimental protocols. This naturally occurring genetic variation can be exploited to identify novel relationships between specific genes and disease-related phenotypes[93,94] or to study the phenotypic consequences of variation in genes that have already been implicated in human disease risk[95–97]. Variation in *OPRM1*, which encodes the mu-opioid receptor, illustrates the parallels in monkeys and humans, as naturally occurring nonsynonymous variation in rhesus macaque *OPRM1* influences both the behavioural response of animals to alcohol consumption and the pharmacogenetic response to treatment, which is similar to the effects of nonsynonymous variation in the same gene in humans[96]. Large-scale DNA resequencing of macaques, baboons, African green monkeys, marmosets and other laboratory primates will undoubtedly identify many functionally important genetic variants that will be useful for investigating genetic mechanisms of disease in experimentally controlled primate models[33,41].

Transcriptomics. Analyses of gene expression in primate models of disease will be fundamental to future studies; for example, primates are crucial for the development and testing of new drugs. Expression of drug-metabolizing P450 genes and some amino acid sequences differ between cynomolgus and rhesus macaques[98], which has implications for pharmacokinetics. Using linkage analyses and quantified differences in gene expression among pedigreed African green monkeys, researchers have mapped expression quantitative trait loci (eQTLs)[99]. Analyses of rhesus macaques show that gene expression in the immune system is sensitive to differences in social dominance rank — a fundamental aspect of macaque behaviour[100] — which indicates that common social interactions can influence gene function that is related to immunology and disease. Primate microRNAs (miRNAs) will be crucial for understanding disease models[101] and evolutionary adaptation[102,103]. Species differences in miRNA expression may affect expression of transcription factors[104], which has important consequences for several pathways.

Future directions

The initial draft assemblies for non-human primates all provide much useful information but are not complete or reliable enough to support all current scientific goals[21,22]. One limitation of draft genomes is the presence of gaps in chromosomal sequences, which results in missing exons or genes. Although growing use of RNA sequencing to identify transcribed genes is improving the completeness and annotation of non-human genome assemblies, the available assemblies still contain gaps. For example, the recent assembly of the gorilla genome incorporates 2.8 Gb of sequence into contigs[5]. However, when the investigators aligned the human, chimpanzee, gorilla and orang-utan genome assemblies in order to carry out whole-genome analyses of sequence differences, they were only able to produce a four-way 'great ape plus human' alignment that included 2.0 Gb[5] owing partly to gaps and other problems among the ape assemblies. Another crucial issue that affects the ability to carry out comprehensive analyses is problems with identifying and properly assembling segmental duplications and gene copy-number differences among species[21]. Misassemblies are also a recurrent problem among draft assemblies produced using only next-generation short-read technologies[22].

Improved assemblies with longer contigs and more complete coverage in high-quality sequence data (that is, comprehensive delineation of segmental duplications, and fewer genes with gaps and errors) are needed. Deeper sequence coverage will improve some assemblies, but new technologies that provide longer reads will yield better assemblies by filling remaining gaps. The Pacific Biosciences RS II platform is one plausible option for upgrading the quality of primate genomes[105].

In addition, annotation of functional elements can improve with contiguity and quality of the reference genome, as well as with access to transcript data. A high priority is to identify and validate transcripts for both protein-coding and other transcribed sequences. Long non-coding RNAs, miRNAs and other genome features are currently poorly annotated for most non-human primates. Experimental study of those genomic elements in primate model systems is likely to produce substantial dividends for both biomedical and evolutionary studies.

With the sequencing technologies now available, researchers are able to generate large amounts of DNA and RNA sequence data rapidly. This is creating an increasing need for software tools to process comparative data and to speed up interpretation. The natural emphasis among researchers in human genetics has been the development of computational tools that are specifically designed to analyse human genomes, some of which are not easily applicable to non-human species. However, some new tools are readily useful in analyses of non-human primates[106,107], and several online databases are collecting, organizing and synthesizing comparative genomic data. These include the Great Ape Genome Project, RhesusBase, the UCSC genome browser and Ensembl genome browser. However, the speed with which comparative data are being generated creates an ever-growing need for additional computational tools that are designed to meet the needs of comparative analyses.

Most effort in primate sequencing so far has been directed towards the great apes; this is natural given their phylogenetic relationships to humans. The sequencing of species from other branches of the primate evolutionary tree, in particular the New World monkeys and

Catarrhine primates
Members of the primate evolutionary lineage that includes Old World monkeys (superfamily Cercopithecoidea) or hominoids (superfamily Hominoidea). The catarrhines include all extant apes, anthropoid monkeys native to Asia and Africa, and humans.

strepsirrhine primates, will provide increased power to identify conserved genomic segments that are unique to primates or to subsets of primates (for example, catarrhine primates). Each newly sequenced species adds evolutionary perspective and generates new potential models of human genetic disorders.

Little is known about genetic variation in most primate species, although they generally show as much variation as, or more variation than, humans. Resequencing in commonly used laboratory primates will discover new variants of interest for biomedical research. Furthermore, there is substantial opportunity to use this naturally occurring functional variation to explore gene–gene or gene–environment interactions[108,109].

Finally, non-human primates can facilitate investigation of epigenetic control of genome function. Experimental manipulation of environmental factors that influence the human epigenome will be feasible in better characterized primate genomes. Detailed analyses and manipulation of the primate microbiome may also have a substantial effect.

Conclusions

Comparative primate genomics is in a phase of rapid growth, as information about transcriptomes, intra-species polymorphism and other aspects of genomics is being generated at a fast pace. The major impact so far has been to provide novel information on the history and mechanisms of human genome evolution, including evidence for a complex history of genetic divergence and exchange among ancestral evolutionary lineages (FIG. 3). Non-human primate genomics is also expanding the scope of biomedical research with innovative analyses of primate models of human disease. Despite recent progress, both evolutionary and biomedical studies would substantially benefit from additional information. There is real opportunity to examine the continuum from microevolutionary processes that control intra-species variation (for example, positive and negative fitness effects of segregating polymorphisms within species) to macroevolutionary processes that affect inter-species differences.

Both from the perspective of understanding the origin of humans and from that of elucidating the genetic basis of human disease, non-human primates are indispensable resources for comparative and experimental study. Genomics is now central to all of biology, and it is therefore both sensible and timely that comparative primate genomics is receiving increased attention. Analyses so far have provided valuable and sometimes unexpected results. There will be many further advances, including a few more surprises, and ultimately a much richer understanding of genome structure, function and dynamics, as investigators with a wide range of interests continue to generate new information on the genomes of non-human primates.

1. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
2. Gibbs, R. A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
3. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
4. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
5. Scally, A. *et al.* Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169–175 (2012).
6. Langergraber, K. E. *et al.* Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl Acad. Sci. USA* **109**, 15716–15721 (2012).
7. Steiper, M. E. & Seiffert, E. R. Evidence for a convergent slowdown in primate molecular rates and its implications for the timing of early primate evolution. *Proc. Natl Acad. Sci. USA* **109**, 6006–6011 (2012).
8. Li, W. H. & Tanimura, M. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* **326**, 93–96 (1987).
9. Perry, G. H. *et al.* A genome sequence resource for the aye-aye (*Daubentonia madagascariensis*), a nocturnal lemur from Madagascar. *Genome Biol. Evol.* **4**, 126–135 (2012).
10. Britten, R. J., Rowen, L., Williams, J. & Cameron, R. A. Majority of divergence between closely related DNA samples is due to indels. *Proc. Natl Acad. Sci. USA* **100**, 4661–4665 (2003).
11. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
This paper presents an exceptional example of the power of comparative genomics to identify novel conserved genomic regions that evolve slowly across species as a result of shared functional importance.
12. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
13. Locke, D. P. *et al.* Comparative and demographic analysis of orang-utan genomes. *Nature* **469**, 529–533 (2011).
14. Gokcumen, O. *et al.* Primate genome architecture influences structural variation mechanisms and functional consequences. *Proc. Natl Acad. Sci. USA* **110**, 15764–15769 (2013).
15. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Rev. Genet.* **10**, 691–703 (2009).
16. Daza-Vamenta, R., Glusman, G., Rowen, L., Guthrie, B. & Geraghty, D. E. Genetic divergence of the rhesus macaque major histocompatibility complex. *Genome Res.* **14**, 1501–1515 (2004).
17. Nowick, K. *et al.* Gain, loss and divergence in primate zinc-finger genes: a rich resource for evolution of gene regulatory differences between species. *PLoS ONE* **6**, e21553 (2011).
18. Popesco, M. C. *et al.* Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* **313**, 1304–1307 (2006).
19. Dumas, L. J. *et al.* DUF1220-domain copy number implicated in human brain-size pathology and evolution. *Am. J. Hum. Genet.* **91**, 444–454 (2012).
20. O'Bleness, M. S. *et al.* Evolutionary history and genome organization of DUF1220 protein domains. *G3 (Bethesda)* **2**, 977–986 (2012).
21. Alkan, C., Sajjadian, S. & Eichler, E. E. Limitations of next-generation genome sequence assembly. *Nature Methods* **8**, 61–65 (2011).
22. Zhang, X., Goodsell, J. & Norgren, R. B. Jr. Limitations of the rhesus macaque draft genome assembly and annotation. *BMC Genomics* **13**, 206 (2012).
23. Stankiewicz, P. & Lupski, J. R. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* **61**, 437–455 (2010).
24. Marques-Bonet, T., Ryder, O. A. & Eichler, E. E. Sequencing primate genomes: what have we learned? *Annu. Rev. Genom. Hum. Genet.* **10**, 355–386 (2009).
25. Marques-Bonet, T. *et al.* A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**, 877–881 (2009).
26. Jiang, Z. *et al.* Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. *Nature Genet.* **39**, 1361–1368 (2007).
27. Dumas, L. *et al.* Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res.* **17**, 1266–1277 (2007).
28. Gazave, E. *et al.* Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. *Genome Res.* **21**, 1626–1639 (2011).
29. Johnson, M. E. *et al.* Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**, 514–519 (2001).
This is one of the first studies to identify changes in gene copy number among humans and the great apes that seem to be driven by positive selection, with evidence for adaptive changes in both copy number and nucleotide sequences.
30. Lorente-Galdos, B. *et al.* Accelerated exon evolution within primate segmental duplications. *Genome Biol.* **14**, R9 (2013).
31. Fortna, A. *et al.* Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol.* **2**, e207 (2004).
32. Jensen-Seaman, M. I., Deinard, A. S. & Kidd, K. K. Modern African ape populations as genetic and demographic models of the last common ancestor of humans, chimpanzees, and gorillas. *J. Hered.* **92**, 475–480 (2001).
33. Fawcett, G. L. *et al.* Characterization of single-nucleotide variation in Indian-origin rhesus macaques (*Macaca mulatta*). *BMC Genomics* **12**, 311 (2011).
34. Hernandez, R. D. *et al.* Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. *Science* **316**, 240–243 (2007).
35. Smith, D. G., McDonough, J. W. & George, D. A. Mitochondrial DNA variation within and among regional populations of longtail macaques (*Macaca fascicularis*) in relation to other species of the *fascicularis* group of macaques. *Am. J. Primatol.* **69**, 182–198 (2007).
36. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nature* **499**, 471–475 (2013).
This paper presents a substantial amount of genomic information concerning species and subspecies of great apes, which provides important new insights into the evolution of these lineages.

37. Ferguson, B. *et al.* Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). *BMC Genomics* **8**, 43 (2007).
38. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
39. Perry, G. H. *et al.* Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res.* **22**, 602–610 (2012).
This paper reports the first genome-scale analysis of several threatened or endangered primates; it documents unexpected patterns of intra-species variability and ancient selection on protein-coding genes.
40. Vallender, E. J. Expanding whole exome resequencing into non-human primates. *Genome Biol.* **12**, R87 (2011).
41. Yuan, Q. *et al.* The rhesus macaque is three times as diverse but more closely equivalent in damaging coding variation as compared to the human. *BMC Genet.* **13**, 52 (2012).
42. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
In this classic and prescient paper that was written long before researchers had access to substantial amounts of DNA sequence data, the authors used information about protein sequence differences and dissociation temperatures for hybrid human–chimpanzee DNA molecules to correctly infer that much of the anatomical and physiological difference between humans and chimpanzees is due to changes in gene regulation rather than to changes in protein sequence.
43. Wilson, A. C., Maxson, L. R. & Sarich, V. M. Two types of molecular evolution. Evidence from studies of interspecific hybridization. *Proc. Natl Acad. Sci. USA* **71**, 2843–2847 (1974).
44. Arbiza, L. *et al.* Genome-wide inference of natural selection on human transcription factor binding sites. *Nature Genet.* **45**, 723–729 (2013).
This paper describes an innovative analysis of evolutionary changes in transcription factor binding sites, which found that natural selection has exerted marked effects of these regulatory sequences during recent human evolution.
45. Pipes, L. *et al.* The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics. *Nucleic Acids Res.* **41**, D906–D914 (2013).
46. Blekhman, R., Oshlack, A., Chabot, A. E., Smyth, G. K. & Gilad, Y. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genet.* **4**, e1000271 (2008).
47. Blekhman, R., Marioni, J. C., Zumbo, P., Stephens, M. & Gilad, Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* **20**, 180–189 (2010).
This report describes notable data concerning differences in gene splicing — a potentially important mechanism for rapid evolutionary change — among non-human primates and humans.
48. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
49. Calarco, J. A. *et al.* Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev.* **21**, 2963–2975 (2007).
50. Zeng, J. *et al.* Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am. J. Hum. Genet.* **91**, 455–465 (2012).
51. Babbitt, C. C., Tung, J., Wray, G. A. & Alberts, S. C. Changes in gene expression associated with reproductive maturation in wild female baboons. *Genome Biol. Evol.* **4**, 102–109 (2012).
52. Mailund, T., A. E. Halager, Scally, A. A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. *PLoS Genet.* **8**, e1003125 (2012).
This paper reports a highly informative analysis of genomic differentiation among ancestral populations of hominoids, including analysis of the process of divergence that produced the evolutionary separation of the ancestors of humans, chimpanzees and gorillas.
53. Reich, S. P. L. P. The date of interbreeding between Neanderthals and modern humans. *PLoS Genet.* **8**, e1002947 (2012).
54. Sankararaman, S., Patterson, N., Li, H., Paabo, S. & Reich, D. The date of interbreeding between Neandertals and modern humans. *PLoS Genet.* **8**, e1002947 (2012).
55. Auton, A. *et al.* A fine-scale chimpanzee genetic map from population sequencing. *Science* **336**, 193–198 (2012).
56. George, R. D. *et al. Trans* genomic capture and sequencing of primate exomes reveals new targets of positive selection. *Genome Res.* **21**, 1686–1694 (2011).
57. Harris, R. A. *et al.* Evolutionary genetics and implications of small size and twinning in callitrichine primates. *Proc. Natl Acad. Sci. USA* **111**, 1467–1472 (2014).
This report describes genetic evolution in an unusual group of non-human primates that show unique adaptations for reproduction and identifies specific sequence changes that may contribute to those adaptations.
58. O'Bleness, M., Searles, V. B., Varki, A., Gagneux, P. & Sikela, J. M. Evolution of genetic and genomic features unique to the human lineage. *Nature Rev. Genet.* **13**, 853–866 (2012).
59. McLean, C. Y. *et al.* Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–219 (2011).
60. Prabhakar, S. *et al.* Human-specific gain of function in a developmental enhancer. *Science* **321**, 1346–1350 (2008).
61. Charrier, C. *et al.* Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. *Cell* **149**, 923–935 (2012).
62. Pamilo, P. & Nei, M. Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5**, 568–583 (1988).
63. Rogers, J. Levels of the genealogical hierarchy and the problem of homonoid phylogeny. *Amer. J. Phys. Anthropol.* **94**, 81–88 (1994).
64. Ruvolo, M. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Mol. Biol. Evol.* **14**, 248–265 (1997).
65. Prufer, K. *et al.* The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**, 527–531 (2012).
66. Wall, J. D. *et al.* Incomplete lineage sorting is common in extant gibbon genera. *PLoS ONE* **8**, e53682 (2013).
67. Zinner, D., Arnold, M. L. & Roos, C. The strange blood: natural hybridization in primates. *Evol. Anthropol.* **20**, 96–103 (2011).
68. Detwiler, K. M., Burrell, A. S. & Jolly, C. J. Conservation implications of hybridization in African cercopithecine monkeys. *Int. J. Primatol.* **26**, 661–684 (2005).
69. Jolly, C. J. A proper study for mankind: analogies from the Papionin monkeys and their implications for human evolution. *Am. J. Phys. Anthropol. Suppl.* **33**, 177–204 (2001).
This wide-ranging review discusses issues related to the complexity of reconstructing ancient evolutionary processes and the value of information on the population biology of and hybridization among non-human primate species for understanding aspects of human evolution.
70. Jolly, C. J. in *Species, Species Concepts and Primate Evolution* (eds Kimbel, W. H. & Martin, L. B.) 67–107 (Plenum Press, 1993).
71. Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L. F. & Roos, C. Baboon phylogeny as inferred from complete mitochondrial genomes. *Am. J. Phys. Anthropol.* **150**, 133–140 (2013).
72. Groves, C. *Primate Taxonomy* (Smithsonian Institution Press, 2001).
73. Jolly, C. J., Burrell, A. S., Phillips-Conroy, J. E., Bergey, C. & Rogers, J. Kinda baboons (*Papio kindae*) and grayfoot chacma baboons (*P. ursinus griseipes*) hybridize in the Kafue river valley, Zambia. *Am. J. Primatol.* **73**, 291–303 (2011).
74. Alberts, S. C. & Altmann, J. Immigration and hybridization patterns of yellow and anubis baboons in and around Amboseli, Kenya. *Am. J. Primatol.* **53**, 139–154 (2001).
75. Bergman, T. J., Phillips-Conroy, J. E. & Jolly, C. J. Behavioral variation and reproductive success of male baboons (*Papio anubis*×*Papio hamadryas*) in a hybrid social group. *Am. J. Primatol.* **70**, 136–147 (2008).
76. Ackermann, R. R., Rogers, J. & Cheverud, J. M. Identifying the morphological signatures of hybridization in primate and human evolution. *J. Hum. Evol.* **51**, 632–645 (2006).
77. Burrell, A. S. *Phylogenetics and population genetics of central African baboons*. Thesis, New York Univ. (2009).
78. Charpentier, M. J. *et al.* Genetic structure in a dynamic baboon hybrid zone corroborates behavioural observations in a hybrid population. *Mol. Ecol.* **21**, 715–731 (2012).
79. Tosi, A. J., Morales, J. C. & Melnick, D. J. Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution* **57**, 1419–1435 (2003).
80. Stevison, L. S. & Kohn, M. H. Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. *Mol. Ecol.* **18**, 2457–2475 (2009).
81. Burrell, A. S., Jolly, C. J., Tosi, A. J. & Disotell, T. R. Mitochondrial evidence for the hybrid origin of the kipunji, *Rungwecebus kipunji* (Primates: Papionini). *Mol. Phylogenet Evol.* **51**, 340–348 (2009).
82. Jones, T. *et al.* The highland mangabey *Lophocebus kipunji*: a new species of African monkey. *Science* **308**, 1161–1164 (2005).
83. Jaffe, K. E. & Isbell, L. A. in *Primates in Perspective* 2nd edn Ch. 16 (eds Campbell, C. J., Fuentes, A., MacKinnon, K. C., Bearder, S. K. & Stumpf, R. M.) (Oxford Univ. Press, 2011).
84. Xing, J. *et al.* A mobile element-based evolutionary history of guenons (tribe Cercopithecini). *BMC Biol.* **5**, 5 (2007).
85. Moulin, S., Gerbault-Seureau, M., Dutrillaux, B. & Richard, F. A. Phylogenomics of African guenons. *Chromosome Res.* **16**, 783–799 (2008).
86. Gombert, Z., Parchman, T. L. & Buerkle, C. A. Genomics of isolation in hybrids. *Phil. Trans. R. Soc. B* **367**, 439–450 (2012).
87. Network, M. C. S. What do we need to know about speciation? *Trends Ecol. Evol.* **27**, 27–39 (2012).
88. Nosil, P. & Feder, J. L. Genomic divergence during speciation: causes and consequences. *Phil. Trans. R. Soc. B* **367**, 332–342 (2012).
89. Abbott, R. *et al.* Hybridization and speciation. *J. Evol. Biol.* **26**, 229–246 (2013).
90. Chahroudi, A., Bosinger, S. E., Vanderford, T. H., Paiardini, M. & Silvestri, G. Natural SIV hosts: showing AIDS the door. *Science* **335**, 1188–1193 (2012).
91. Ebeling, M. *et al.* Genome-based analysis of the nonhuman primate *Macaca fascicularis* as a model for drug safety assessment. *Genome Res.* **21**, 1746–1756 (2011).
92. Yan, G. *et al.* Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nature Biotech.* **29**, 1019–1023 (2011).
93. Cox, L. A. *et al.* Identification of promoter variants in baboon endothelial lipase that regulate high-density lipoprotein cholesterol levels. *Circulation* **116**, 1185–1195 (2007).
94. Rogers, J. *et al. CRHR1* genotypes, neural circuits and the diathesis for anxiety and depression. *Mol. Psychiatry* **18**, 700–707 (2013).
95. Francis, P. J. *et al.* Rhesus monkeys and humans share common susceptibility genes for age-related macular disease. *Hum. Mol. Genet.* **17**, 2673–2680 (2008).
96. Vallender, E. J., Ruedi-Bettschen, D., Miller, G. M. & Platt, D. M. A pharmacogenetic model of naltrexone-induced attenuation of alcohol consumption in rhesus monkeys. *Drug Alcohol Depend* **109**, 252–256 (2010).
97. Barr, C. S. *et al.* Rearing condition and rh5-HTTLPR interact to influence limbic-hypothalamic-pituitary-adrenal axis response to stress in infant macaques. *Biol. Psychiatry* **55**, 733–738 (2004).
98. Ise, R. *et al.* Expression profile of hepatic genes in cynomolgus macaques bred in Cambodia, China, and Indonesia: implications for cytochrome P450 genes. *Drug Metab. Pharmacokinet.* **27**, 307–316 (2012).
99. Jasinska, A. J. *et al.* A non-human primate system for large-scale genetic studies of complex traits. *Hum. Mol. Genet.* **21**, 3307–3316 (2012).
100. Tung, J. *et al.* Social environment is associated with gene regulatory variation in the rhesus macaque immune system. *Proc. Natl Acad. Sci. USA* **109**, 6490–6495 (2012).
101. Karere, G. M., Glenn, J. P., VandeBerg, J. L. & Cox, L. A. Differential microRNA response to a high-cholesterol, high-fat diet in livers of low and high LDL-C baboons. *BMC Genomics* **13**, 320 (2012).

102. Hu, H. Y. *et al.* MicroRNA expression and regulation in human, chimpanzee, and macaque brains. *PLoS Genet.* **7**, e1002327 (2011).
103. Somel, M. *et al.* MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol.* **9**, e1001214 (2011).
104. Dannemann, M. *et al.* Transcription factors are targeted by differentially expressed miRNAs in primates. *Genome Biol. Evol.* **4**, 552–564 (2012).
105. English, A. C. *et al.* Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE* **7**, e47768 (2012).
106. Wang, Y., Lu, J., Yu, J., Gibbs, R. A. & Yu, F. An integrative variant analysis pipeline for accurate genotype/haplotype inference in population NGS data. *Genome Res.* **23**, 833–842 (2013).
107. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
108. Barr, C. S. *et al.* The utility of the non-human primate; model for studying gene by environment interactions in behavioral research. *Genes Brain Behav.* **2**, 336–340 (2003).
109. Barr, C. S. *et al.* Interaction between serotonin transporter gene variation and rearing condition in alcohol preference and consumption in female primates. *Arch. Gen. Psychiatry* **61**, 1146–1152 (2004).
110. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
111. Mayr, E. *Systematics and the Origin of Species* (Columbia Univ. Press, 1942).
112. Gould, S. J. & Eldridge, N. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* **3**, 115–151 (1977).
113. Pagel, M., Venditti, C. & Meade, A. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* **314**, 119–121 (2006).
114. Prüfer, K. *et al.* The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**, 527–531 (2012).
115. Higashino, A. *et al.* Whole-genome sequencing and analysis of the Malaysian cynomolgus macaque (*Macaca fascicularis*) genome. *Genome Biol.* **13**, R58 (2012).
116. Pan_troglodytes-2.1.4 assembly. *National Center for Biotechnology Information* [online], www.ncbi.nlm.nih.gov/assembly/GCF_000001515.5 (2011).
117. OtoGar3 assembly. *National Center for Biotechnology Information* [online], www.ncbi.nlm.nih.gov/assembly/396188 (2011).
118. Tarsius_syrichta-2.0.1 assembly. *National Center for Biotechnology Information* [online], www.ncbi.nlm.nih.gov/assembly/64501 (2013).
119. Fang, X. *et al.* Genome sequence and global sequence variation map with 5.5 million SNPs in Chinese rhesus macaque. *Genome Biol.* **12**, R63 (2011).

Acknowledgements
The authors thank K. Worley, S. Richards, M. Raveendran, G. Fawcett, D. Rio Deiros and F. Yu for discussion, and three anonymous reviewers for their comments. This work was supported by the US National Institutes of Health grants U54-HG006484 and R24-OD011173.

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION

Broad Institute Genome Center: http://www.broadinstitute.org
Ensembl genome browser: http://www.ensembl.org
Genome Institute at Washington University: http://genome.wustl.edu/
Great Ape Genome Project: http://biologiaevolutiva.org/greatape
Human Genome Sequencing Center: www.hgsc.bcm.edu
Red List of Endangered Species: www.iucnredlist.org
RhesusBase: http://www.rhesusbase.org
UCSC genome browser: http://genome.ucsc.edu/
US National Institutes of Health Blueprint Non-Human Primate Atlas: http://www.blueprintnhpatlas.org

SUPPLEMENTARY INFORMATION

See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF