The new field of genomics, studying the architecture of the human and other species' genomes, has been driven by the technology and promise of the Human Genome Project. No field of biology is unaffected by the biotechnology revolution and the majority of biologists' thinking has been influenced by genome analysis. The principal goals of the Human Genome Project, to produce a genetic map, physical maps and the complete DNA sequence of human chromosomes, should realistically be accomplished by the original projection date of 2005 (Refs 1–5). Upon completion, a comprehensive human gene map would have two general uses: first as a resource for locating the genetic blueprints for all inheritable characteristics, behaviors and phenotypes; second as a template for resolving the evolutionary heritage of the human species.

Full achievement of both these goals depends in no small part upon the parallel implementation of gene mapping projects in other species. Valuable inference emerges not only from analysis of nominated model species for genomics, such as mouse, Caenorhabditis elegans, Drosophila melanogaster and yeast, but also from precise comparisons of genome organization of other mammals and vertebrates. The rationale for these model species, particularly in revealing the mechanisms of gene action and development, is well described, while the value of mapping additional mammals is not always apparent. In this review, we highlight the applications for comparative gene mapping in mammals and sketch the approaches taken to assemble comparative genetic maps, both as a resource for gene localization and as an evolutionary element. To illustrate the progress, we describe genome inference drawn from our experience with a feline genome project, although the same principles apply with other mammal or vertebrate mapping projects such as those of cattle, sheep, pig, deer and chicken.

Why comparative genomics?

The construction of high resolution gene-dense maps for two mammalian species, human and mouse, has already occurred. The human gene map includes nearly 6000 identified mapped genes plus over 16 000 expressed sequence tags (ESTs) placed in linear order on 23 autosomes by linkage mapping, radiation hybrids, or both.1,5,14 The mouse map contains some 7000 genes plus over 7000 simple sequence length polymorphisms (SSLPs) placed on linkage maps largely through interspecies backcrosses.6,7,23,24 Although gene mapping projects have grown in several other mammalian species, particularly domestic farm and companion animals, the gene density is not likely to reach that of the mouse or human maps for some time because of fiscal constraints. Nonetheless, moderate resolution gene maps of domesticated species are valuable for locating specific phenotypes important for the species and they can be aligned using comparative anchor loci to more dense maps of other species, effectively connecting mapping information between mammalian orders. Thirty years ago, Drosophila geneticists located phenotypes to chromosomal sites and then consulted the 'Red Book' of mapped Drosophila mutations to inspect scores of genes in a region that might encode the variant. Computer databases of the high resolution human and mouse maps now provide the same function for mammalian species whereby conserved chromosomal segments are connected by homologous anchor loci.22,26,27

There are good reasons why comparative gene mapping of vertebrate species is valuable.

• Domestic animal species are a source of thousands of hereditary diseases that are explicit analogs to human hereditary defects.26,27 Their identification by veterinarians provides not only physiological detail about pathogenesis but also laboratory models for testing therapeutic agents.

• The maps benefit domestic animal species by offering proven human therapies for genetically homologous diseases, plus allowing the genetic tracking of economically valuable traits (such as fertility, weight, faster racehorses and many others).14,15,17,19,20

• Animal genomes hold undiscovered adaptive solutions to many incurable human diseases that have been perpetuated in modern species through natural selection.28 Consider that over 100 000 mammalian species have occurred since the origin of mammals in the Mesozoic era (some 100–150 million years ago), but fewer than 5000 exist today. Most animal species have encountered the same diseases as we have, but living mammals are the survivors of cancers, degenerative diseases, such as Alzheimer disease, arthritis, multiple sclerosis, or debilitating epidemics, such as AIDS, influenza and bubonic plague. Variants in immune response, viral restriction, and tumor suppressor genes have been retained by natural selection and passed down to living species as a protection against the scourges of their ancestors. Understanding the process of historic gene adaptation provides clues to design of treatments, particularly gene therapy, and drug design targeting specific gene products.

• As the comparative maps of index species representing the 20 modern orders of mammals are assembled, the opportunity to reconstruct the constraints, pattern,
tempo and mode of chromosome segment exchange will become a reality. Furthermore, the phylogenetic reconstruction of gene segment changes that have occurred would probably permit a robust phylogenetic hierarchy or evolutionary topology of the mammalian radiations. Because chromosomal exchange is precisely tractable and exceedingly slow (see below), the disposition of conserved chromosomal segments might represent the most powerful suite of phylogenetic (cladistic) characters ever encountered for resolving the precise hierarchy of mammalian evolution.

- The comparative genome maps affirm the hope to identify genetic determinants that drive and reinforce mammalian specification. Connecting the gene sequence to the magnificient adaptations that have led to bats, whales, amni-los and humans is not an unrealistic expectation for 21st century genetic researchers.

The feline genome project

The present gene map of the cat (Felis catus) was developed principally by use of a rodent × cat somatic cell hybrid panel plus fluorescent in situ hybridization (FISH) with molecular clones of certain cat genes. From this, 185 gene markers (97 of which have identifiable human homologs) have been assigned to the 19 cat chromosomes (Fig. 1). When the chromosomal position of the human homology is identified, one is struck by the extensive syntenic conservation (i.e., location of linked homologous genes on the same chromosome in both species) across the feline genome map. In cases of high marker density, for example, cat chromosomes A3, B2, B4, D1 and X, extensive strings of genes have conserved synteny. At least two chromosomes (D1 and X) appear to be conserved on blue. Chromosomal order of cat genes is not yet determined, although their physical order is established for the human homologs (Fig. 2). Both Fig. 1 and Fig. 2 affirm extensive syntenic conservation between cat and human, in most cases involving entire human chromosomal arms and spanning the centromere across several human chromosomes (2, 4, 6, 7, 11, 12 and X).

TIG October 1997 Vol. 13 No. 10

394
FIGURE 2. Abridged human gene map including 56 loci mapped in the cat (Fig. 1). Most human gene markers are syntenically assigned to a cytogenetic segment. The chromosome position of the feline homolog is indicated to the left of each gene symbol. Colored boxes to the right indicate the region of human chromosome that hybridized the chromosome painting to specific cat chromosome libraries.

TIG October 1997 Vol. 13 No. 10

395
The technology of interspecies chromosome painting (also called Zoo-FISH) has allowed us to inspect the patterns of genome exchange virtually by direct observation (Fig. 3). In this procedure, DOP-PCR amplification of flow-sorted single metaphase chromosomes are used as a fluorescent probe for in situ hybridization of distantly related species. In Fig. 3, we illustrate the result of single chromosome painting experiments comparing human and cat genomes, and summarize the whole genome comparisons in Fig. 1 (for human single chromosome probes painted on cat metaphase spreads) and Fig. 2 (for cat chromosome probes painted on human metaphase preparations).

Reciprocal human-cat chromosome paints (performed using isolated chromosomes from both species) have provided dramatic affirmation of several tentative conclusions drawn by comparative mapping. (1) Reciprocal painting decorates chromosomes or parts of chromosomes predicted by over 90% of the gene mapping assignments. (2) Painting physically extends the homology stretches to over 90 percent of the two species genomes, a significant advance over human-cat gene mapping comparisons, which cover 50-60% of the chromosome lengths. Although the resolution of painting is limited to the cytogenetic demarcation, the procedure offers a rapid and falsifiable (by gene mapping) glimpse of the extent and character of genomic conservation between distantly related species. (3) The results confirm and extend the remarkable degree of conservation of genome organization between cat and human predicted by comparative gene mapping. Sixteen of 23 human chromosomes have all their homology segments localized on single cat chromosomes (Fig. 1). The seven other human chromosomes are split between two cat chromosomes and six of the seven (all except human chromosome 4) are homologous to uninterrupted contiguos segments of their feline homolog (Fig. 1).

We recognized 30 homology segments defined by cat chromosome probes painting human chromosomes (Fig. 2) and 32 homology segments in the cat genome identified by human paint probes (Fig. 1). The comparable value for human-pig is 47 (Refs 42, 43, 45), for human-cow is 50 (Ref. 38) and for human-mouse (derived from counting homology segments) is 120 (Refs 7, 47, 48). Because carnivores and primates shared a common ancestor on the order of 65-80 million years ago, the number of genomic breaks computes to a rate of a single translocation every 10-12 million years, a dramatically slow rate of chromosomal evolution. The small number of homology segments that occur in cat-human genome comparisons does not consider inversions within the segments, although we suspect there are few of these since only two conserved syntenies (cat chromosome B1 and human chromosome 3) showed intercal homology segments with segments from another chromosome (Figs 1, 2).

For most human chromosomes syntenic conservation with cat homologs is greater than human-mouse conservation. Extensive comparison of some 1800 human-mouse homologous genes reveal 112 conserved segments, usually 1-10 cM in length (Refs 7, 22, 48). This number might actually rise with increasing map advancement to the theoretical prediction of 180 conserved segments estimated a decade ago. Consider human chromosome 11, which appears conserved as a single chromosome (B4) in cats; mouse segments homologous to human chromosome 11 are dispersed to a minimum of five separate segments on four mouse chromosomes (7, 2, 19, 7, 9).

When we look closely at the comparative association of homologous genes with unambiguous knowledge of their order in two species, the picture becomes even more complex due to cryptic inversions and translocations during the ancestry of the genomes. Genes are ordered in mouse by linkage mapping of interspecies backcrosses, but human linkage mapping has lagged behind until recently. The high resolution of radiation hybrids has resulted in the physical ordering of some 5000 human coding loci. R. Elliot compared the order of genes on human chromosome 11 to the placement of their mouse homologues and identified 20 distinct ordered linkage segments that were homologous to noncontiguous counterparts in the mouse genomes. This level of reassortment is also being revealed among other mouse chromosomes affining the general notion that the mouse genome is remarkably shuffled relative to that of human but, importantly, not beyond resolution of homology segments.

**Integrating the gene maps of mammals**

Once the level of conservation between species is determined for target mammal species, important details of genome alignment require comment. First, the principal categories of homology segments we encounter must be considered. As illustrated in Fig. 4, these include: (1) conserved synteny; (2) conserved segment with an undetermined gene order; (3) conserved linkage, unconserved gene order; (4) conserved linkage with conserved gene order; and (5) smallest conserved evolutionary unit segment (SCUES). Each of these can be resolved explicitly once both chromosomal assignment and physical order of genes is known for compared species.

It is important to realize that two very different classes of gene markers are being placed on mammalian gene
Type I markers are coding genes, which, through DNA sequence comparison and comparative mapping, offer precise recognition of homology, essential for genome comparisons. But intraspecies polymorphism is limited among Type I markers, making them of little use in pedigree or family-based gene localization. Type II markers, hypervariable microsatellite (also called short tandem repeat polymorphisms (STRP) or simple sequence length polymorphism (SSLP)) are 2-6 bp tandem repeats dispersed at random throughout vertebrate genomes. Type II loci are ideal for pedigree mapping because they are numerous (estimated as 1-200,000 per genome) and highly polymorphic. Assembly of Type II linkage maps has occurred in several mammalian species, including the cat, mouse, and cattle. The disadvantage of Type II loci is that they are seldom conserved beyond the family level so they are useless for comparative inference. The costly truth is that efficacious gene maps of index mammalian species must contain Type II loci (for pedigree mapping of traits) and Type I loci (to connect maps in a comparative sense to the gene dense human-mouse maps).

For maximum comparative information, Type I gene maps for different species should include the same homologous anchored reference loci in each species. A list of 321 such loci was selected three years ago based upon 5-10 cM spacing in the human and mouse maps, inclusion in developing cat and bovine maps, availability of clone probes for each marker, and including nearly all previously defined SCEUS segments among human, mouse, cat and cattle gene maps. The concept of comparative anchor loci was embraced by the comparative mapping community, but progress had been slow because heterologous probing of Southern blots, using largely human and mouse clones, was not always technically feasible.

Recently, we proposed a second version of comparative anchor loci, but with important improvements. Using powerful sequence analysis software linked together by a computer script, the 540,000 vertebrate sequences in GenBank were employed to align gene sequences from different mammalian orders. Conserved PCR primers were designed from adjacent conserved exons of nominated anchor loci with intention of amplifying a portion of two exons plus an intron of the gene homolog in virtually any mammal species. The short exon stretch (25-200 bp) of the PCR product would allow for gene identity verification and the intron sequence would be a source of DNA sequence divergence in gene mapping analyses. The cat genome provided a field trial for universality of the primers because few GenBank sequences are derived from species of the order Carnivora, which includes Felis catus.

A total of 537 loci was selected for PCR primer design including the 321 original comparative anchor markers. Primers were designed successfully for 410 of these genes and 318 were optimized empirically because they produce a single PCR product from domestic cat DNA (Fig. 5). Eighty-three percent of feline PCR products matched the original gene of primer design upon a homology database search. A preliminary PCR screening of the same primers using DNA of 20 mammalian species (most with developing gene maps) from 11 mammalian orders revealed that 25-52% of the 318 primers yielded a single PCR product using optimized cat PCR conditions. We predict that about 75% of the primers would be successfully optimized in any mammalian species. The new markers and their primers comprise the comparative equivalent of sequence tagged sites (STS) or ESTs, and have been named comparative anchor tagged sequences (CATS).

The principal gene mapping resources for mammal species include somatic cell hybrid panels, radiation hybrids, pedigrees and interspecies backcrosses. For the feline map we have concentrated historically on cell hybrid panels, but more recently an interspecies backcross between domestic cat and Asian leopard cat (Prionailurus bengalensis) has been developed. This cross, derived from natural breeding and artificial insemination, has yielded some 60 backcross offspring sufficient for 5 cM resolution linkage maps. We are now in the process of integrating all 318 Type I CATS markers plus a group of 250 microsatellite loci to a feline linkage map with maximal comparative value and polymorphism information content. Similar strategies are being applied with much promise in mouse, cattle, pig, sheep, and deer.

**Conclusions**

Rapid technological advances driven by human and mouse gene mapping projects have clearly facilitated...
The Human Genome Project has made significant progress in other species. If the Human Genome Project had been chosen, the ancestral genome organization for all modern mammals. The cooperation is a wealth of evolutionary context by which gene action and adaptation might be resolved. A byproduct would be revealing the phylogentic topology of the mammalian hierarchy. The units of genomic phylogeny are gene sequence organization and the SCEUS (Fig. 4). Like blocks or elements of genome exchange, defining SCEUSs among mammals and the itineraries they traverse in different mammal lineages provides a challenge rich with promise in resolving species' natural history.

There is a great deal yet to be done in this field. At least ten mammal orders have little or no gene mapping activity whatsoever. These include Cetacean (whales), Chiropterans (bats), Pholidota ( pangolins), Pinnipeds (seals), Proboscidea (elephants), Hyracoidea (hyraxes), Sirenia (manatees), Dermoptera (flying lemurs), Scandentia (tree shrews) and Xenarthra (shrews). These groups hold untapped reservoirs of comparative genomics, and even within domestic index species the connections are just beginning. Yet genetic technologies will certainly lead to a network of comparative genetics in the near future as powerful theory and computer algorithms will organize and interpret the enormous data sets. With increasing homology contacts the exchange of information from human to animal and back to human will address biological questions unimagined until very recently.

References
7 Copeland, N.G. et al. (1993) Science 262, 57–60
9 C. elegans database Genome Informatics Group, University of Maryland, College Park, MD (http://probe.nal.usda.gov)
15 Bishop, M.D. et al. (1994) Genetics 136, 619–639
18 Tate, D.W. et al. (1995) Genetics 139, 1383–1391
22 http://www.informatics.jax.org/

**Figure 5.** Comparative anchor tagged sequence (CATS) with tagged universal mammalian PCR primers maps based upon human cytogenetic position. CATS assigned in cow, mouse and cat gene maps are indicated by diamonds and when adjacent homologs are syntenic (located on the same chromosome), diamonds are connected. Different chromosomes are shaded in different colors for the three species.
The first microbial genome to be sequenced in its entirety was the 5.3 Mb of bacteriophage \(\Phi X 174\) in 1978 (Ref. 1). Between that time and 1995, the only other genomes sequenced were viral; the \textit{Escherichia coli} sequencing project had begun in the mid-1980s and it was widely expected that this would be the first complete bacterial sequence determined. However, in July 1995, to the surprise of many, the 1.83 Mb \textit{Haemophilus influenzae}\textsuperscript{strain Rd} sequence was completed and published\textsuperscript{2}. The remarkable achievement of this work is that it demonstrated that whole bacterial genomes can be sequenced quickly and relatively inexpensively. The entirely novel aspect of the approach was the focus of research over the next few years. In this article we outline what has been learned from this and other genome sequencing projects, and discuss some of the potential avenues of investigation that will follow in the 'post-genome era'.

\textbf{Haemophilus influenza: the impact of whole genome sequencing on microbiology}

\textbf{CHRISTOPH M. TANG (christoph.tang@paediatrics.oxford.ac.uk)}

\textbf{DEREK W. HOOD (dhood@work.nlbio.ox.ac.uk)}

\textbf{E. RICHARD MOXON (richard.moxon@paediatrics.oxford.ac.uk)}

\texttt{TIG October 1997 Vol. 13 No. 10}

Copyright \textcopyright 1997 Elsevier Science Ltd. All rights reserved. 0140-2658/97/$17.00