

# Conservation genomics: applying whole genome studies to species conservation efforts

O.A. Ryder

Center for Reproduction of Endangered Species, Zoological Society of San Diego, San Diego, CA (USA)

**Abstract.** Studies of complete genomes are leading to a new understanding of the biology of mammals and providing ongoing insights into the fundamental aspects of the organization and evolution of biological systems. Comparison of primate genomes can identify aspects of their organization, regulation and function that appeared during the primate radiation, but without comparison to more evolutionarily distant mammals and other vertebrates, highly conserved aspects of genome architecture will not be accurately identified nor will the lineage-specific changes be identified as such. Many species of primates face risks of extinction; yet the knowledge of their genomes will provide a deeper understanding of primate adaptations, human origins, and provide the framework for discoveries anticipated to improve human medicine. The great apes,

the closest relatives of the human species, are among the most vulnerable and most important for human medical studies. However, apes are not the only species whose genomic information will enrich humankind. Comparative genomic studies of endangered species can benefit conservation efforts on their behalf. Increased knowledge of genome makeup and variation in endangered species finds conservation application in population evaluation monitoring and management, understanding phylogeography, can enhance wildlife health management, identify risk factors for genetic disorders, and provide insights into demographic management of small populations in the wild and in captivity.

Copyright © 2005 S. Karger AG, Basel

Genomic studies analyzing intra-specific and inter-specific differences offer to become an effective and significant conservation tool. The variation disclosed by these studies can provide a basis for understanding the geographical distribution of variation, infer demographic events such as bottlenecks and population expansions, identify loci under selection, and provide assessments of population structure, behavioral ecology and epidemiology in the recent to distant past. Improved con-

servation assessments and enhanced methods of medical diagnosis and population management incorporating data from genome studies may contribute to increasing the chances for survival of some endangered species.

The proposal to sequence the human genome and the subsequent unexpectedly rapid assembly of the draft human genome sequence created new expectations in human medicine and biology (Collins, 1999). Recently described aspects of the structure of the human genome, such as the partition of genes that are abundantly expressed into relatively GC-rich, SINE-rich and LINE-poor “ridges” (Versteeg et al., 2003) adds to earlier unexpected findings such as the estimated number of genes in the human genome (Lander et al., 2001; Venter et al., 2001). The trend for new, previously unanticipated, insights into the architecture of the human genome continues as the genomes of additional mammals are completed and applied to comparisons of overall syntenic structure and fine-scale organization in specific chromosomal regions (Hardison et al., 2003; Pevzner and Tesler, 2003; Thomas et al., 2003).

Supported by NSF 0094993, the Cesar Kleberg Wildlife Foundation, the John and Beverly Stauffer Foundation and the Zoological Society of San Diego.

Received 24 November 2003; manuscript accepted 11 March 2004.

Request reprints from: Oliver A. Ryder  
Center for Reproduction of Endangered Species  
Zoological Society of San Diego  
P.O. Box 120551, San Diego, CA 92112-0551 (USA)  
telephone: 619 557-3950; fax: 619 557-3958; e-mail: oryder@sandiegozoo.org

Comparative studies of genome sequences are leading to the identification of functional elements through identification of conserved sequence regions (Gumucio et al., 1992; Boffelli et al., 2003). Pairwise comparison of genome sequences identifies features held in common, but falls short of identifying all regions in each separate genome that are functional elements (Boffelli et al., 2003; Kirkness et al., 2003). Multi-species comparisons to identify functional elements require an evolutionary perspective in order to efficiently derive functional aspects of genome architecture (Eisen and Fraser, 2003).

Phylogenetic studies have been utilized to identify genome segments and gene regions that evolve more rapidly (Stankiewicz et al., 2001; Enard et al., 2002a, b; Stankiewicz and Lupski, 2002; Hellman et al., 2003) or to identify genome segments conserved in the radiation of mammals (Gumucio et al., 1992) and within the order Primates (Boffelli et al., 2003). The location of these conserved regions is consistent with a role as *cis*-acting regulatory elements that influence developmental programming (Davidson et al., 2003; Nobrega et al., 2003).

Studies of variation in the human mitochondrial (mt) genome led over the course of the last two decades to a view of human origins in and dispersal from Africa (Cann et al., 1987) that was supported by subsequent studies of Y chromosome variation (Hammer et al., 2001) and single nucleotide polymorphisms (SNPs) (Yu et al., 2002). In many ways, the utility of the early RFLP and sequencing studies helped produce expectations that fundamental attributes of human biology and evolution would be discovered by more ambitious DNA sequencing efforts. With the expansion of mtDNA studies to other hominoid species, the chimpanzees, bonobo, gorillas and orangutans, clearer definition of the evolutionary relationships among these species unfolded, though not without accompanying controversy. An early finding (Ferris et al., 1981), subsequently upheld (Gagneux et al., 1999), was that species of great apes displayed greater diversity in their mtDNA than did humans. The development of larger data sets and new analytical tools led to the suggestion of a population bottleneck in *Homo sapiens* in its relatively recent evolutionary past, e.g., within the last 50,000 – 200,000 years (Marth et al., 2003). Again, in anticipation of new findings and increasing resolving power that could be obtained by larger scale efforts at sequencing mt and nuclear loci, the importance of comparative studies in elucidating the history and definition of the human genome gained broad recognition. Similar analyses for chimpanzees, bonobos, gorillas and orangutans could greatly add to knowledge of the geographic distribution of the genetic variation within the species; such information could play a crucial role in recognizing conservation units (Morin et al., 1992; Ryder and Chemnick, 1993; Garner and Ryder, 1996; Zhi et al., 1996; Deinard and Kidd, 2000). The role of duplication of chromosomal segments in the evolution of mammalian (Kent et al., 2003; Pennacchio, 2003) and hominoid genomes (Stankiewicz et al., 2001; Locke et al., 2003) has been established and suggests that additional comparative studies will shed further light on the incidence and causes of this dynamic aspect of genome architecture.

In an effort to identify and understand the full spectrum of functional elements of the human genome and document the

differences between humans and their closest relatives, efforts are underway to sequence the chimpanzee genome (Fujiyama et al., 2002; Olson and Varki, 2003). A parallel effort to broaden the impact of comparative analyses utilizing the mouse genome sequence involves comparison with the genome sequence of a related rodent model organism, the laboratory rat.

While the understanding of the functional elements of the human genome involves comparison to other species, the choice of species and even which individual within the species has potential conservation relevance. Anticipating the analyses that can become feasible through genomic studies and identifying opportunities for collection of data and their subsequent use in conservation efforts is a process still in its infancy. While all information regarding genetic variation is potentially relevant, the data gain value when associated with knowledge of phenotype, provenance of origin of the sampled individuals within the range of the species, and background genetic and demographic data.

Although expanding, complete genome studies are currently few and the allocation of funding for sequencing projects reflects a parsimonious process for understanding the human genome. With the release of the chimpanzee genome, the first genome sequence from an endangered species will become available. Endangered status does not contribute in a significant way to the sequencing decision process, but it is worthwhile to note that knowledge of the genomes of endangered primates, specifically, and other endangered mammals representing primates and other major mammalian lineages has humanitarian value.

#### **SNP analysis, ancestral character state and patterns of selection in genome evolution**

The comparative analysis of mammalian genomes to evaluate the mode and tempo of changes in exons, intron structure and sequence has provided an outline of some of the potential ways that comparative genomics can provide context for evaluating the human genome and its variation (Eichler and Sankoff, 2003; Navarro and Barton, 2003; Shi et al., 2003; Thomas et al., 2003).

With the generation of partial sequences of the human genome, the importance of evaluating the variation intrinsic in the most simple sequence polymorphisms, the SNPs or single nucleotide polymorphisms, offered to provide tools for effectively achieving one of the primary goals of the Human Genome Project (HGP): identification of the DNA changes that contributed to disease, especially those that were due to heritable risk factors (Collins, 1999) be it from simple (single gene) or more complex (multi-factorial) factors (Meyerson, 2003).

SNP discovery and interpretation of SNP data required methods for obtaining information on frequency of these polymorphisms and determination of the ancestral character state of such variation within the human genome. Re-sequencing of loci and the use of oligonucleotide chips demonstrated that SNPs were only rarely shared between humans, chimpanzees and gorillas and in proportions consistent with the overall aver-

age nucleotide divergence between these species. For the majority of human SNPs, the most common nucleotide at an SNP site was the same base that was present in the chimpanzee and accordingly the newly substituted nucleotide was generally less abundant, consistent with expectations of neutral theory (Hacia et al., 1999; Kennedy et al., 2003).

Targeted sequencing of homologous regions of other species and their comparison to the human sequence provides information about conservation of aspects of gene structure such as exon-intron boundaries, the presence of insertions and deletions (indels) of varying magnitude and allows for the identification of conserved regions that serve as candidates for sequences imbued with functions not involving coding (Britten, 2002; Waterson et al., 2002; Weber et al., 2002; Yu et al., 2002; Frazer et al., 2003; Kirkness et al., 2003).

Through the use of phylogenetic comparison of DNA sequences, presumptive regulatory elements conserved between higher categories of systematic classification, such as families, orders and supraordinal lineages could be identified. The comparison of human, mouse, and dog genome sequences becomes a crucial tool in genome annotation and enriches the understanding of the organization and workings of the human genome (Cooper et al., 2003; Kirkness et al., 2003).

Analysis of genomic variation within a species can reveal evidence of population processes that cannot be uncovered by behavioral observations alone, or would require prohibitively lengthy monitoring (Buchan et al., 2003). Investigating polymorphisms among populations can shed light on the basis of prezygotic and postzygotic isolation and the relative importance of sex-linked characters in these processes, as has been demonstrated for a small sampling of autosomal loci and sex-linked SNPs in *Ficedula* flycatchers (Saetre et al., 2003). SNP data are becoming available as a result of whole genome sequencing projects, resequencing efforts and through strategies designed to collect locus-specific SNP data in a variety of species (Primmer et al., 2002; Navarro and Barton, 2003; Shi et al., 2003; Yu et al., 2003).

Of what utility will these data be for conservation efforts? Construction of SNP-based maps has become a target for domesticated species and many model organisms. SNPs in coding sequences are subject to evaluation by a growing number of methodologies and models (see Fay and Wu, 2003; Luikart et al., 2003 for reviews). Comparative sequence analysis of human and chimpanzee genes partitioned into presumptive functional subsets has led to the suggestion that loci encoding olfactory and nuclear transport functions have undergone accelerated evolution since chimpanzees and humans diverged from a common ancestor (Clark et al., 2003). Sequellae of these substitutions may be physiological adaptations that distinguish the two species.

In lieu of standing aside while large societal investments in genomics are made, conservationists may consider assembling strategies for incorporating the forthcoming effort in evolutionary genomics (Eisen and Fraser, 2003) into an enlarged effort to describe, better understand and conserve biodiversity.

## Phylogenetic footprints and shadows

Genome annotation for a richer diversity of species is becoming available through web browsers ([www.genome.ucsc.edu](http://www.genome.ucsc.edu)). Phylogenetic footprinting (Gumucio et al., 1992) and phylogenetic shadowing (Boffelli et al., 2003) require such in silico data resources. These phylogenetic analyses align homologous stretches of DNA that, through inclusion of representative organisms from different taxonomic levels or representing a range of times of divergence, infer or identify non-coding sequences, including DNA-binding motifs with roles in the regulation of gene activity (Banerjee et al., 2002).

While there is an understandable focus on comparative analysis of primate genomes, outgroups to primates will need to be studied in order to identify those genome changes that are unique to the primate lineage. In order to identify the functional significance of each nucleotide in the human genome, data should be derived from organisms whose common ancestor with humans occurred early in the phylogeny of mammals (and even other vertebrates, e.g., pufferfish). Currently, consideration of the amount of total sequencing data required across the mammalian phylogenetic tree to assemble the required number of nucleotide replacements has led to the suggestion that complete genome sequences from an additional five to eight species of mammals will need to be collected (Cooper et al., 2003), although suggestions for more extensive taxon sampling are also persuasive (O'Brien and Murphy, 2003).

Identification of the functional basis of conserved sequence blocks within mammalian genomes as well as the specializations that have been produced uniquely or in a convergent manner in this vertebrate class is a basic research question in biology. The notable specializations of humans, including the capacity for learning and acquisition of language, and versatility of cultural practices beg for greater understanding. Early searches for novel genes in humans have produced sets of candidate loci for contributing to a deeper understanding of genetic differences between humans and their closest relatives (Gagneux and Varki, 2001). Comparative investigations will be facilitated in new ways with reference to the human as the model system. Indeed, the chimpanzee genome sequencing effort will rely on alignment to the human genome for assembly (Fujiiyama et al., 2002; Olson and Varki, 2003).

Yet, there are other considerations in choosing additional primate and non-primate species for sequencing that address both the imperative for a new vision of medicine utilizing a deepening understanding of the human genome, while engaging in fundamental aspects of intellectual inquiry into the science of life. Losses of biological diversity indicate that, even as we anticipate ongoing technical advances and tools for data analysis, the genetic, organismal and ecosystem diversity that is amenable to such investigations is being lost. These losses are occurring as the result of human causes, direct and indirect (Wilson, 1992), even as the genome pioneers begin to uncover richness of insight from this emerging era of biology. The great apes are all endangered species suffering from fragmentation and loss of habitat, direct predation for bush meat, and disease risks (Walsh et al., 2003).

Conservation genomics continues and extends the consideration of evolutionary structuring of the gene pools of endangered species as part of the effort to conserve biological diversity. The concept of evolutionary significant units (ESUs) was originally brought forward to incorporate a variety of data into conservation management decision making, including but not restricted to genetic data (Ryder, 1986; Crandall et al., 2000). What is the range of genetic changes that are involved in subspecies differences and how do these reflect processes that influence both *ex situ* and *in situ* conservation options? How, armed with knowledge of the substance and variety of genetic differences contributing to the adaptive differentiation within populations of a species and between species, would efforts to conserve biological diversity and the evolutionary potential of biotic communities and ecosystems be modified by this information? Comparative studies examining levels and tissue specificity of gene expression will ultimately provide a basis for interpretation of specific phenotypic differences relevant to local adaptation and provide insights linking life history strategies with gene expression studies (Enard et al., 2002a, b; Gibson, 2002; Karaman et al., 2003).

### Primate adaptations and human origins

The special adaptations that contribute to human uniqueness, including complex verbal language and communication skills, diversity of cultural developments, cognitive skills, etc, require consideration from an evolutionary perspective. Primates have notable adaptations, some of which served to prepare for the emergence of the human species (Table 1). The evolution of primate adaptations facilitated the origin of humans but the appearance of humans was not an inevitable consequence of biotic evolution on earth. The understanding of human origins will be facilitated by the purposeful investigation of pertinent genomic, neurophysiological, morphological and other changes that have taken place along the primate lineage. Studies of gene expression differences between human and other primate species involving tissue-specific comparisons of brain (Enard et al., 2002a; Caceres et al., 2003) and models such as fibroblasts (Karaman et al., 2003) serve to provide candidate gene loci for the genetic basis of adaptive modifications in hominid evolution. Studies of other primates including wild populations offer definition and clarification of the context of traits that contribute to the human phenomenon. Parental care by male baboons for their own offspring in preference over young baboons not their offspring, could only be evidenced through genetic analyses (Buchan et al., 2003). The further unraveling of this phenomenon and the details of the system of preferences resulting in choice of mates, will undoubtedly involve contributions from larger scale genome studies.

Duplications associated with genome regions of recurrent translocations may foster the appearance of novel gene sequences (Chai et al., 2003; Nahon, 2003). Establishment of novel physiological capabilities through gain of function involving divergence of duplicated genes is becoming more richly exemplified. A prominent example includes genes involved in digestive adaptations associated with folivory in primates

**Table 1.** Selection and adaptations amenable to investigation with comparative genomic approaches

Some mammalian adaptations	
varied natural history	mass, generation time, cerebration
locomotory adaptations	flight, swimming, skeletal loading
sensory adaptations	olfaction, echolocation, color vision, transduction
dietary adaptations	myrmecophagy, gut fermentation, dentition, cranial morphology
behavioral adaptations	parental investment, sociality

(Zhang et al., 2002). The RNase1 gene duplication in the leaf-eating douc langur is an important example in the conservation context because of the endangerment of leaf-eating primates in general, and this species in particular (IUCN, 2001). The convergent evolution of gut lysozyme in ruminant artiodactyls and colobine primates is a comparable example (Stewart et al., 1987; Messier et al., 1997). In zoos, insufficient success in captive propagation has been due, at least in part, to gastric pathology, potentially associated with an inappropriate dietary regimen that may constitute a significant factor contributing to lack of population growth (Ensley et al., 1982).

Loss of function of duplicated loci identifies relaxation of selective constraint, as in the differential fate of olfactory receptor gene families in rodents, carnivores and primates (Young and Trask, 2002; Whinnett and Mundy, 2003; Kirkness et al., 2003). Though inactive, vestigial gene sequences and pseudogenes may still influence organismal fitness (Giglio et al., 2001; Stankiewicz and Lupski, 2002).

Genes and regulatory pathways that are involved in reproductive isolation may also be identified by genomic approaches. Evaluation of selective forces and identification of genome segments that are resistant to modification by recombination or conversion as the result of inversion, for example, provides approaches for identifying candidate loci involved in the speciation process crucial to increasing biodiversity (Rieseberg, 2001; Hey, 2003; Navarro and Barton, 2003; Wang et al., 2003).

There is fundamental and practical interest in understanding the underlying genetic mechanisms that result in the morphological changes that differentiate species. Developing a framework for interpreting genetic differences that contribute to morphological changes portends a new descriptive metric to add to the historical framework of comparative anatomy and morphological systematics.

Obesity and cardiovascular disease are among the significant medical problems of gorillas in captivity. Cholesterol/HDL ratios increase with age and are associated with incidence of cardiovascular disease in gorillas (Baichman et al., 2003), despite the vegetarian diet of these apes. Gorillas may develop cardiovascular symptoms that, in humans, are associated with morbidity and mortality, but which may favorably be influenced by management of blood lipid components. Investigations of genetic factors that impact lipid metabolism in humans may both gain benefit from an evolutionary perspective and simultaneously provide a model for investigating the physiology of related species.

**Table 2.** Opportunities in primate genomics relevant to conservation

Impact category	Investigation	Specific studies
immune function and regulation	organization and allelic diversity of MHC and other loci affecting immune function	pathogen resistance; co-evolution with parasites/pathogens
genetic risk factors in small populations	characterization candidate gene studies mapping studies	albinism; syndactyly; trisomy and other aneuploidies diabetes, cardiovascular disease SNP development; haplotype analysis; linkage and LD analysis
understanding specializations	dietary tolerances, e.g., tannins, foregut fermentation	digestive enzymes, e.g., RnaseI; binding proteins; liver microsomal enzymes
reproduction	identification and characterization of genes regulating reproduction, e.g., spermatogenesis and ovulation, placentation	loci influenced by light cycle, olfactory and auditory stimuli; imprinted genes, MHC impacts on mate choice
gene pool conservation	phylogeography; evolutionary patterns of migration and dispersal; demographic history	mitochondrial, autosomal and Y-chromosome variation
parental care and behavior	parentage analyses; mate choice associations; haplotype structure	SSR analyses; MHC structure and variation; genome-wide scans for LD

### Conservation genomics: examples and opportunities

For most endangered species, ignorance of basic aspects of their biology impedes conservation efforts. Application of genomics information can influence species survival through improvements in wildlife health and by structuring assessments of and intervention for population viability and recovery. Genome comparisons may provide otherwise unobtainable clues to physiology and medicine for some endangered species (Table 2).

Genomic studies undertaken in other organisms required development of associated resources in order to produce genetic maps, annotate genomes, and assist with alignments and assembly. DNA from well-characterized individuals, including cell lines and large insert libraries, in addition to a source of blood or other tissue for shotgun sequencing efforts is required. For endangered species, and especially those with small body mass and therefore unable to provide large amounts of DNA from blood samples, this may pose tactical difficulties and require careful consideration for appropriate access to specimens. Population samples, from both captive and wild populations, if possible, provide enhanced conservation benefits through comparative analyses. Demographic, dietary and disease information for individuals whose genomes may be sequenced add value as additional individuals and/or populations are studied. Samples from families (such as have been provided by CEPH for humans) will be generally useful for linkage studies and, from kindreds segregating phenotypes of interest, will provide an essential tool for identifying genetic disease loci in small populations of endangered species, e.g., California condor (Ralls et al., 2000).

While advances in genetic medicine of domestic animals have occurred, in part because of their value as models for human disease, genetic approaches to understanding disease susceptibility, the emergence of genetic diseases consequent to inbreeding in small populations (Ryder, 1988), as well as the broader aspects of genetic risk factors associated with the dynamic architecture of mammalian genomes (Stankiewicz

and Lupski, 2002) are amenable to analysis and interpretation through comparative genomic studies.

For endangered species, especially those with no close domestic relatives, opportunities for invasive studies of endocrinology, immune function, physiological chemistry, etc. are limited. Comparative genomic studies offer potential new ways to unravel the significance of environmental factors that may contribute to the health and growth of small populations. Zoo and wildlife medicine have contributed substantively to identifying risk factors and significant biomedical issues that affect populations in captivity and in the wild, e.g. impact of Ebola virus on in situ populations of African apes (Leroy et al., 2004). Further advances in zoo and wildlife medicine will play an increasingly crucial role in conservation efforts. Numerous studies have helped identify idiopathic conditions for which comparative genomic studies may help identify causal factors (Table 3).

### Whole genome sequencing projects and conservation genomics

It may seem that genome sequencing projects are too expensive to be justified for species solely because they are endangered. While this may be true at present, the cost of genome sequencing has been decreasing substantially and if predictions of future advances in DNA sequencing technology are realized, the costs may become less restrictive. For this reason alone, some discussion at this time regarding benefits and priorities is appropriate.

The importance of comparative genome sequencing efforts focuses attention on the rationale for choosing genomes for sequencing studies. O'Brien and colleagues (2001) suggested a rationale for consideration of priorities for genome sequencing projects for placental mammals that considered phylogenetic exigencies for identifying lineage-specific genome changes within the panoply of mammalian genomic architectures. Their rationale also considered the importance of species that serve



**Table 3.** Potential studies facilitated by genome sequence information

Species	Research activity
cheetah	risk factors for infectious peritonitis and cerebral spongiform disease; immune system gene organization and variation
black rhinoceros	viral diseases; adaptations in oxidative phosphorylation pathways; olfactory mediation of social behavior; susceptibility to trypanosomiasis
lowland and mountain gorilla	immune system gene organization and variation; mate choice and MHC variation; susceptibility to pathogens of human origin (measles, helminths); consanguinity estimation
bonobo	MHC structure and variation, cardiovascular disease risk factors, susceptibility to dematases; papiloma virus infection
ruffed lemur	hemosiderosis and hemochromatosis
giant panda	olfactory mediation of sexual behavior; twinning; delayed implantation; fetal loss
bighorn sheep	susceptibility to pathogens commensal with domestic sheep

as medical models (e.g., dog, cat, rabbit, rat), have conserved (insectivore-like, e.g., tenrec, tree shrew) or highly derived morphological features (whale, elephant, horse, bat) or whose karyotype is highly derived (dog) or considered to retain ancestral features (whale, bat). Interestingly, several of the species that arose in their analysis were endangered species: the elephant, chimpanzee, and orangutan. Endangered species fall within every mammalian order, with the exception of the Tubulidentata, of which the aardvark is the single extant species.

In the current era of genome exploration, an overriding consideration has been the elucidation of knowledge applicable to the management of human health and disease. Currently, relatively little consideration of the health of other species nor of biotic communities and ecosystems is evident. Some efforts now to reconcile these views and broaden the scope of the effort launched by the Human Genome Project will benefit human society, present and future, through both an increased understanding of the human organism in its evolutionary context, and through the preservation of that context itself.

A useful initial step involves consideration of the allocation of current genome sequencing efforts. As the choices are considered for selection of taxa for whole genome sequencing projects, selection of existing experimental model organisms will not provide the desired phylogenetic diversity and complement of nucleotide replacements that will allow the functional interpretation of site-specific changes in the human genome (Sidow, 2002). The complete genome sequencing of organisms which are not traditional objects of research in medicine and physiology is an emergent phenomenon.

A conservation-oriented rationale for allocation of mammalian genome sequencing efforts has much in common with a rationale developed to elucidate the functional elements of the human genome. Appropriate phylogenetic diversity and the extent of accumulated substitutions must be incorporated. For each candidate taxon for genome sequencing, consideration may be given to whether a species with direct conservation application might be substituted or included at a low level of sequencing effort. 1–2× sequences may align sufficiently well to provide a level of coverage that will immediately facilitate further studies (Cooper et al., 2003), as has been argued for the 1.5× genome sequence of the domestic dog (Kirkness et al., 2003).

While the alternate choice of organisms on the same evolutionary lineage may have relatively little impact on the replace-

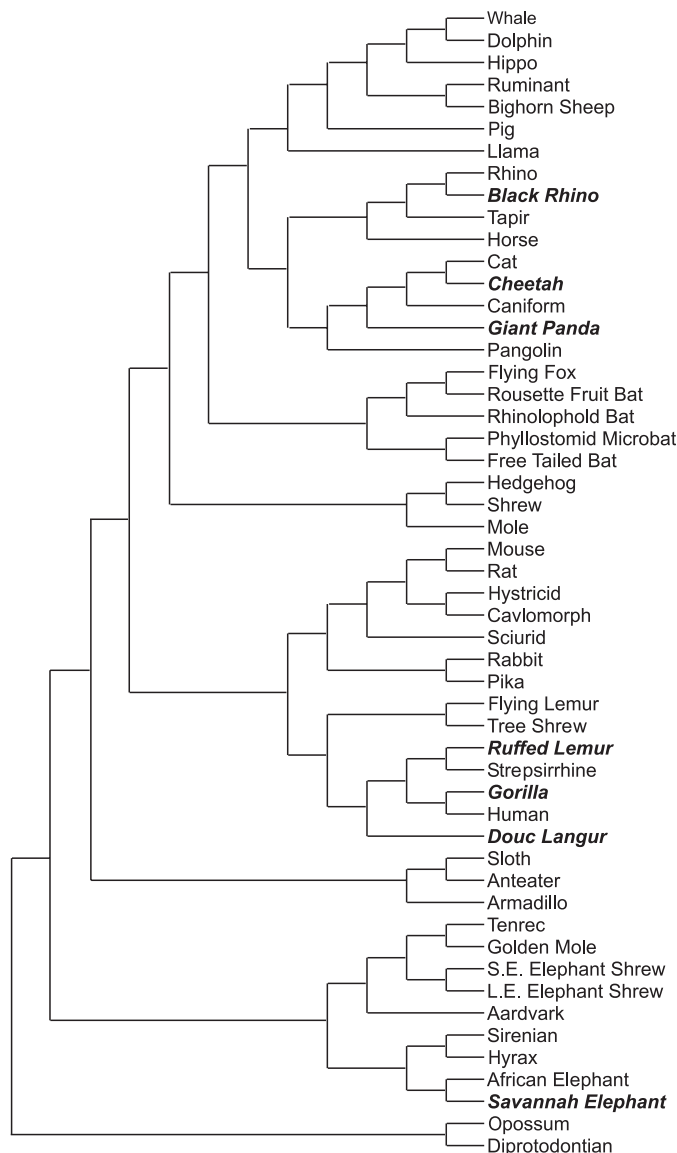
ment tree data set, the impact of adding genome sequence information for understanding the natural history, evolutionary genetics, ecology and health of a species will change the context for its conservation and the prospects for constructive intervention for survival.

Considering the phylogeny of eutherian mammals, from which species most of the next set of high priority species for whole genome sequencing will be identified, there are several endangered species that could provide the utilitarian comparative data for increasing the number of mutational differences in the compiled tree (Table 3).

Potential candidate species may be placed in an evolutionary context with other candidate eutherian taxa (O'Brien et al., 2001) and include an appropriate marsupial outgroup (Wakefield and Graves, 2003) (Fig. 1).

Opportunities for genome sequence information to shed light on organismal fitness fall into some broad general categories such as sensory adaptations, physiology and disease, and examination of the nature and extent of genetic diversity for analysis of demographic history, inference of gene flow, evaluation effective population size and for evaluating hypotheses in behavioral ecology (Table 2).

With relatively few mammalian genomes substantially or completely sequenced, evolutionary comparisons are still in early stages. Organization of repetitive sequences, including highly-repetitious simple sequence DNA, moderately repetitive elements, and transposable elements participate significantly in identified aspects of genome evolution (e.g. Kirkness et al., 2003; Thomas et al., 2003), but rather more limited data exists regarding variation within clades. Similarly, small insertion/deletion events have been shown to contribute to the divergence of genomes (Britten, 2002; Frazer et al., 2003), but the rate and extent of these changes across mammalian phylogeny is poorly understood. This current state-of-the-art suggests that rather complete sequencing of additional mammalian genomes will be required in order to shed light on these aspects of genome evolution and their functional relevance (Cooper et al., 2003). It is a comfortable prediction that many new insights will be gained through this process. Ultimately, suites of genes and regulatory regions may be identified, the screening of which can extract significant amounts of useful information relative to understanding physiology and evolutionary adaptations in different lineages. However, at the present time, this approach would not be expected to yield the same information that complete genome sequencing efforts can amass. Most like-



**Fig. 1.** Molecular phylogeny of eutherian mammals with inclusion of endangered species suitable as candidates for genome studies. The endangered species are in boldface.

ly, comparative studies of genes involved in the diversity of mammalian adaptations (e.g. reproduction, immunity and pathogen resistance, sensory adaptations, etc.) will be studied even as additional efforts in comparative genome sequencing provide information to help refine the understanding of the involved gene complexes and their regulatory networks.

### Conservation bioinformatics

A bioinformatics capability commensurate with amount of available data is necessary to productively utilize the sequencing information generated and will require customized tools for

many conservation applications. For human genetic data, confidentiality and patient anonymity are required. However, the very opposite holds for studies in other species for which documentation of individual genotypes and allele frequencies in populations facilitates rigorous study design, interpretation of data and also contributes to a more complete knowledge of the population genetics of the species being sampled. Such data have obvious relevance for conservation efforts through insights provided into effective population size (Hayes et al., 2003; Roman and Palumbi, 2003), estimates of gene flow, inferences of population structure and as they reflect behavior of populations (Altmann et al., 1996).

In many earlier comparative genetic studies of humans and other primates, and still in recent work, the source of the studied materials is incompletely documented. The institution providing the sample may be acknowledged, but individual identification is often not provided, resulting in several undesirable consequences. Findings may not be able to be placed in a genealogical or population perspective. Variation may be overestimated or underestimated because kinship (or identity) is not possible to reference. Potential polymorphisms and their origin may be tested in pedigree studies, but without linking the individual data to a standard identifier, such as a studbook number or ISIS identifier ([www.isis.org](http://www.isis.org)), important data are being lost. The same animal may be repeatedly investigated, even in the same study without recognition. Such errors can be compounded in the literature further exacerbating the situation. Opportunities to build databases with individual genotypes and/or haplotypes are lost.

The launch by the U.S. National Science Foundation of the Integrated Non-human Primate Biomaterials and Information Resource (IPBIR) represents a substantial effort to assemble, characterize, and distribute high-quality DNA samples of known provenance with accompanying demographic, geographic, and behavioral information in order to stimulate and facilitate research in primate genetic diversity and evolution, comparative genomics, and population genetics ([www.ipbir.org](http://www.ipbir.org)). Complementary efforts are also underway within the European community, with the initiation of INPRIMAT ([www.inprimate.org](http://www.inprimate.org)).

Expanded biomaterials banking efforts for a wider range of taxa than is now underway are urgently needed to facilitate comparative studies and to make the best use of genomic sequence information through providing resource curation and integrating associated genealogical and demographic data.

### Conservation relevance and ethical concerns

The genetic refugium of many endangered species is contained in a small number of individuals, typically with unknown kinship and uncertain demographic characterization. Managing population growth and retention of genetic variation are central features of captive management such as that undertaken in zoos and free-ranging populations of endangered species for which recovery is actively being attempted (Foose et al., 1995; Ballou and Foose, 1996). Comparative mt genomic studies are now providing insights into demographic histories of

species that may be incorporated into conservation analyses (Caccone et al., 2002). The contribution of disease outbreaks as a demographic event is also being identified for known pathogens such as malaria (Wang et al., 2003). Demographic inferences are also providing the basis for postulating disease outbreaks in the history of some species of mammals, such as cheetahs.

Species must have sufficient suitable habitat. They must have ecosystem resources that allow for population persistence over extended periods. Normally, these periods are in evolutionary time scales, e.g., hundreds to hundreds of thousands of generations. Currently, concern for many species is focused on the next several generations as the trends for their populations in the face of human-caused endangerment forebode their critical endangerment and extinction. Genome sequencing studies cannot be expected to impact directly habitat sufficiency for endangered species. However, the ability to more effectively evaluate demographic history and genetic variation in populations through genomic analyses may find conservation application, increase the effectiveness of residual habitat fragments for population persistence, and provide more quantitative evaluation of levels of threat or progress to recovery than are currently available.

The humanitarian value of studies of comparative genomics of mammals, generally, and primates, specifically, stands to become more fully realized. The fact that, according to current estimates, approximately one-fourth of mammal species face levels of extinction risk leading to inclusion in the Red List of Threatened Species (IUCN, 2001) may greatly impair the feasibility, and ethical pursuit of such studies in the future.

Currently, whole genome studies are available for only a few species and until the chimpanzee (*Pan troglodytes*) genome is

completed, none will represent endangered species. Although it is estimated that nearly one-third of all mammal species will face a level of extinction threat resulting in listing as endangered by recognized bodies such as the IUCN (www.iucn.org), gathering knowledge of the genomes of these species is not yet a priority.

For primates, 118 extant species or approximately one-third of the 628 recognized species are considered endangered or critically endangered, including all the species of great apes (IUCN, 2001). Even as new species of primates are being identified, catastrophic declines are currently being noted for African apes as a result of predation and in association with zoonotic disease outbreaks (Walsh et al., 2003; Leroy, et al., 2004).

One outcome of comparative genomics generally, and primate genomics specifically, will be to assess at the genetic level responses of living systems to environmental change. New insights into the origins of humans, their traits and culture and advances in human medical care are anticipated as genomic technologies advance and are applied to a wider diversity of species. Genome biologists can promote the utilization of genomic analysis tools to contribute to the welfare and persistence of animal populations and, particularly, endangered species, recognizing that even as technical capabilities grow, biodiversity is being lost.

## Acknowledgements

I thank many colleagues for helpful discussions, including William Murphy, Stephen O'Brien, Kurt Benirschke, Maryellen Ruvolo, Suellen Charter, Leona Chemnick and Marlys Houck. Susan Hansen and Elisabeth Baum helped with the preparation of the manuscript.

## References

- Altmann J, Alberts SC, Haines SA, Dubach J, Muruthi P, Coote T, Geffen E, Cheesman DJ, Mututua RS, Saiyalel SN, Wayne RK, Lacy RC, Bruford MW: Behavior predicts genetic structure in a wild primate group. *Proc Natl Acad Sci, USA* 93:5797–5801 (1996).
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE: Recent segmental duplications in the human genome. *Science* 297:1003–1007 (2002).
- Baitechman EJ, Goldberger MH, Nanna M, Calle PP, Clippinger TL, Deem SL, James SB, Raphael BL, Cook RA: Evaluation of blood lipid profiles as part of a complete cardiac assessment in western lowland gorillas (*Gorilla gorilla gorilla*). *Proc Am Assoc Zoo Vets* 72–74 (2003).
- Ballou JD, Foose TJ: Demographic and genetic management of captive populations, in Kleiman DG, Allen ME, Thompson KV, Lumpkin S (eds): *Wild Mammals in Captivity: Principles and Techniques*, pp 263–283 (University of Chicago Press, Chicago 1996).
- Banerjee P, Bahlo M, Schwartz JR, Loots GG, Houston KA, Dubchak I, Speed TP, Rubin EM: SNPs in putative regulatory regions identified by human mouse comparative sequencing and transcription factor binding site data. *Mamm Genome* 13:554–557 (2002).
- Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, Rubin EM: Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* 299:1391–1394 (2003).
- Britten RJ: Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc Natl Acad Sci, USA* 99:13633–13635 (2002).
- Buchan JC, Alberts SC, Silk JB, Altmann J: True paternal care in a multi-male primate society. *Nature* 425:179–181 (2003).
- Caccone A, Gentile G, Gibbs JP, Fritts TH, Snell HL, Betts J, Powell JR: Phylogeography and history of giant Galapagos tortoises. *Evolution* 56:2052–2066 (2002).
- Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, Lockhart DJ, Preuss TM, Barlow C: Elevated gene expression levels distinguish human from non-human primate brains. *Proc Natl Acad Sci, USA* 100:13030–13035 (2003).
- Cann RL, Stoneking M, Wilson AC: Mitochondrial DNA and human evolution. *Nature* 325:31–36 (1987).
- Chai JH, Locke DP, Greally JM, Knoll JH, Ohta T, Dunai J, Yavor A, Eichler EE, Nicholls RD: Identification of four highly conserved genes between breakpoint hotspots BP1 and BP2 of the Prader-Willi/Angelman syndromes deletion region that have undergone evolutionary transposition mediated by flanking duplicons. *Am J Hum Genet* 73:898–925 (2003).
- Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferriera S, Wang G, Zheng X, White TJ, Sninsky JJ, Adams MD, Cargill M: Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* 302:1876–1877 (2003).
- Collins FS: Shattuck Lecture: Medical care and societal consequences of the Human Genome Project. *New Engl J Med* 341:28–37 (1999).
- Cooper GM, Brudno M, NISC Comparative Sequencing Program, Green ED, Batzoglu S, Sidow A: Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Res* 13:813–820 (2003).
- Crandall KA, Bininda-Emonds OR, Mace GM, Wayne RK: Considering evolutionary processes in conservation biology. *Trends Ecol Evol* 15:290–295 (2000).



- Davidson EH, McClay DR, Hood L: Gene networks and the properties of the developmental process. *Proc Natl Acad Sci, USA* 100:1475–1480 (2003).
- Deinard AS, Kidd K: Identifying conservation units within captive chimpanzee populations. *Am J Phys Anthropol* 111:25–44 (2000).
- Eichler EE, Sankoff D: Structural dynamics of eukaryotic chromosome evolution. *Science* 301:793–797 (2003).
- Eisen JA, Fraser CM: Phylogenomics: intersection of evolution and genomics. *Science* 300:1706–1707 (2003).
- Enard W, Khaitovich P, Klose J, Zöllner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, Pääbo S: Intra- and interspecific variation in primate gene expression patterns. *Science* 296:340–343 (2002a).
- Enard W, Przeworski M, Fisher SE, Lai CSL, Wiebe V, Kitano T, Monaco AP, Pääbo S: Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418:869–872 (2002b).
- Ensley PK, Rost TL, Anderson M, Benirschke K, Brockman D, Ullrey D: Intestinal obstruction and perforation caused by undigested *Acacia* sp leaves in langur monkeys. *J Am Vet Med Assoc* 181:1351–1354 (1982).
- Fay JC, Wu CI: Sequence divergence, functional constraint, and selection in protein evolution. *Ann Rev Genomics Hum Genet* 4:213–235 (2003).
- Ferris SD, Brown WM, Davidson WS, Wilson AC: Extensive polymorphism in the mitochondrial DNA of apes. *Proc Natl Acad Sci, USA* 78:6319–6323 (1981).
- Foose TJ, De Boer L, Seal US, Lande R: Conservation management strategies based on viable populations, in Ballou JD, Gilpin M, Foose TJ (eds): *Methods and Cases in Conservation Science; Population Management for Survival and Recovery: Analytical Methods and Strategies in Small Population Conservation*, pp 273–294 (Columbia University Press, New York 1995).
- Frazer KA, Chen X, Hinds DA, Pant PV, Patil N, Cox DR: Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Res* 13:341–346 (2003).
- Fujiyama A, Watanabe H, Toyoda A, Taylor TD, Itoh T, Tsai SF, Park HS, Yaspo ML, Lehrach H, Chen Z, Fu G, Saitou N, Osoegawa K, de Jong PJ, Suto Y, Hattori M, Sakaki Y: Construction and analysis of a human-chimpanzee comparative clone map. *Science* 295:131–134 (2002).
- Gagneux P, Varki A: Genetic differences between humans and great apes. *Mol Phylogenet Evol* 18:2–13 (2001).
- Gagneux P, Wills C, Gerloff U, Tautz D, Morin PA, Boesch C, Fruth B, Hohmann G, Ryder OA, Woodruff DS: Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc Natl Acad Sci, USA* 96:5077–5082 (1999).
- Garner KJ, Ryder OA: Mitochondrial DNA diversity in gorillas. *Mol Phylogenet Evol* 6:39–48 (1996).
- Gibson G: Microarrays in ecology and evolution: a preview. *Mol Ecol* 11:17–24 (2002).
- Giglio S, Broman KW, Matsumoto N, Calvari V, Gimmelli G, Neumann T, Ohashi H, Voullaire L, Larizza D, Giorda R, Weber JL, Ledbetter DH, Zuffardi O: Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. *Am J Hum Genet* 68:874–883 (2001).
- Gumucio DL, Heilstedt-Williamson H, Gray TA, Tarle SA, Shelton DA, Tagle DA, Slightom JL, Goodman M, Collins FS: Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes. *Mol Cell Biol* 12:4919–4929 (1992).
- Hacia JG, Fan JB, Ryder OA, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SPA, Collins FS: Determination of ancestral alleles for human single nucleotide polymorphisms using high density oligonucleotide arrays. *Nat Genet* 22:164–167 (1999).
- Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL: Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 18:1189–1203 (2001).
- Hardison RC, Roskin KM, Yang S, Diekhans M, Kent WJ, et al: Co-variation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res* 13:13–26 (2003).
- Hayes BJ, Visscher PM, McPartlan HC, Goddard ME: Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res* 13:635–643 (2003).
- Hellmann I, Zollner S, Enard W, Ebersberger I, Nickel B, Pääbo S: Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res* 13:831–837 (2003).
- Hey J: Speciation and inversions: chimps and humans. *Bioessays* 25:825–828 (2003).
- IUCN. The 2001 Red List of Endangered Species. International Union for Conservation of Nature and Natural Resources. Gland, Switzerland (2001).
- Karaman MW, Houck ML, Chemnick LG, Nagpal S, Chawannakul D, Sudano D, Pike BL, Ho VV, Ryder OA, Hacia JG: Comparative analysis of gene-expression patterns in human and African great ape cell lines. *Genome Res* 13:1619–1630 (2003).
- Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, et al: Large-scale genotyping of complex DNA. *Nat Biotechnol* 21:1233–1237 (2003).
- Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D: Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci, USA* 100:11484–11489 (2003).
- Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC: The dog genome: survey sequencing and comparative analysis. *Science* 301:1898–1903 (2003).
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al: Initial sequencing and analysis of the human genome. *Nature* 409:860–921 (2001).
- Leroy EM, Rouquet P, Formenty P, Souquiere S, Kilbourne A, Froment JM, Bermejo M, et al: Multiple Ebola virus transmission events and rapid decline of central African wildlife. *Science* 303:387–390 (2004).
- Locke DP, Archidiacono N, Miscow D, Cardone MF, Deschamps S, Roe B, Rocchi M, Eichler EE: Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. *Genome Biol* 4:R50 (2003).
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P: The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* 4:981–994 (2003).
- Marth G, Schuler G, Yeh R, Davenport R, Agarwala R, et al: Sequence variations in the public human genome data reflect a bottlenecked population history. *Proc Natl Acad Sci, USA* 100:376–381 (2003).
- Messier W, Stewart CB: Episodic adaptive evolution of primate lysozymes. *Nature* 385:151–154 (1997).
- Meyerson M: Human genetic variation and disease. *Lancet* 362:259–260 (2003).
- Morin PA, Moore JJ, Woodruff DS: Identification of chimpanzee subspecies with DNA from hair and allele-specific probes. *Proc R Soc Lond B Biol Sci* 249:293–297 (1992).
- Nahon JL: Birth of “human-specific” genes during primate evolution. *Genetica* 118:193–208 (2003).
- Navarro A, Barton NH: Chromosomal speciation and molecular divergence – accelerated evolution in rearranged chromosomes. *Science* 300:321–324 (2003).
- Nobrega MA, Ovcharenko I, Afzal V, Rubin EM: Scanning human gene deserts for long-range enhancers. *Science* 302:413 (2003).
- O'Brien SJ, Murphy WJ: Genomics. A dog's breakfast? *Science* 26:1854–1855 (2003).
- O'Brien SJ, Eizirik E, Murphy WJ: Genomics. On choosing mammalian genomes for sequencing. *Science* 292:2264–2266 (2001).
- Olson MV, Varki A: Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat Rev Genet* 4:20–28 (2003).
- Pennacchio LA: Insights from human/mouse genome comparisons. *Mamm Genome* 14:429–436 (2003).
- Pevzner P, Tesler G: Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proc Natl Acad Sci, USA* 100:7672–7677 (2003).
- Primmer CR, Borge T, Lindell J, Saetre G-P: Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Mol Ecol* 11:603–612 (2002).
- Ralls K, Ballou JD, Rideout BA, Frankham R: Genetic management of chondrodystrophy in California condors. *Animal Conservation* 3:145–153 (2000).
- Rieseberg LH: Chromosomal rearrangements and speciation. *Trends Ecol Evol* 16:351–358 (2001).
- Roman J, Palumbi SR: Whales before whaling in the North Atlantic. *Science* 301:508–510 (2003).
- Ryder OA: Systematics and conservation: The dilemma of subspecies. *Trends Ecol Evol* 1:8 (1986).
- Ryder OA: Founder effects and endangered species. *Nature* 331:396 (1988).
- Ryder OA, Chemnick LG: Chromosomal and mitochondrial DNA variation in orangutans. *J Hered* 84:405–409 (1993).
- Saetre GP, Borge T, Lindroos K, Haavie J, Sheldon BC, Primmer C, Syvanen AC: Sex chromosome evolution and speciation in *Ficedula* flycatchers. *Proc R Soc Lond B Biol Sci* 270:53–59 (2003).
- Shi J, Xi H, Wang Y, Zhang C, Jiang Z, et al: Divergence of the genes on human chromosome 21 between human and other hominoids and variation of substitution rates among transcription units. *Proc Natl Acad Sci, USA* 100:8331–8336 (2003).
- Sidow S: Sequences first. Ask questions later. *Cell* 111:13–16 (2002).
- Stankiewicz P, Lupski JR: Molecular-evolutionary mechanisms for genomic disorders. *Curr Opin Genetics Dev* 12:312–319 (2002).
- Stankiewicz P, Park SS, Inoue K, Lupski JR: The evolutionary chromosome translocation 4;19 in *Gorilla gorilla* is associated with microduplication of the chromosome fragment syntenic to sequences surrounding the human proximal CMT1A-REP. *Genome Res* 11:1205–1210 (2001).
- Stewart CB, Schilling JW, Wilson AC: Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature* 330:401–404 (1987).
- Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, et al: Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 424:788–793 (2003).
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al: The sequence of the human genome. *Science* 291:1304–1351 (2001).

- Versteeg R, van Schaik BDC, van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ, van Kampen AHC: The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res* 13: 1998–2004 (2003).
- Wakefield MJ, Graves JA: The kangaroo genome. Leaps and bounds in comparative genomics. *EMBO R* 4:143–147 (2003).
- Walsh PD, Abernethy KA, Bermejo M, Beyers R, De Wachter P, et al: Catastrophic ape decline in western equatorial Africa. *Nature* 422:611–614 (2003).
- Wang HY, Tang H, Shen CK, Wu CI: Rapidly evolving genes in human I. The glycoporphins and their possible role in evading malaria parasites. *Mol Biol Evol* 20:1795–1804 (2003).
- Waterston RH, Lindblad-Toh K, Bienwyl E, Rogers J, Abril JF, et al: Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562 (2002).
- Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G: Human diallelic insertion/deletion polymorphisms. *Am J Hum Genet* 71:854–862 (2002).
- Whinnett A, Mundy NI: Isolation of novel olfactory receptor genes in marmosets (*Callithrix*): insights into pseudogene formation and evidence for functional degeneracy in non-human primates. *Gene* 304:87–96 (2003).
- Wilson EO: *The diversity of life*; Edward O. Wilson. Published Cambridge, MA (Belknap Press of Harvard University Press, Boston 1992).
- Young JM, Trask BJ: The sense of smell: genomics of vertebrate odorant receptors. *Hum Mol Genet* 11:1153–1160 (2002).
- Yu N, Chen F-C, Ota S, Jorde LB, Pamilo P, Patthy L, Ramsay M, Jenkins T, Shyue S-K, Li W-H: Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* 161:269–274 (2002).
- Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder OA, Kidd KK, Li WH: Low nucleotide diversity in chimpanzees and bonobos. *Genetics* 164:1511–1518 (2003).
- Zhang J, Zhang YP: Pseudogenization of the tumor-growth promoter angiogenin in a leaf-eating monkey. *Gene* 308:95–101 (2003).
- Zhang J, Zhang YP, Rosenberg HF: Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet* 30:411–415 (2002).
- Zhi L, Karesh WB, Janczewski DN, Frazier-Taylor H, Sajuthi D, Gombek F, Andau M, Martenson JS, O'Brien SJ: Genomic differentiation among natural populations of orang-utan (*Pongo pygmaeus*). *Curr Biol* 6:1326–1336 (1996).