Annu. Rev. Genomics Hum. Genet. 2004. 5:351–78
doi: 10.1146/annurev.genom.5.061903.180040

First published online as a Review in Advance on May 26, 2004

COMPARATIVE PRIMATE GENOMICS

Wolfgang Enard and Svante Pääbo
Max-Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany; email: enard@eva.mpg.de

Key Words chimpanzee, apes, gene expression, molecular evolution, selection

■ **Abstract** With the completion of the human genome sequence and the advent of technologies to study functional aspects of genomes, molecular comparisons between humans and other primates have gained momentum. The comparison of the human genome to the genomes of species closely related to humans allows the identification of genomic features that set primates apart from other mammals and of features that set certain primates—notably humans—apart from other primates. In this article, we review recent progress in these areas with an emphasis on how comparative approaches may be used to identify functionally relevant features unique to the human genome.

INTRODUCTION

Among the ∼300 primate species that currently live on Earth, humans are clearly the most dominant. They achieve their dominance by an unparalleled capacity for culture, which has allowed them to grow in number, to extend their range to almost all regions of the planet, and to impact—for better and for worse—the lives of many other animals, plants, and ecosystems. This begs the question of what features caused such a huge difference in the evolutionary trajectory of humans compared with other closely related primates.

Another aspect of human uniqueness among primates is that they contemplate questions such as the one posed above. An important aspect of this human enquiry into human nature is the exploration of the human genome, which culminated in the finished human genome sequence presented in the spring of 2003. Thus, humans as a species now have the unique challenge of understanding the functions and origins of their own genome.

One of the most important ways to meet this challenge is to compare the human genome with the genomes of other species. Hitherto, the limited availability of complete genome sequences confined such comparisons to species such as mice, fish, flies, and worms, which are all distantly related to humans and allow genomic features to be identified due to their conservation over long times. Such features are relevant to understanding what is common to all mammals, all vertebrates, or all animals, but uninformative with regard to what is unique to primates. It is especially uninformative with respect to what most interests the congenitally

narcissistic human species, i.e., "what sets us apart from other species?" This situation is changing with the imminent completion of a draft sequence of the chimpanzee genome, which will allow us to list most genomic features by which we differ from our closest relative among primates. Thus, we are at the wake of a new era that will allow us to explore the comparative genomics of primates. Below we summarize some evolutionary aspects of our current knowledge of primate genomes, and we are painfully aware that this review will likely become obsolete faster than most reviews in the rapidly changing field of genomics. However, perhaps one day, after its short period of possible relevance, it may be of interest to historians as a document of our monumental ignorance at this point in time.

PHYLOGENIES

To understand how the structural and functional features of genomes evolve, it is of paramount importance to know how the species carrying the genomes are related to each other, i.e., their phylogeny.

Primates are generally classified into groups for which there is evidence that their ancestry traces back to a single common ancestor. The six major goups are: lemurs that live on Madagascar; loriformes, such as the galagos, that live in Africa and Asia; tarsiers that dwell on the islands of southeastern Asia; New World monkeys; Old World monkeys; and hominoids, i.e., apes and humans (38). The apes include the gibbons and orangutans in Asia, and the gorillas, chimpanzees, and bonobos in Africa. The phylogeny of these groups is generally accepted (Figure 1). Only the tarsiers are controversial because of the support for a closer relationship between them and the common ancestor of Old and New World monkeys and apes has been found (49, 103), as well as support for a common ancestry of tarsiers and the lemurs (93). Because three Alu elements were found at the same sites in the genomes of tarsiers, New World monkeys, and Old World monkeys, a common ancestry of these groups seems more likely (116). From a human perspective, our closest relatives are the apes, followed by the progressively more distant relatives: the Old Word monkeys, the New World monkeys, the tarsiers, and the lemurs.

Among the apes, the African apes are more closely related to humans than the gibbons and orangutans. For example, in 1863, Thomas Huxley (59) stated that "it is quite certain that the Ape which most nearly approaches Man, in the totality of its organization, is either the Chimpanzee or the Gorilla." However, whether chimpanzees or gorillas are closer to humans, or whether the human lineage split off before the gorilla and chimpanzee lineages separated, remained unclear for a long time. Because DNA sequences from different parts of the genomes of humans chimpanzees and gorillas are now available, one can address this question from a genomic perspective [for an introduction to phylogenetic analyses, see, e.g., (96)]. For example, out of 53 short intergenic, ~500-bp long DNA sequences, 31 support the human-chimpanzee group, 10 support the human-gorilla group, and 12 support the chimpanzee-gorilla group (15). In most cases, the analyses of longer DNA sequences indicate that chimpanzees and humans share a more recent ancestor,

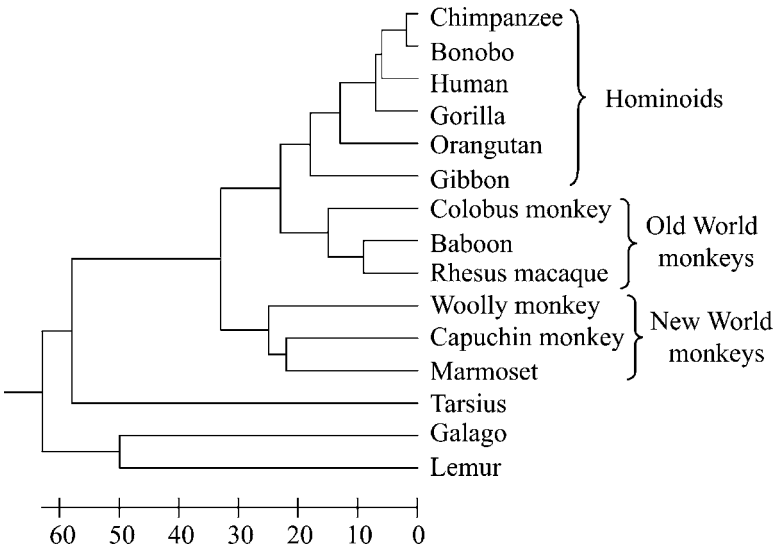

Figure 1 Phylogeny of a few primate species according to Goodman (48). Estimates of divergence times are from Goodman (48), except for the divergence times between humans and the great apes, the Old World monkeys, and the New World monkeys (46) and the chimpanzee-bonobo split (140). Note that these are divergence times of DNA sequences and that species divergences can be considerably smaller (see Figure 3).

but a few indicate other phylogenies (e.g., 113). The reason that most regions of the genome indicate a human-chimpanzee relationship, whereas other regions indicate the other two possible phylogenies, is probably that the gorilla lineage diverged shortly before the lineages of humans and chimpanzees diverged. Therefore, polymorphisms in the ancestral population of all three species could persist from the first divergence to the second divergence. The random fixation of these alleles in the three lineages led to a situation where different nucleotide positions in the three species are related by different phylogenies (Figure 2). The smaller the genetic distance between two nucleotide positions, the higher the probability that they will share the same phylogeny. Thus, single genes tend to support one phylogeny and the human genome as a whole can be seen as a "mosaic," in that different regions are differently related to chimpanzees and gorillas.

The close phylogenetic relationship between humans and both chimpanzees and gorillas means that the true phylogeny can not be known for any particular genomic region if its phylogeny has not been analyzed in the context of at least one species in addition to humans, chimpanzees, and gorillas. When comparing the human and chimpanzee genomes, the ancestral state cannot be reliably inferred using the gorilla as an outgroup. A species more distantly related to humans and the African apes needs to be used. The closest possibility is the orangutan, whose lineage branched off so long ago that any region not affected by special phenomena

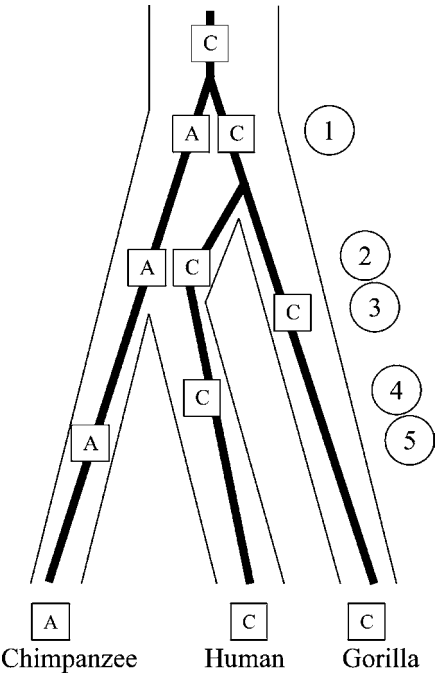

Figure 2 Illustration of how ancestral polymorphisms can lead to a gene tree that is different from the species tree. The genealogy of a neutral polymorphism is depicted. The ancestral allele C mutates to A, and the A/C polymorphism segregates in the ancestral population of humans, chimpanzees, and gorillas (1) and is still segregating in the ancestral population of humans and chimpanzees (2). Whereas the C-allele becomes fixed on the gorilla lineage (3) and the human lineage (4), the A-allele gets fixed on the lineage leading to the chimpanzee (5). Therefore, the common ancestor of the site is more recent for humans and gorillas, and the gene tree is different from the species tree.

such as balancing selection is expected to represent a true outgroup to humans and chimpanzees

Divergence Times

The estimation of divergence times between primate groups is more controversial than the phylogeny, largely due to the uncertainties associated with fossil calibration points. For example, the divergence of the lineage leading to lemurs and the lineage to all other primates has been dated to 63 mya (49) or to 80 mya (118). Using two different fossil calibration points and testing several statistical methods, Glazko & Nei (46) estimated the divergence times between humans and other primates. According to them, the human lineage diverged from that of the chimpanzee approximately 5–7 mya, that of the gorilla 6–8 mya, that of the

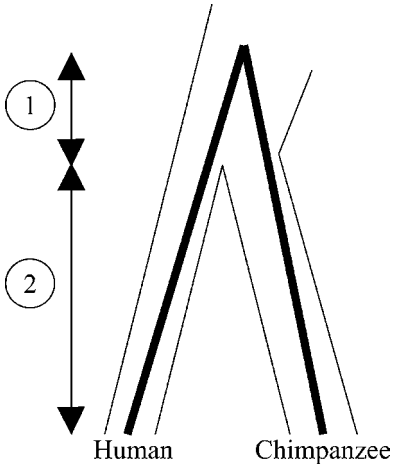

Figure 3 The divergence time of two orthologous DNA sequences consists of the time when the sequences segregated in the ancestral population (1) and the time after the species split (2). Thus, divergence times of DNA sequences always predate the species split. This can constitute a substantial fraction of the divergence time when the species split is recent and the ancestral population is size large.

orangutan 12–15 mya, that of Old World monkeys 21–25 mya, and that of New World monkeys 32–36 mya (Figure 1). These dates are in accord with previous studies (e.g., 15, 48), even if they are disputed by some (2). The divergence times of DNA sequences from bonobos and chimpanzees are less well dated and range from 0.9 million years (65) to 2.5 million years (43), with the largest dataset collected so far suggesting 1.8 million years (141).

One issue often not considered when estimating the divergence times of closely related species such as humans and the great apes is that the divergence of a DNA sequence is composed of the time after the species split as well as the time before that when the two sequences segregated in the ancestral population (Figure 3). If the split time is recent, and the ancestral population size is large, the divergence times of DNA sequences is substantially larger than the divergence of the species (109). For example, the ancestral effective population size of humans and chimpanzees is estimated to have been four to seven times larger than the current effective human population size of ∼10,000 (126). Considering this, the time of the species divergence of humans and chimpanzees has been estimated to 2.9–4.3 mya (126). Similarly, an estimate of the population divergence time of bonobos and chimpanzees suggests a divergence time of 0.8 million years, which is not much different from the divergence estimates of 0.4–0.65 million years between the populations of chimpanzees in central Africa and western Africa (37). Note that these estimates are based on a model in which populations are immediately isolated. If gene flow occurs between populations for some time, this would overestimate the size of the ancestral population and underestimate the divergence time (126).

Molecular information is available about one extinct primate group's phylogenetic relationship and divergence from humans. The Neandertals lived in Europe and western Asia until about 30,000 years ago, when they disappear from the fossil record. Thus, they are the group most closely related to contemporary humans. Determining the mitochondrial DNA sequences from four Neandertals (73, 74, 102, 117) has shown that they carried mitochondrial DNA sequences that fall outside the variation of modern humans and that diverged a little more than 500,000 years ago. This fits well with archaeological and paleontological estimates of a divergence between anatomically modern humans and Neandertals more than 350,000 years ago (8). However, just because Neandertals did not contribute mitochondrial DNA to the contemporary human gene pool does not preclude that they might have contributed other genes. A recent study of several Neandertals and early modern human remains concluded that the Neandertal contribution to the early modern gene pool is limited to 0% to 25% (117a). Thus, although there is no positive genetic evidence for any interbreeding between humans and their closest extinct relatives, it can not be excluded that it took place to some extent.

Intraspecific Diversity

There are vast amounts of data on intraspecific genetic diversity of humans. By comparison, few studies have been devoted to the genetic diversity of chimpanzees (e.g., 20, 37, 43, 45, 65, 70, 120, 141), and even fewer to gorillas, orangutans (20, 43, 66), and other primates.

What is clear is that humans differ from chimpanzees, gorillas, and orangutans by being less diverse (for reviews, see 63, 64). Only bonobos may carry about as much, or a little more, diversity than humans (21, 65, 141). Thus, humans differ from almost all the apes by having low levels of intraspecific diversity (Figure 4). This is surprising because humans are orders of magnitude more numerous than the

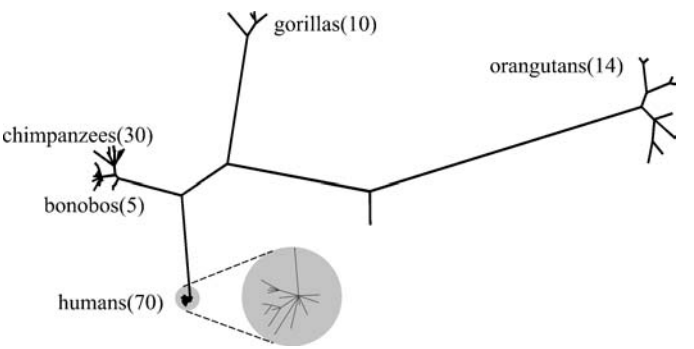

Figure 4 Gene tree of a X-chromosomal region (66). The number of chromosomes analyzed is given in parantheses. Figure is modified from Kaessmann et al. (66).

great apes. It is also surprising given the fact that humans are distributed over the entire world whereas apes are more restricted in their distribution. One explanation is that humans go back to a small population that expanded in the relatively recent past (53). However, other factors such as mating systems and population structure might also play a role.

An observation compatible with a recent human expansion from a small population is that humans have little genetic differences between populations whereas ape populations differ a lot. For example, the difference between chimpanzees from central Africa and chimpanzees from West Africa is three to four times greater than the differences among human continental populations (37). There is a similar trend—although supported by less data—in gorillas and orangutans (64).

Splitting or Lumping Primates?

In conjunction with the extent of genetic differentiation seen between populations, there is frequent discussion of the classification of various regional populations of apes into subspecies or even different species. Two trends occur among scientists interested in such classificatory issues. Some argue that gorillas, for example, should be split into different species based on behavioral, ecological, and genetic differences. A reflection of this trend is the increasing number of recognized African primate species from 63 to 79 in the last decade (50). Conversely, others stress the great similarity between humans and chimpanzees and suggest the fusion of humans and chimpanzees into one genus containing the species *Pan sapiens*, *Pan troglodytes* (chimpanzees), and *Pan paniscus* (bonobos) (135). Although the former stress the uniqueness of different ape populations, often in a laudable zeal to argue for their protection, the latter stress our close evolutionary connection to the apes, also often to argue for their protection.

In our opinion, so-called subspecies of chimpanzees, gorillas, and orangutans share closely related or even identical genomic DNA sequences in parts of their genomes (64). Thus, they are not genetically distinct in any absolute sense. From a genomic perspective, it is not clear if it is sensible to regard them as subspecies or even species, as some advocate (e.g., 145). In fact, various ape populations carry no or very marginal differences in their appearance and behavior when compared to the corresponding differences seen in humans from different continents. Because there is no good reason to talk about subspecies or species for humans, it seems absurd to use these categories for apes. In the end, the discussion of such classifications itself has limited biological relevance even if they may have profound political and legal implications.

GENOME EVOLUTION

Hitherto, our understanding of the evolution of primate genomes has been severely hampered by a lack of data from nonhuman primates. However, this is now changing as more and more DNA sequences from primates are accumulating. Most

excitingly, the call for primate genome projects (e.g., 89, 124) has materialized with the sequencing of the genome of a common chimpanzee and the rhesus macaque to follow (63). Other large-scale efforts include the private attempt to sequence more than 200,000 protein-coding exons in the chimpanzee (18), the chromosome 21 ortholog in the chimpanzee (129), the more than 100,000 end reads from a chimpanzee bacterial artificial chromosome (BAC) library (40), the chimpanzee's MHC class I region (1), and the parts of the chimpanzee's Y chromosome (111). In addition, BAC libraries from 13 nonhuman primates are available (http://bacpac.chori.org/) (25) and are used, for example, by the comparative vertebrate sequencing initiative (http://www.nisc.nih.gov/).

The availability of genome sequences from the chimpanzee and other primates will greatly increase our knowledge of how genomes, chromosomes, and genes evolve over relatively short evolutionary timescales. This is important because knowledge about these phenomena forms the basis for identifying functionally conserved genomic features as well as functionally relevant genomic changes among primate species (see below).

Chromosome Evolution

Karyotypes have changed little during primate evolution (e.g., 92), although some groups, such as gibbons, owl monkeys, or lemurs (99), show faster rates of chromosomal reorganization. The human and chimpanzee karyotypes differ by 10 euchromatic rearrangements seen by G-banding: the telemore fusion of two chromosomes resulting in human chromosome 2 and nine pericentric inversions (142). This fusion site (33), as well as a few other rearrangement sites (e.g., 68, 83), were mapped at the molecular level, but no obvious functional consequences were detected.

Chromosomal rearrangements could promote speciation by several mechanisms (107). Navarro & Barton recently proposed a model where the blocked gene flow in rearranged chromosomal regions would facilitate the evolution of alleles compatible within—but incompatible between—species (94). This model was supported by the finding that ratios of amino acid substitutions to silent substitutions (*ka*/*ks* ratios, see below) between human and chimpanzee genes were higher in rearranged chromosomes than in nonrearranged chromosomes, indicating more occurrences of positive selection on rearranged chromosomes (95). However, this finding was likely caused by the biased set of genes analyzed (86). Thus, the role of chromosomal rearrangements for speciation in primates remains unclear.

Segmental Duplication

About 5% of the human genome is located in so-called segmental duplications (3), i.e., relatively large (1- to >200-kbp) regions that are present in at least two copies per haploid genome and are 90% to 100% identical to each other (reviewed in 114). Among the fully sequenced genomes, the human genome has the highest amount of segmental duplications. For example, only ~2% and ~3% of the mouse

and rat genome, respectively, is found in segmental duplications (6a, 16, 123a). It has been suggested that the increased spread of retroposed copies of Alu elements in the ancestor of Old World monkeys and New World monkeys (see below) facilitated the expansion of segmental duplication by recombination between Alu elements. Once established, the duplications may facilitate further recombination due to their high level of sequence similarity (4). There are several reasons why segmental duplications are relevant when comparing primate genomes. First, due to their high sequence identity and large size, segmental duplications are difficult to correctly identify and localize in a genome (5, 77). Second, they can lead to the alignment of paralogous, instead of orthologous, regions, i.e., the comparison of regions that share a common ancestor due to a gene duplication instead of due to a speciation event. Indeed, a lower divergence between humans and chimpanzees has been found when omitting comparisons in duplicated regions (82). Third, it seems that segmental duplications are hotspots for chromosomal rearrangements (119). For example, a chimpanzee-specific inversion is located in a cluster of segmental duplications (83), and the chromosome fusion region on human chromosome 2 is highly enriched for segmental duplications (32). Fourth, they have the potential to create "new" genes and could therefore play an important role for species-specific phenotypes. For example, analysis of duplications on human chromosome 22 revealed 11 cases of transcribed genes that were created either by whole gene duplications, were modified by segmental duplications, or contained exons derived from different duplication events (6). The potential to generate new genes by this process is intriguing.

Deletions of genomic regions are much less studied than duplications because it is not possible to identify deletions by studying a single genome or by comparing two genomes. However, the sequencing of several primate genomes and comparative genomic hybridization (CGH) techniques (84) will make it possible to study also deletions on a genome-wide scale in the near future.

Mobile Elements

Interspersed repeats derived from mobile elements that propagate either via a DNA intermediate (DNA transposons) or via a RNA intermediate (retroposons) account for at least 45% of the human genome (77). Especially relevant for comparisons among primates are the two currently active retroposons, the long interspersed element (LINE) L1 and the short interspersed element (SINE) Alu, which make up ~25% of the human genome (for reviews on mobile elements see 7, 23, 104).

Alu elements arose with the radiation of primates (146) and, with more than a million copies, make up more than 10% of the human genome (77). By analyzing the divergence between different subfamilies of Alu elements in the human genome, researchers showed that the rate of Alu insertions has changed (7). Most elements were inserted in the common ancestor of Old World and New World monkeys when an average of one Alu insertion occurred per birth, whereas today about one in 200 births occur in humans (22). Liu et al. observed more Alu and L1 elements in human and baboon DNA sequences than in lemur DNA sequences,

which suggests that a higher retroposition activity caused the genomes of humans and baboons to be larger than that of the lemur (82). Although the rate of Alu insertions has stayed relatively constant on the human lineage since the divergence from Old World monkeys, it has increased on the baboon lineage and decreased on the chimpanzee lineage (82). Together with more insertions of other retroposons, this has probably caused the human genome to become 1% larger than the chimpanzee genome (82, 129).

The insertion of mobile elements can have functional consequences by inactivating genes, inserting new coding sequences into genes, and changing the regulation of genes (11, 12, 97). Transposing L1 elements can lead to the cotransposition of the DNA sequence adjacent to the element and thus to exon shuffling (91). Unequal recombination between Alu elements can cause the deletion of genes or exons. This has been proposed for the human-specific deletion of exon 34 in the tropoelastin gene (122). Gene conversion can also occur between different Alu elements, and this may be responsible for the deletion of one exon in the gene encoding CMP-N-acetylneuraminic acid hydroxylase (CMAH) in humans (54), which led to the absence of a hydroxylated form of sialic acid in humans (17, 60). How often insertions of mobile elements have functional consequences such as these awaits further investigation.

Insertions and Deletions

Differences in the length of two homologous sequences (often called indels) can arise by different mutational mechanisms such as insertions of mobile elements (see above), deletions of a region by unequal recombination, or errors of the DNA replication machinery that cause smaller indels. The latter type of indels is the most frequent one (129) and is observed once per 1000 compared sites between human and chimpanzee DNA sequences (24).

Overall, indels are responsible for a large proportion of sequence differences between species. Using oligonucleotide arrays based on the human sequence of chromosome 21, Frazer et al. (39) found that ~9% of the sequence is deleted in either chimpanzee, orangutan, rhesus macaque, or a New World monkey. Watanabe et al. (129) identified ~68,000 indels among the ~3.3 million compared base pairs of the human chromosome 21 and its chimpanzee ortholog. A lineage-specific analysis of the larger indels showed that 32 kbp were gained and 39 kbp were lost on the human lineage, compared to 25 kbp and 53 kbp on the chimpanzee lineage. When BAC sequences are compared between humans and chimpanzees, 2.3% (82) to 3.4% (10) of the aligned sequence consists of indels, whereas point substitutions contribute ~1.2% of all sequence differences. However, the mutational mechanisms between different indels and between indels and point substitutions are different, so a comparison of such percentages is not necessarily meaningful.

Microsatellites

Indels occur especially frequently in microsatellites [i.e., tandemly repeated short (1–6 bp) DNA motifs], and the high mutation rate and correspondingly high

variation of microsatellites make them useful, for example, in linkage mapping (47) and paternity testing (125). Human microsatellites are generally longer than their chimpanzee orthologs, which implies that the mutation process can be different even in closely related species (112). Although the original analysis that this suggestion was based on was confounded by ascertainment bias (26, 132), a recent study using 5.1 Mb of genomic DNA alignments between humans and chimpanzees showed that at least dinculeotide repeats are on average ~2 bp longer in humans than in chimpanzees (132). A comparison of microsatellites between humans and baboons suggested that this property might have changed on the human lineage and, thus, that the mutation patterns of microsatellites may change rapidly during primate evolution (132).

Nucleotide Substitutions

Nucleotide substitutions are the most frequent type of genetic change and probably the best-understood mutational process (27). However, comparing sequences in closely related species such as humans and chimpanzees has the advantage of less ambiguous alignments and less frequent multiple substitutions and, thus, one can better discern the substitutional processes. At neutral sites (i.e., at sites where changes do not influence the fitness of the organism), one can use the observed substitution rate between species to estimate the mutation rate (80). Although the mutation rate per year is relatively constant among the great apes (e.g., 15, 66, 133), evidence shows a ~30% higher mutation rate in Old World monkeys (82, 138), a twice-as-high rate in lemurs (82), and a fivefold higher rate in mice (131). This might partially be explained by differences in generation times between the different species (81, 100), although the magnitude of this effect is debated (75).

Substitution rates within a genome vary at different scales from whole chromosomes to the dinucleotide context (27). At the chromosomal scale, the Y chromosome evolves more quickly than the autosomes, whereas the X chromosome evolves more slowly. This is probably largely caused by a ~fivefold higher mutation rate in the male germline, where Y chromosomes spend all their time, the autosomes spend half their time, and the X chromosomes spend one third of their time (87). However, the autosomes also differ significantly in substitution rates between humans and chimpanzees (24, 129), and these differences might even be conserved in the mouse (79). The reason for this is not clear. Other factors, such as GC content of a region (133) or the recombination rate (55), are also positively correlated with substitution rates in primates. The conservation of such factors may at least partly explain the differences in substitution rates at a chromosomal scale.

Interestingly, an excess of GC to AT substitutions exists in regions of high GC content in humans and chimpanzees (133). This means that the GC content of a region is not in equilibrium between mutation and drift, as previously thought, but that regions of high GC content are decreasing their GC content.

When discussing substitution rates, an important factor is the fraction of the DNA sequence comprised of CpG dinucleotides. Because CpG dinucleotides can

be methylated at the 5′ position of the cytosine residue, deamination results in bona fide mispaired thymine residue. The inefficient restitution to the original sequence of such perturbations in the DNA molecule by repair mechanisms makes CpG sites hotspots of mutation in mammalian genomes. This is evident because 28% of all substitutions observed between alignments of human and chimpanzee DNA sequences are transitions at CpG sites (24). Thus, CpG sites have a 23-fold higher transition rate than average sites It is important to consider this, for example, when estimating mutation rates at silent sites (56, 121).

Protein-Coding Regions

Protein-coding regions are of obvious importance for the organism. They contain sites at which mutations do not change the amino acid sequence of the encoded protein (silent sites) and these sites evolve neutrally or nearly neutrally. This is advantageous for analyzing protein-coding regions because the number of differences per silent sites (ks) can be used to gauge the putatively neutral substitution rate for the gene in question. This rate can then be compared to the number of substitutions per nonsilent sites at which mutations do change the encoded amino acid (ka). The ratio ka/ks can thus be used to gain an overall view of the extent and mode of selection affecting the gene. If the ratio is above one, positive selection affected a substantial number of positions; if the ratio is close to one, the gene evolves as if the amino acid sequence was not functionally significant; if the ratio is below one, functional constraints limit the extent to which amino acid changes affected the protein (for a recent review, see e.g., 35). Using this approach, ∼70% of all nucleotide substitutions leading to amino acid changes are deleterious in humans and chimpanzees (56).

Information about intraspecific variation adds additional value to such analyses. For example, Fay et al. (36) used the ka/ks ratio of polymorphisms at intermediate frequency to estimate the extent of neutral amino acid changes within humans. They excluded rare polymorphisms because these could be slightly deleterious and therefore not neutral. They found ∼35% more amino acid changes between humans and Old World monkeys than expected, suggesting that a surprisingly large proportion of amino acid differences between species would have been driven to fixation by positive selection. However, because the genes in their dataset on divergence and the genes used to collect the polymorphism data were not identical, this estimate may need to be revised. Fortunately, larger datasets that will enable researchers to perform this type of analysis on a truly genomic scale will soon be available.

Besides nucleotide substitutions and small insertions and deletions, gene duplications, exon shuffling, retrotransposition, and gene fusions can change protein-coding regions in a way that can allow the emergence of "new genes." Examples of genes that arose by such mechanisms during primate evolution are accumulating (reviewed in 85). Comparing primate genome sequences will reveal how frequently such events occur.

Non-Genic Conserved Elements

Identifying genomic regions that are conserved between species allows the identification of regions that are functionally important outside of classical genes. Regulatory regions are an important category of such regions (19). However, if a regulatory function is conserved only in primates, it cannot be studied in distantly related organisms such as the mouse, but can be studied only in primates. However, primates are relatively closely related, so functionally important DNA sequences cannot be identified by virtue of their conservation between any two species because nonconstrained regions have not diverged enough. Bofelli et al. (9) recently showed how to overcome this by using "phylogenetic shadowing," in which many primate species are sequenced to identify DNA sequence motifs that do not change within a group of related organisms. For example, the gene encoding apolipoprotein (a) (apo(a)) has orthologs only in Old World monkeys (78). By sequencing an upstream region of the apo(a) gene in 18 Old World monkeys, Bofelli et al. (9) identified several regulatory elements that were functionally relevant for the human apo(a) gene. When several primate genomes become available, researchers will be able to identify primate-specific conserved elements based on this approach.

Recombination

Recombination rates vary across the genome at a megabase scale and differ between the female and male germ lines (e.g., 71). The finding that the genetic map of the baboon is 28% smaller than the human map (108) shows that global recombination rates can change relatively quickly during evolution. By contrast, little is known about how recombination rates are distributed at a fine scale across the genome and how these frequencies change over time. In a few cases, recombination hotspots of approximately 1–2 kbp have been defined by sperm typing (e.g., 61). Certain alleles at such hotspots tend to initiate recombination (62). Because this leads to the conversion of that DNA sequence to the sequence of its recombination partner (62), hotspots may have a short half-life in the genome. The finding that two recombination hotspots in humans are not hotspots in chimpanzees (106a, 127) indicates that this may be true. However, we only begin to understand to what extent recombination is localized to hot spots of this type in the genome and to what extent it occurs at more random locations (90a). Comparing the haplotype landscapes of human and chimpanzee chromosomes will be one way to resolve this question.

MOLECULAR PHENOTYPES

Although linear DNA sequences are not influenced by the environment, gene expression occurs as the result of influences originating both within the organism and outside the organism. In this respect, gene expression and associated phenomena

such as DNA methylation and chromatin structure are similar to phenotypes such as morphology and behavior. Thus, it is useful to think of such phenomena as "molecular phenotypes."

The development of microarrays has revolutionized the analysis of molecular phenotypes in that the transcripts of essentially all genes in a mammalian tissue can be quantitated in a single experiment. To a certain degree, arrays developed to analyze human genes can also be used to analyze gene expression in closely related primates. Thus, as is true for the study of DNA sequences from entire genomes, array-based transcriptome analyses allow insights into general evolutionary mechanisms. They enable the identification of genes that are outliers and therefore potentially responsible for adaptations of the organism.

Three studies have used high-density oligonucleotide arrays to compare gene expression profiles between humans and chimpanzees (13, 29, 67). They have shown that many species-specific gene expression differences exist in all tissues analyzed (brain, liver, fibroblasts, heart). Exact numbers depend on the criteria used to define an expression difference and on the power of the particular study to distinguish real differences from false positives. However, approximately 3–7% of all reliably detected transcripts are expressed differently in a given tissue (e.g., 13, 29). Environmental factors such as postmortem conditions will likely impact the gene expression profile of an individual. However, at least in one dataset (29), less than 10% of the expression differences consistently observed between species can be explained by such factors, as indicated by permutation analyses (W. Enard, unpublished results). Out of 169 genes identified as expressed differently between human and chimpanzee brain samples in one study, 167 showed a qualitatively similar expression difference in another study (13). In addition, there were species-specific differences to a similar extent between two mouse species for which the environmental conditions were controlled as much as possible (29). This suggests that most of the consistently observed expression differences between, e.g., humans and chimpanzees are not caused by random or systematic differences in the environment, and likely have a genetic basis.

A trivial genetic cause for gene expression differences are nucleotide differences between the species, leading to a lower hybridization efficiency of mRNAs from nonhuman primates to the human arrays. Approximately 20% of the genes identified as differentially expressed between humans and chimpanzees on the oligonucleotide arrays could not be confirmed by independent analyses such as Northern blots or RT-PCR (13, 29, 67). An analysis using only array oligonucleotides that have no sequence difference between humans and chimpanzees similarly suggests that hybridization differences cause ~20% of genes to be differently expressed (I. Hellmann & P. Khaitovich, unpublished observation). Other possible genetic causes include gene deletions, gene duplications, differences in alternative splicing, use of different polyadenylation sites (e.g., 13), changes in the regulatory region of the gene, changes in RNA stability, different effective concentrations of proteins regulating the gene, or genetic changes leading to a different cellular composition in the tissue analyzed. The relative extents to

which these genetic causes contribute to the observed differences are currently unknown.

It is also unclear what proportion of interspecies expression differences have functional consequences in that they became fixed by positive selection rather than by neutral drift. Because recent analyses indicate that most interspecies expression differences are probably neutral (68a), the null hypothesis for any particular expression difference should be functional neutrality. In this context it is also important to note that almost all interspecies expression differences are shared among cerebral cortex regions, most with other brain regions, and a large fraction even with tissues from other organs (13, 67; P. Khaitovich, personal communication). Therefore, even if a gene expression difference is the result of positive selection, it may have no functional consequences in the particular tissue examined.

From the data collected so far, it seems that more gene expression changes occurred in the human brain than in the chimpanzee brain (29), that differences in the brain are more often due to a higher expression in humans than are due to a lower expression in humans (13, 51), and that more differences are observed in liver than in brain (29, 57). More individuals and more tissues need to be examined to fully understand how the primate transcriptome evolves. The availability of the chimpanzee genome, the development of algorithms that detect oligonucleotides whose hybridization are influenced by sequence differences (93a), and arrays designed for various primate species will greatly facilitate this endeavor.

The evolution of other molecular phenotypes is of equal or even greater importance than that of the transcriptome. A first small-scale study using microarrays to examine differences in DNA methylation suggests that CpGs might generally be more methylated in human brains than in chimpanzee brains (28). Also, scientists have taken the first steps to analyze differences in protein expression using two-dimensional gel (29, 41). However, a global understanding of these phenomena awaits the development of high-throughput approaches that allow them to be studied at a genome-wide scale.

HUMAN-SPECIFIC ADAPTIVE CHANGES

Perhaps the most interesting application of comparisons between primate genomes is the identification of genetic changes that have functional consequences—especially those that may be responsible for adaptive changes in the human phenotype (see 14, 42, 52, 101, 124 for recent reviews). Therefore, we use humans as an example to discuss how to identify species-specific genetic changes of functional relevance.

Fixation

The definition of a human-specific change is that it took place on the human evolutionary lineage and that it it occurs in all currently living humans (i.e., that it is fixed among humans).

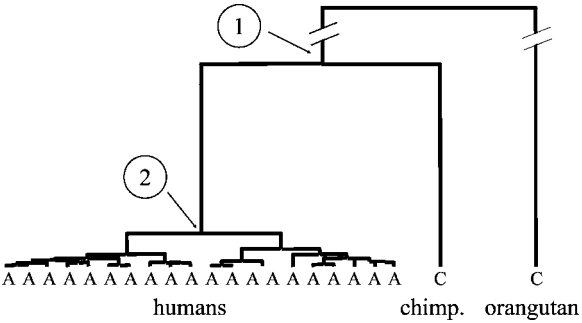

Figure 5 Genealogy of a human-specific change as expected under the standard neutral model. A genetic change can be defined as human-specific if it occurred between time points (1) and (2). Assuming that (1) and (2) are six million and one million years ago, respectively, a change is specific to humans in more than 80% of the cases if one human is analyzed. Observing the change in a few humans considerably increases this chance.

By analyzing one human and one chimpanzee DNA sequence, as well as one or preferably more outgroups such as the orangutan and the macaque, one can establish that a change occurred on the human lineage. That a change is fixed among humans can be established by analyzing several humans from various regions of the world. How many individuals need to be analyzed can be gauged from the statistical framework provided by the coalescent theory (see e.g., 110 for a review). If the most recent common ancestors (MRCAs) of human and chimpanzee autosomal DNA sequences go back on average 6 million years (46), and the MRCA of human autosomal DNA sequences is ~1 million years old (31), a randomly genetic change on the human lineage has a probability of more than 80% to be fixed among humans (Figure 5). This probability increases drastically if the change is observed in several humans. Under the standard neutral model of a randomly mating population of constant size, the probability that n samples from the population have the same MRCA as the whole population is $(n-1)/(n+1)$ (115). Therefore 2, 10, and 100 samples have the same MRCA as the whole population in 33%, 82%, and 99% of the cases, respectively, increasing the chance that a genetic change is fixed in currently living humans to 86.7%, 96.4%, and 99.6%, respectively. Although these estimates rely on simplifying assumptions, a worldwide sample of only 10 to 100 human chromosomes is enough to establish that any particular genetic variant is likely fixed among humans.

Positively Selected Changes

The hypothesis that a change fixed in all humans is adaptive (i.e., driven to fixation by positive selection during human evolution) is more difficult to test (see e.g., 35, 72, 98 for reviews). When comparing DNA sequences between species, positive selection can be inferred by testing whether more changes occurred in a putatively

positively selected category of sites than expected from the neutral mutation rate. The most common methods of this kind are versions of the ka/ks test in which positive selection is assumed if the rate of amino acid substitutions (ka) is greater than the rate of neutral substitutions, as estimated from substitutions at silent sites (ks). This test is robust against violations of demographic assumptions, but can only detect positive selection in a lineage if it acted at several sites of the protein. A couple of such fast-evolving genes were found in primates, many of which are involved in reproduction and immunity (136).

The McDonald-Kreitman test (90) is a variant of this test that infers positive selection if the ka/ks ratio between species is higher than would be expected from the ka/ks ratio seen within the species. Using a variant of this test, Gilad et al. (45) found suggestive evidence that positive selection acted on olfactory receptor genes in humans but not in chimpanzees.

Yang & Nielsen (137) recently developed a method in which interspecies comparisons are used to estimate the number of sites under constraint. In a study that heralds the new era of genome-wide screens for selected genes, Clark et al. (18) used a variant of this method to analyze 7645 orthologous gene trios from human, chimpanzee, and mouse to infer positive selection on the human lineage. They identified 178 genes in which more amino acid changes took place on the human lineage than expected ($p < 0.01$). However, because the amount of constraint is inferred mainly from the lineage leading to the mouse, a relaxation of constraint on the human and/or the chimpanzee lineage can also lead to a low p-value in this approach. Nevertheless, Clark et al. identified several genes and categories of genes that allow researchers to formulate interesting hypotheses about selected phenotypes during human evolution.

Evidence for positive selection can also be found by studying the patterns of intraspecific variation in DNA sequences affected by a selective sweep due to the so-called hitchhiking effect (88). The basis of this is that DNA is inherited from one generation to the next in blocks so that if a new advantageous allele arises and spreads through the population, adjacent regions of DNA hitchhike with the advantageous site. After such a selective sweep (i.e., when all individuals carry the advantageous allele), adjacent regions have a genealogy different from what is expected in the absence of selection (Figure 6). There are several statistical tests that can detect the effects of a selective sweep in such data by comparing the observed pattern to neutral expectations. For example, the Tajima's (123) *D* statistic analyzes the excess of rare DNA sequence variants created by a selective sweep, Fay & Wu's (34) *H* can detect an excess of high-frequency variants, and the Hudson-Kreitman-Aguadé (HKA) test (58) asks whether the level of diversity is lower than expected for a region not affected by a selective sweep. Recently, maximum-likelihood methods that combine these different sources of information have been developed (69, 106). They may currently be the most reliable approach to detect selective sweeps in the human genome.

It is important to realize that violations of demographic assumptions (for example, population growth or negative selection acting on many sites in the analyzed

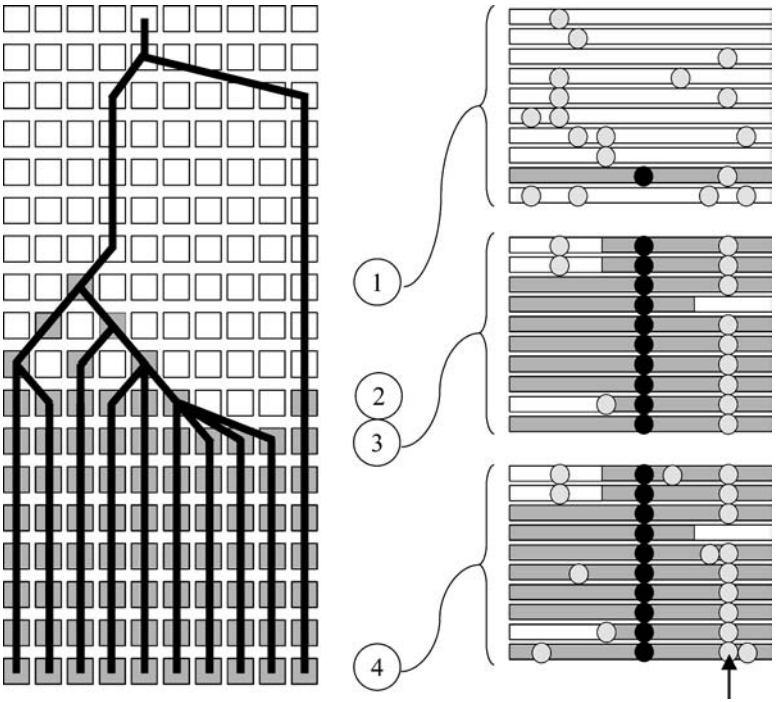

Figure 6 The hitchhiking effect of a positively selected site under the standard neutral model. The favorable variant (*black dot*) (1) is advantageous in individuals carrying it (*gray squares*) and spreads through the population by positive selection until becoming fixed (3). Recombination occurs during the sweep (2). The observed pattern of variation in the population after such a sweep (4) differs from what is seen without the sweep (1). For example, more low-frequency-derived polymorphisms *(light gray dots*) are expected [two alleles are at 10% frequency in (1) and six alleles are at 10% frequency in (4)]. In addition, derived polymorphisms that were on the same background as the favorable variant will get to high frequencies due to the sweep. Because of recombination, this can lead to derived polymorphisms at a high frequency after the sweep (*arrow*). Modified from a figure courtesy of Molly Przeworski.

region), can lead to false support for a selective sweep. Fortunately, increasing amounts of information on DNA sequence variation among humans from different genomic locations will make it possible to estimate these effects much better in the near future and thus will increase the accuracy and the power of these tests.

A fundamental limitation of all methods based on the study of intraspecific variation is that the signature of a selective sweep can be detected only for a short time period after fixation of the favorable variant. Under certain simplifying assumptions, a fixation that happened more than 8000 generations or ~200,000 years ago will not likely be detected in humans (105). Nevertheless, the first example of a human-specific selective sweep could recently be identified in the FOXP2 gene,

in which the analysis of human sequence variation next to two human-specific amino acid changes demonstrated that a selective sweep likely occurred (30, 143). Because FOXP2 is the first gene known to specifically affect the development of speech and language in humans (76), it is possible that the selected phenotype was related to speech and language.

Hypothesis Testing?

So far, single genes have been tested for positive selection, generally because they are involved in a phenotype that was known or assumed to have been selected. The availability of complete genome sequences will soon offer the new and exciting possibility to screen whole genomes for genes under positive selection based both on interspecies comparisons and intraspecific variation (e.g., 18). As with other discovery-driven approaches, this will produce lists of candidate genes. These can be used to generate hypotheses about selected phenotypes based on the known functions of the genes. However, direct tests of such hypotheses are generally impossible because transgenic techniques are currently not available in any primate species. In fact, ethical concerns might also prevent this in the future. Therefore, the challenge is to develop ways in which circumstantial evidence for hypotheses regarding primate phenotypes can be gathered in vitro and in other model organisms such as the mouse. This is going to be difficult. However, examples such as the duplication and point mutations in opsin genes that are coupled to the evolution of trichromatic color vision (e.g., 139), and the duplication of the RNAse1 gene and the subsequent adaptations of one copy to digest bacterial RNA in the foregut of leaf-eating monkeys (144), show that important insights can be gained from circumstantial evidence.

In this context, direct evidence for a hypothesized phenotypic change on the human evolutionary lineage can be obtained if scientists can find human individuals carrying back mutations to the ancestral state. In principle, the human population is large enough for this to be accomplished. For example, the frequency of transitional substitutions between humans and chimpanzees is 0.0087 (24). Assuming 11 million years of divergence between humans and chimpanzees and a generation time of 20 years, this results in a back mutation rate of 1.58×10^{-8} per generation for a human-specific transition. Among 6 billion people, one expects 190 novel back mutations to occur per generation. It will not be easy to find an individual with a particular back mutation, but it is comforting that natural variation in humans is large enough to allow some hypotheses about the functional consequences of human-specific changes to be directly tested, at least in principle.

OUTLOOK

In the nearer future, the comparison of primate genome sequences will open numerous novel perspectives. First, the comparison of closely related primate genomes will enable the development of realistic models for how various types of genomic changes such as indels, gene inactivations, or gene transpositions occur. It will also

be possible to develop models for substitutional changes that take into account the mutation spectrum in different genomic regions and on different evolutionary lineages. Such models will allow the identification of individual genomic events that are exceptional and thus likely candidates for positive selection. Second, as the efficient integration of different genomic data on gene expression, DNA sequence variation, and functional annotation of genomes progresses, the power to detect genetic changes that are relevant to the phenotype will increase. Third, progress in other fields will increase the knowledge of how genomic changes cause phenotypic changes. Thus, we will increasingly be able to interpret how genetic differences observed between primate species may affect the organism.

As this work proceeds, things other than the availability of genome sequences will start to limit progress. Perhaps the largest limitation will be the dearth of truly comparative studies of the phenotypes of humans and the great apes. For example, there are indications that chimpanzees are less susceptible for epithelial neoplasms (124), but rigorous studies are lacking mainly through the limited number of older chimpanzees analyzed. Clearly there is a need for more extensive comparative studies of relevant phenotypes, as was recently proposed (124a). It will be important that cognitive scientists, primatologists, and geneticists work together to design studies that are comparative in nature and amenable to genetic interpretation.

Once genes that carry differences of possible importance for a human-specific trait are identified, a serious limitation will be that there are—per definition—no animal models available for the study of human-specific traits. Possible approaches to these questions will include cell culture systems, transgenic mice, and perhaps large screens for back mutations in humans. It is encouraging that relevant insights into, e.g., human neurological diseases can be gained from animal models (130). Nevertheless, a huge challenge will be to design models for human-specific traits in experimental systems such as the mouse.

Last but not least, a serious limitation to our ability to perform comparative phenotypic work of human-specific traits is that the survival of all great apes in the wild is acutely threatened (128). This is not only a tragedy for the apes and for humans who enjoy seeing and studying them in the wild, it severely limits our ability to define the scope of the variability of their cultural and cognitive abilities. Ape populations vary greatly in behavior, and these differences are transmitted from generation to generation as cultural traditions (134). Most likely, these traditions developed over hundreds or even thousands of years. This means that to truly understand some of the most interesting aspects of the similarities and differences between humans and apes, it is important that apes exist in varied and undisturbed environments. Is it possible that the increasing interest in the genomes of primates can result in an increased support for the conservation of their African and Asian habitats?

ACKNOWLEDGMENTS

We apologize to all whose work is not cited due to space limitations or oversight on our part. We are greatly indebted to previous and current members of our laboratory and to the MPI-EVA for input and discussion. Our work is

funded by the Max Planck Society and the Bundesministerium für Bildung und Forschung.

The *Annual Review of Genomics and Human Genetics* is online at http://genom.annualreviews.org

LITERATURE CITED

1. Anzai T, Shiina T, Kimura N, Yanagiya K, Kohara S, et al. 2003. Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. *Proc. Natl. Acad. Sci. USA* 100:7708–13
2. Arnason U, Gullberg A, Burguete AS, Janke A. 2000. Molecular estimates of primate divergences and new hypotheses for primate dispersal and the origin of modern humans. *Hereditas* 133:217–28
3. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, et al. 2002. Recent segmental duplications in the human genome. *Science* 297:1003–7
4. Bailey JA, Liu G, Eichler EE. 2003. An Alu transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.* 73:823–34
5. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* 11:1005–17
6. Bailey JA, Yavor AM, Viggiano L, Misceo D, Horvath JE, et al. 2002. Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am. J. Hum. Genet.* 70:83–100

6a. Bailey JA, Church DM, Ventura M, Rocchi M, Eichler EE. 2004. Analysis of segmental duplications and genome assembly in the mouse. *Genome Res.* 14: 789–801

7. Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3:370–79
8. Bischoff JL, Shamp DD. 2003. The Sima de los Huesos hominids date to beyond U/Th equilibrium (>350 kyr) and perhaps to 400–500 kyr: new radiometric dates. *J. Archaeol. Sci.* 30:275–80
9. Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, et al. 2003. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* 299:1391–94
10. Britten RJ. 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl. Acad. Sci. USA* 99:13633–35
11. Brosius J. 1999. Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* 107: 209–38
12. Brosius J. 2003. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* 118:99–116
13. Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, et al. 2003. Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl. Acad. Sci. USA* 100:13030–35
14. Carroll SB. 2003. Genetics and the making of Homo sapiens. *Nature* 422:849–57
15. Chen F-C, Li W-H. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* 68:444–56
16. Cheung J, Wilson MD, Zhang J, Khaja R, MacDonald JR, et al. 2003. Recent

segmental and gene duplications in the mouse genome. *Genome Biol.* 4:R47
17. Chou HH, Takematsu H, Diaz S, Iber J, Nickerson E, et al. 1998. A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proc. Natl. Acad. Sci. USA* 95:11751–56
18. Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, et al. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* 302: 1960–63
19. Cooper GM, Sidow A. 2003. Genomic regulatory regions: insights from comparative sequence analysis. *Curr. Opin. Genet. Dev.* 13:604–10
20. Deinard A, Kidd K. 1999. Evolution of a HOXB6 intergenic region within the great apes and humans. *J. Hum. Evol.* 36: 687–703
21. Deinard AS, Kidd K. 2000. Identifying conservation units within captive chimpanzee populations. *Am. J. Phys. Anthropol.* 111:25–44
22. Deininger PL, Batzer MA. 1999. Alu repeats and human disease. *Mol. Genet. Metab.* 67:183–93
23. Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr. 2003. Mobile elements and mammalian genome evolution. *Curr. Opin. Genet. Dev.* 13:651–58
24. Ebersberger I, Metzler D, Schwarz C, Paabo S. 2002. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am. J. Hum. Genet.* 70:1490–97
25. Eichler EE, DeJong PJ. 2002. Biomedical applications and studies of molecular evolution: a proposal for a primate genomic library resource. *Genome Res.* 12:673–78
26. Ellegren H, Primmer CR, Sheldon BC. 1995. Microsatellite 'evolution': directionality or bias? *Nat. Genet.* 11:360–62
27. Ellegren H, Smith NG, Webster MT. 2003. Mutation rate variation in the mammalian genome. *Curr. Opin. Genet. Dev.* 13:562–68
28. Enard W, Fassbender A, Model F, Adorjan P, Paabo S, Olek A. 2004. Differences in DNA methylation patterns between humans and chimpanzees. *Curr. Biol.* 14: R148–49
29. Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, et al. 2002. Intra- and interspecific variation in primate gene expression patterns. *Science* 296:340–43
30. Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, et al. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418:869–72
31. Excoffier L. 2002. Human demographic history: refining the recent African origin model. *Curr. Opin. Genet. Dev.* 12:675–82
32. Fan Y, Linardopoulou E, Friedman C, Williams E, Trask BJ. 2002. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes. *Genome Res.* 12:1651–62
33. Fan Y, Newman T, Linardopoulou E, Trask BJ. 2002. Gene content and function of the ancestral chromosome fusion site in human chromosome 2q13-2q14.1 and paralogous regions. *Genome Res.* 12:1663–72
34. Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–13
35. Fay JC, Wu CI. 2003. Sequence divergence, functional constraint, and selection in protein evolution. *Annu. Rev. Genomics Hum. Genet.* 4:213–35
36. Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. *Genetics* 158:1227–34
37. Fischer A, Wiebe V, Paabo S, Przeworski M. 2004. Evidence for a complex demographic history of chimpanzees. *Mol. Biol. Evol.* 21: 799–808
38. Fleagle JG. 1999. *Primate Adaptation and Evolution*. San Diego: Academic
39. Frazer KA, Chen X, Hinds DA, Pant PV, Patil N, Cox DR. 2003. Genomic

DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Res.* 13:341–46
40. Fujiyama A, Watanabe H, Toyoda A, Taylor TD, Itoh T, et al. 2002. Construction and analysis of a human-chimpanzee comparative clone map. *Science* 295:131–34
41. Gagneux P, Amess B, Diaz S, Moore S, Patel T, et al. 2001. Proteomic comparison of human and great ape blood plasma reveals conserved glycosylation and differences in thyroid hormone metabolism. *Am. J. Phys. Anthropol.* 115:99–109
42. Gagneux P, Varki A. 2001. Genetic differences between humans and great apes. *Mol. Phylogenet. Evol.* 18:2–13
43. Gagneux P, Wills C, Gerloff U, Tautz D, Morin PA, et al. 1999. Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc. Natl. Acad. Sci. USA* 96:5077–82
44. Deleted in proof
45. Gilad Y, Bustamante CD, Lancet D, Paabo S. 2003. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am. J. Hum. Genet.* 73:489–501
46. Glazko GV, Nei M. 2003. Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* 20:424–34
47. Goldstein D, Schlotterer C. 1999. *Microsatellites: Evolution and Applications.* Oxford: Oxford Univ. Press
48. Goodman M. 1999. The genomic record of Humankind's evolutionary roots. *Am. J. Hum. Genet.* 64:31–39
49. Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, et al. 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* 9:585–98
50. Grubb P, Butynski TM, Oates JF, Bearder SK, Disotell TR, et al. 2003. Assessment of the diversity of African primates. *Int. J. Primatol.* 24:1301–57
51. Gu J, Gu X. 2003. Induced gene expression in human brain after the split from chimpanzee. *Trends Genet.* 19:63–65
52. Hacia JG. 2001. Genome of the apes. *Trends Genet.* 17:637–45
53. Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST. 1998. Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA* 95:1961–67
54. Hayakawa T, Satta Y, Gagneux P, Varki A, Takahata N. 2001. Alu-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. *Proc. Natl. Acad. Sci. USA* 98:11399–404
55. Hellmann I, Ebersberger I, Ptak SE, Paabo S, Przeworski M. 2003. A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* 72:1527–35
56. Hellmann I, Zollner S, Enard W, Ebersberger I, Nickel B, Paabo S. 2003. Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res.* 13:831–37
57. Hsieh WP, Chu TM, Wolfinger RD, Gibson G. 2003. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* 165:747–57
58. Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–59
59. Huxley TH. 1863. *Evidence as to Man's Place in Nature*. New York: D. Appleton
60. Irie A, Koyama S, Kozutsumi Y, Kawasaki T, Suzuki A. 1998. The molecular basis for the absence of N-glycolylneuraminic acid in humans. *J. Biol. Chem.* 273:15866–71
61. Jeffreys AJ, Kauppi L, Neumann R. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.* 29:217–22
62. Jeffreys AJ, Neumann R. 2002. Reciprocal crossover asymmetry and meiotic

drive in a human recombination hot spot. *Nat. Genet.* 31:267–71
63. Jensen-Seaman MI, Deinard AS, Kidd KK. 2001. Modern African ape populations as genetic and demographic models of the last common ancestor of humans, chimpanzees, and gorillas. *J. Hered.* 92:475–80
64. Kaessmann H, Pääbo S. 2002. The genetical history of humans and the great apes. *J. Intern. Med.* 251:1–18
65. Kaessmann H, Wiebe V, Paabo S. 1999. Extensive nuclear DNA sequence diversity among chimpanzees. *Science* 286: 1159–62
66. Kaessmann H, Wiebe V, Weiss G, Paabo S. 2001. Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nat. Genet.* 27:155–56
67. Karaman MW, Houck ML, Chemnick LG, Nagpal S, Chawannakul D, et al. 2003. Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. *Genome Res.* 13:1619–30
68. Kehrer-Sawatzki H, Schreiner B, Tanzer S, Platzer M, Muller S, Hameister H. 2002. Molecular characterization of the pericentric inversion that causes differences between chimpanzee chromosome 19 and human chromosome 17. *Am. J. Hum. Genet.* 71:375–88
68a. Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, et al. 2004. A neutral model of transcriptome evolution. *PLoS Biol.* 2:E132
69. Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160:765–77
70. Kitano T, Schwarz C, Nickel B, Paabo S. 2003. Gene diversity patterns at 10 X-chromosomal loci in humans and chimpanzees. *Mol. Biol. Evol.* 20:1281–89
71. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, et al. 2002. A high-resolution recombination map of the human genome. *Nat. Genet.* 31:241–47
72. Kreitman M. 2000. Methods to detect selection in populations with applications to the human. *Annu. Rev. Genomics Hum. Genet.* 1:539–59
73. Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, et al. 2000. A view of Neandertal genetic diversity. *Nat. Genet.* 26:144–46
74. Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* 90: 19–30
75. Kumar S, Subramanian S. 2002. Mutation rates in mammalian genomes. *Proc. Natl. Acad. Sci. USA* 99:803–8
76. Lai CSL, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519–23
77. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
78. Lawn RM, Boonmark NW, Schwartz K, Lindahl GE, Wade DP, et al. 1995. The recurring evolution of lipoprotein(a). Insights from cloning of hedgehog apolipoprotein(a). *J. Biol. Chem.* 270:24004–9
79. Lercher MJ, Williams EJ, Hurst LD. 2001. Local similarity in evolutionary rates extends over whole chromosomes in human-rodent and mouse-rat comparisons: implications for understanding the mechanistic basis of the male mutation bias. *Mol. Biol. Evol.* 18:2032–39
80. Li W-H. 1997. *Molecular Evolution*. Sunderland, MA: Sinauer
81. Li WH, Ellsworth DL, Krushkal J, Chang BH, Hewett-Emmett D. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol. Phylogenet. Evol.* 5:182–87

82. Liu G, Zhao S, Bailey JA, Sahinalp SC, Alkan C, et al. 2003. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.* 13:358–68
83. Locke DP, Archidiacono N, Misceo D, Cardone MF, Deschamps S, et al. 2003. Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. *Genome Biol.* 4:R50
84. Locke DP, Segraves R, Carbone L, Archidiacono N, Albertson DG, et al. 2003. Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res.* 13:347–57
85. Long M, Betran E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* 4:865–75
86. Lu J, Li WH, Wu CI. 2003. Comment on "chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes." *Science* 302: 988
87. Makova KD, Li WH. 2002. Strong male-driven evolution of DNA sequences in humans and apes. *Nature* 416:624–26
88. Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* 23:23–35
89. McConkey EH, Goodman M. 1997. A human genome evolution project is needed. *Trends Genet.* 13:350–51
90. McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351:652–54
90a. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* 304: 581–84
91. Moran JV, DeBerardinis RJ, Kazazian HH Jr. 1999. Exon shuffling by L1 retrotransposition. *Science* 283:1530–34
92. Muller S, Wienberg J. 2001. "Barcoding" primate chromosomes: molecular cytogenetic screening for the ancestral hominoid karyotype. *Hum. Genet.* 109:85–94
93. Murphy WJ, Stanyon R, O'Brien SJ. 2001. Evolution of mammalian genome organization inferred from comparative gene mapping. *Genome Biol.* 2: REVIEWS0005.1–8
93a. Nagpal S, Karaman MW, Timmerman MM, Ho VV, Pike BL, Hacia JG. 2004. Improving the sensitivity and specificity of gene expression analysis in highly related organisms through the use of electronic masks. *Nucleic Acids Res.* 32:e51
94. Navarro A, Barton NH. 2003. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* 57:447–59
95. Navarro A, Barton NH. 2003. Chromosomal speciation and molecular divergence–accelerated evolution in rearranged chromosomes. *Science* 300: 321–24
96. Nei M, Kumar AM. 2000. *Molecular Evolution and Phylogenetics*. pp. 75–77. Oxford: Oxford Univ. Press
97. Nekrutenko A, Li WH. 2001. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* 17:619–21
98. Nielsen R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* 86:641–47
99. O'Brien SJ, Stanyon R. 1999. Phylogenomics. Ancestral primate viewed. *Nature* 402:365–66
100. Ohta T. 1993. An examination of the generation-time effect on molecular evolution. *Proc. Natl. Acad. Sci. USA* 90: 10676–80
101. Olson MV, Varki A. 2003. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat. Rev. Genet.* 4:20–28
102. Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, Goodwin W. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404:490–93

103. Page SL, Goodman M. 2001. Catarrhine phylogeny: noncoding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. *Mol. Phylogenet. Evol.* 18:14–25
104. Prak ET, Kazazian HH Jr. 2000. Mobile elements and the human genome. *Nat. Rev. Genet.* 1:134–44
105. Przeworski M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* 160:1179–89
106. Przeworski M. 2003. Estimating the time since the fixation of a beneficial allele. *Genetics* 164:1667–76
106a. Ptak SE, Roeder AD, Stephens M, Gilad Y, Paabo S, Przeworski M. 2004. Absence of the TAP2 human recombination hotspot in chimpanzees. *PLoS Biol.* In press
107. Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16:351–58
108. Rogers J, Mahaney MC, Witte SM, Nair S, Newman D, et al. 2000. A genetic linkage map of the baboon (Papio hamadryas) genome based on human microsatellite polymorphisms. *Genomics* 67:237–47
109. Rosenberg HF, Feldmann MW. 2002. *The Relationship Between Coalescence Times and Population Divergence Times.* Oxford: Oxford Univ. Press
110. Rosenberg NA, Nordborg M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* 3:380–90
111. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, et al. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423:873–76
112. Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, et al. 1995. Microsatellite evolution–evidence for directionality and variation in rate between species. *Nat. Genet.* 10:337–43
113. Ruvolo M. 1997. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Mol. Biol. Evol.* 14:248–65
114. Samonte RV, Eichler EE. 2002. Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.* 3:65–72
115. Saunders IW, Tavare S, Watterson GA. 1984. On the genealogy of nested subsamples from a haploid population. *Adv. Appl. Prob.* 16:471–91
116. Schmitz J, Ohme M, Zischler H. 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of Tarsius bancanus to other primates. *Genetics* 157:777–84
117. Schmitz RW, Serre D, Bonani G, Feine S, Hillgruber F, et al. 2002. The Neandertal type site revisited: interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc. Natl. Acad. Sci. USA* 99:13342–47
117a. Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, et al. 2004. No evidence of Neandertal mtDNA contribution to early modern humans. *PLoS Biol.* 2:E57
118. Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci. USA* 100:1056–61
119. Stankiewicz P, Lupski JR. 2002. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* 18:74–82
120. Stone AC, Griffiths RC, Zegura SL, Hammer MF. 2002. High levels of Y-chromosome nucleotide diversity in the genus Pan. *Proc. Natl. Acad. Sci. USA* 99:43–48
121. Subramanian S, Kumar S. 2003. Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. *Genome Res.* 13:838–44
122. Szabo Z, Levi-Minzi SA, Christiano AM, Struminger C, Stoneking M, et al. 1999. Sequential loss of two neighboring exons of the tropoelastin gene during primate evolution. *J. Mol. Evol.* 49:664–71

123. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–95

123a. Tuzun E, Bailey JA, Eichler EE. 2004. Recent segmental duplications in the working draft assembly of the brown Norway rat. *Genome Res.* 14: 493–506

124. Varki A. 2000. A chimpanzee genome project is a biomedical imperative. *Genome Res.* 10:1065–70

124a. Varki A, Wills C, Perlmutter D, Woodruff D, Gage F, et al. 1998. Great ape phenome project? *Science* 282:239–40

125. Vigilant L, Hofreiter M, Siedel H, Boesch C. 2001. Paternity and relatedness in wild chimpanzee communities. *Proc. Natl. Acad. Sci. USA* 98:12890–95

126. Wall JD. 2003. Estimating ancestral population sizes and divergence times. *Genetics* 163:395–404

127. Wall JD, Frisse LA, Hudson RR, Di Rienzo A. 2003. Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates. *Am. J. Hum. Genet.* 73:1330–40

128. Walsh PD, Abernethy KA, Bermejo M, Beyers R, De Wachter P, et al. 2003. Catastrophic ape decline in western equatorial Africa. *Nature* 422:611–14

129. Watanabe H, Fujiyama A, Hattori M, Taylor TD, Toyoda A, et al. 2004. DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature.* 429:382–88

130. Watase K, Zoghbi HY. 2003. Modelling brain diseases in mice: the challenges of design and analysis. *Nat. Rev. Genet.* 4:296–307

131. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–62

132. Webster MT, Smith NG, Ellegren H. 2002. Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments. *Proc. Natl. Acad. Sci. USA* 99:8748–53

133. Webster MT, Smith NG, Ellegren H. 2003. Compositional evolution of non-coding DNA in the human and chimpanzee genomes. *Mol. Biol. Evol.* 20:278–86

134. Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, et al. 1999. Cultures in chimpanzees. *Nature* 399:682–85

135. Wildman DE, Uddin M, Liu G, Grossman LI, Goodman M. 2003. Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus Homo. *Proc. Natl. Acad. Sci. USA* 100:7181–88

136. Wolfe KH, Li WH. 2003. Molecular evolution meets the genomics revolution. *Nat. Genet.* 33(Suppl.):255–65

137. Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19: 908–17

138. Yi S, Ellsworth DL, Li WH. 2002. Slow molecular clocks in Old World monkeys, apes, and humans. *Mol. Biol. Evol.* 19:2191–98

139. Yokoyama S, Starmer WT, Yokoyama R. 1993. Paralogous origin of the red- and green-sensitive visual pigment genes in vertebrates. *Mol. Biol. Evol.* 10:527–38

140. Yu N, Chen FC, Ota S, Jorde LB, Pamilo P, et al. 2002. Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* 161:269–74

141. Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, et al. 2003. Low nucleotide diversity in chimpanzees and bonobos. *Genetics* 164:1511–18

142. Yunis JJ, Prakash O. 1982. The origin of man: a chromosomal pictorial legacy. *Science* 215:1525–30

143. Zhang J, Webb DM, Podlaha O. 2002. Accelerated protein evolution and

origins of human-specific features: Foxp2 as an example. *Genetics* 162: 1825–35

144. Zhang J, Zhang YP, Rosenberg HF. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* 30:411–15
145. Zhi L, Karesh WB, Janczewski DN, Frazier-Taylor H, Sajuthi D, et al. 1996. Genomic differentiation among natural populations of orang-utan (Pongo pygmaeus). *Curr. Biol.* 6:1326–36
146. Zietkiewicz E, Richer C, Sinnett D, Labuda D. 1998. Monophyletic origin of Alu elements in primates. *J. Mol. Evol.* 47:172–82

Annual Review of Genomics and Human Genetics
Volume 5, 2004

CONTENTS