NCBI PubMed National Library of Medicine NLM

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books

Search PubMed for Go Clear

Limits Preview/Index History Clipboard Details

About Entrez

Text Version

Entrez PubMed
Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities

PubMed Services
Journal Browser
MeSH Browser
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
LinkOut
Cubby

Related Resources
Order Documents
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

Privacy Policy

Display Abstract Sort Save Text Clip Add Order

1: Trends Genet 1997 Oct;13(10):393-9 Related Articles, NEW Links

ELSEVIER SCIENCE FULL-TEXT ARTICLE

Comparative genomics: lessons from cats.

O'Brien SJ, Wienberg J, Lyons LA.

Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702-1201, USA. obrien@ncifcrf.gov

The genomics era, spear headed by dazzling technological developments in human and mouse gene mapping, has additionally provoked extensive comparative gene mapping projects for domestic species of several vertebrate orders. As the human genome project promises a one dimensional string of 100,000 genes and sequences, comparative mapping will extend that inference to a second dimension representing index species of the 20 living mammalian orders and to a third dimension by phylogenetic description of the genomes of mammal ancestors. We review here the remarkable extent of genome homology conservation among mammals illustrated by technology applications in the feline genome project.

Publication Types:
- Review
- Review, Tutorial

PMID: 9351340 [PubMed - indexed for MEDLINE]

Display Abstract Sort Save Text Clip Add Order

Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Freedom of Information Act | Disclaimer

i686-pc-linux-gnu Aug 30 2002 15:17:13

SCIENCE DIRECT

Register or Login: user name Password: Go

Brought to you by:
National Institutes of Health Library

Home Publications Search My Alerts My Profile Help

Quick Search: within All Sources Go Search tips

Trends in Genetics
Volume 13, Issue 10, October 1997, Pages 393-399

This Document
- Abstract
- PDF (648 K)

Actions
- Cited By
- Save as Citation Alert
- Export Citation

PII: S0168-9525(97)01297-3

Comparative genomics: lessons from cats

Stephen J. O'Brien, Johannes Wienberg and Leslie A. Lyons

Available online 17 December 1997.

Abstract

The genomics era, spearheaded by dazzling technological developments in human and mouse gene mapping, has additionally provoked extensive comparative gene mapping projects for domestic species of several vertebrate orders. As the human genome project promises a one dimensional string of 100 000 genes and sequences, comparative mapping will extend that inference to a second dimension representing index species of the 20 living mammalian orders and to a third dimension by phylogenetic description of the genomes of mammal ancestors. We review here the remarkable extent of genome homology conservation among mammals illustrated by technology applications in the feline genome project.

Index Terms: species difference; genetic variability

Trends in Genetics
Volume 13, Issue 10, October 1997, Pages 393-399

This Document
- **Abstract**
- PDF (648 K)

Actions
- Cited By
- Save as Citation Alert
- Export Citation

Home | Publications | Search | My Alerts | My Profile | Help | SCIENCE DIRECT

Send feedback to ScienceDirect
Software and compilation © 2002 ScienceDirect. All rights reserved.
ScienceDirect® is an Elsevier Science B.V. registered trademark.

Your use of this service is governed by Terms and Conditions. Please review our Privacy Policy for details on how we protect information that you supply.

Comparative genomics: lessons from cats

STEPHEN J. O'BRIEN (obrien@ncifcrf.gov)

JOHANNES WIENBERG (jw@mole.bio.cam.ac.uk)

LESLIE A. LYONS (lyons@ncifcrf.gov)

The genomics era, spearheaded by dazzling technological developments in human and mouse gene mapping, has additionally provoked extensive comparative gene mapping projects for domestic species of several vertebrate orders. As the human genome project promises a one dimensional string of 100 000 genes and sequences, comparative mapping will extend that inference to a second dimension representing index species of the 20 living mammalian orders and to a third dimension by phylogenetic description of the genomes of mammal ancestors. We review here the remarkable extent of genome homology conservation among mammals illustrated by technology applications in the feline genome project.

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 heritable 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[6–13]. 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 genomic 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[14–16], sheep[17,18], pig[19], deer[18] and chicken[20].

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 chromosomes by linkage mapping, radiation hybrids, or both[2–4,21]. The mouse map contains some 7000 gene loci plus over 7000 simple sequence length polymorphisms (SSLPs) placed on linkage maps largely through interspecies backcrosses[6,7,22,23]. 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[24,25] 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 chromosome 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[28,29]. Their identification by veterinary clinicians 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[30]. Consider that over 100 000 mammalian species have occurred since the origin of mammals in the Mesazoic 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[31].
- As the comparative maps of index species representing the 20 modern orders of mammals are assembled, the opportunity to reconstruct the constraints, pattern,

Copyright © 1997 Elsevier Science Ltd. All rights reserved. 0168-9525/97/$17.00
PII: S0168-9525(97)01297-3

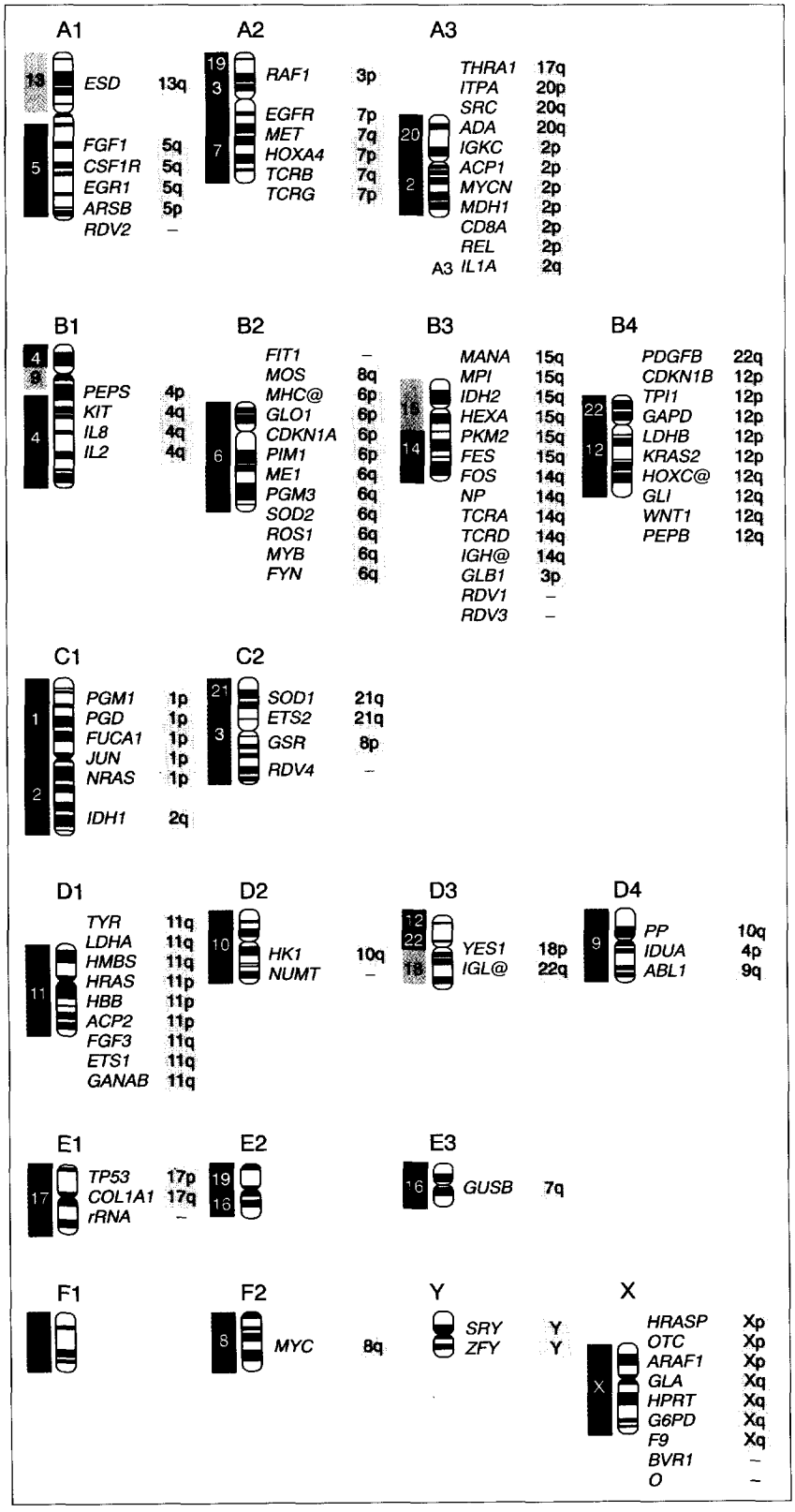

FIGURE 1. Genetic map of the domestic cat. Gene symbols are based, in most cases, on human gene nomenclature. The majority of loci were assigned using isozyme or Southern blot analysis in a panel of rodent × cat somatic cell hybrids although some markers were mapped in linkage crosses or using fluorescent *in situ* hybridization[32,35–37]. The numbers shaded in gray (to the right of each gene symbol) refer to the human chromosome to which the feline gene homolog has been assigned. Colored blocks with numbers (to the left of each chromosome) illustrate the results of chromosome painting of feline metaphase chromosomes using libraries of cell-sorted individual human chromosomes as a hybridization probe[39]. The gene order of these loci has not yet been determined.

tempo and mode of chromosome segment exchange will become a reality[12,26,32–34]. 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 hierachy of mammalian evolution.

- The comparative genome maps affirm the hope to identify genetic determinants that drive and reinforce mammalian speciation. Connecting the gene sequence to the magnificent adaptations that have led to bats, whales, armadillos 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[32,35–37]. From this, 105 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 homolog 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 *en bloc*. 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).

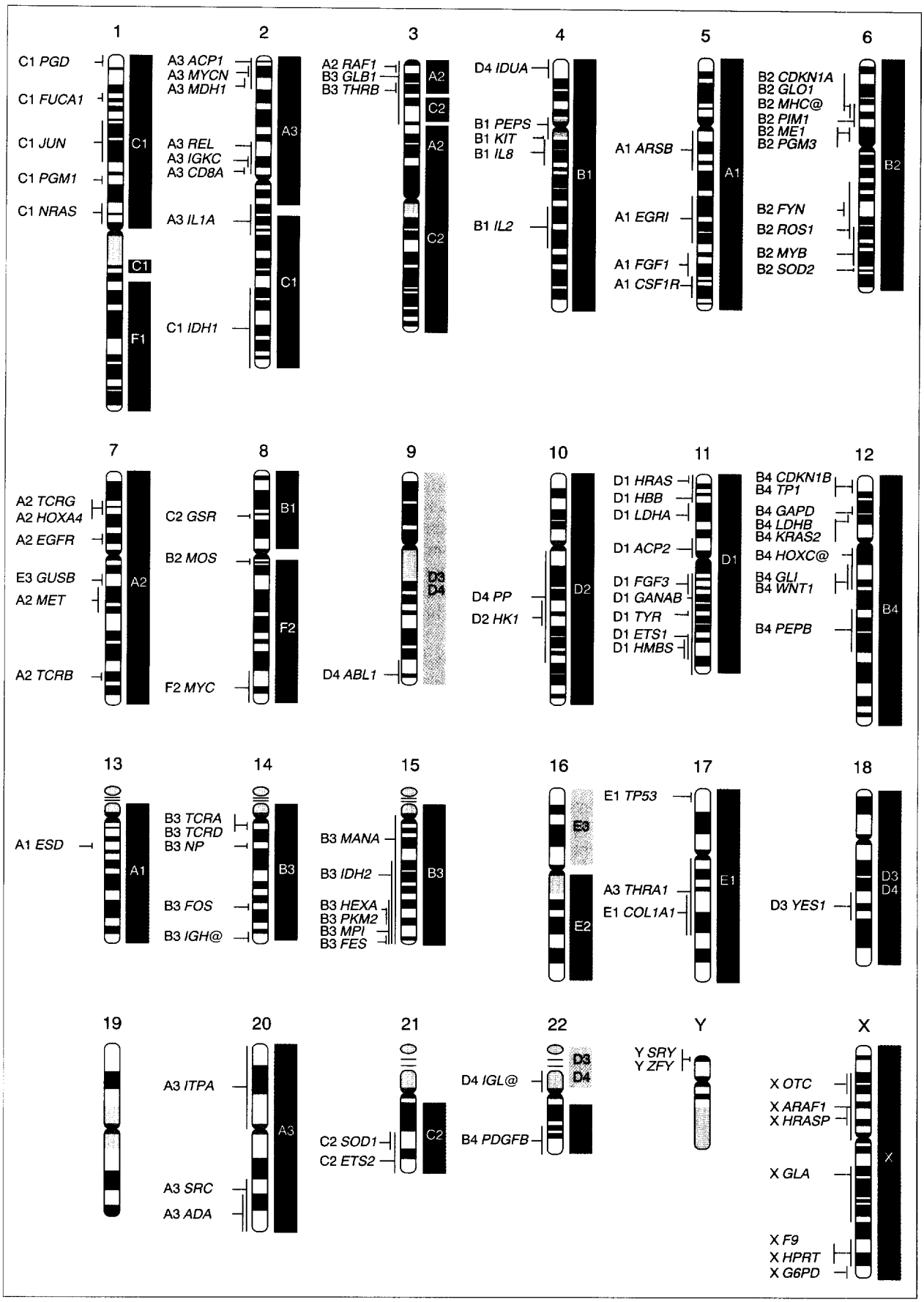

FIGURE 2. Abridged human gene map including 96 loci mapped in the cat (Fig. 1). Most human gene markers are subchromosomally assigned to a cytogenetic segment. The chromosome position of the feline homolog is indicated to the left of each gene symbol. Colored boxes to the right indicate the region of human chromosome that hybridized (by chromosome painting) to specific cat chromosome libraries[39].

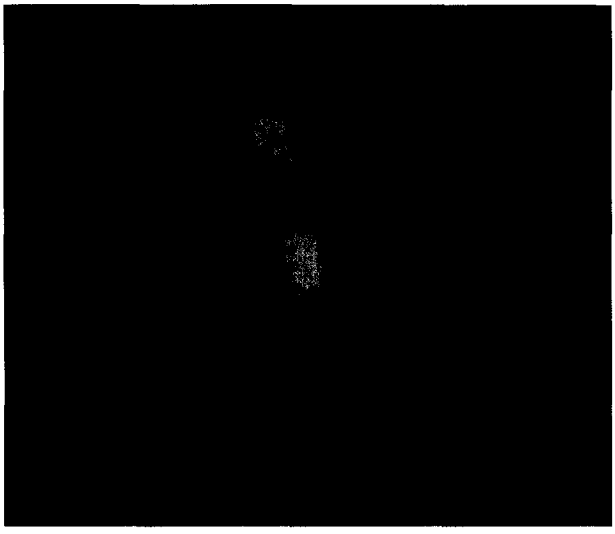

FIGURE 3. Feline metaphase karyotype hybridized by chromosome painting with fluorescent-labeled molecular clones derived from flow-sorted human chromosome 11 decorating the homologous feline chromosome D1.

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[38–45]. 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[39,40,45]. (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[46]. 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 contiguous 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 intercalation of 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[47]. 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)[49].

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[7,48]. 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[3]. 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[49]. This level of reassortment is also being revealed among other mouse chromosomes[48] affirming 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[7,26]. 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 (SCEUS). 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

maps[50]. 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[4,6,51]. 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[4,6,15,17,51–54]. 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)[14,49,50].

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[46]. 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[27]. 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[27,46]. 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[27]. 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).

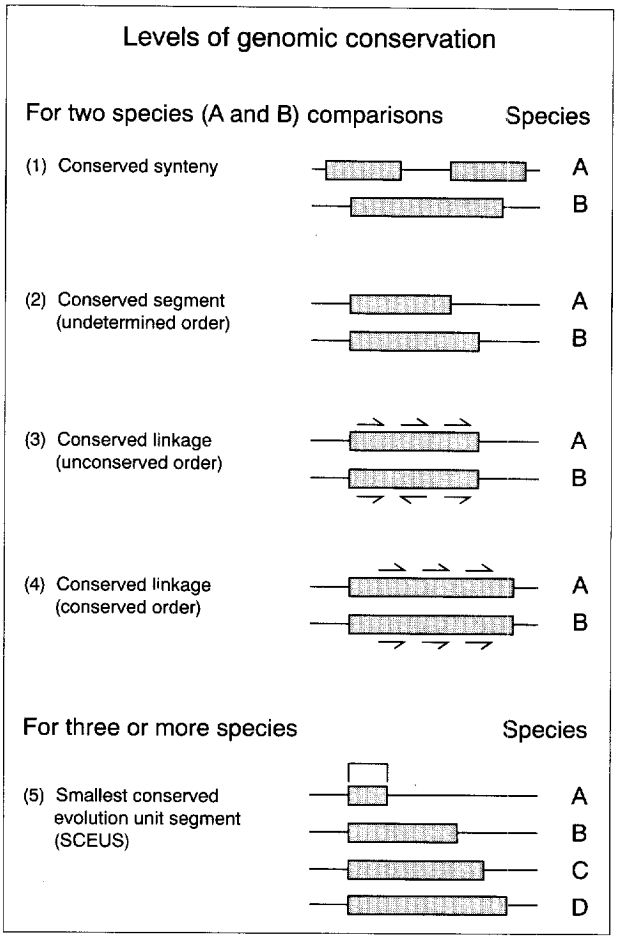

FIGURE 4. Different levels of chromosome segment conservation for comparisons between two or more species gene maps. Categories 3 and 4 are subsets of category 2 that are defined when linkage order of homologous genes in both species is determined.

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 map[52]. 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[6,7], cattle[15,16], pig[19,54], sheep[17] and deer[18].

Conclusions

Rapid technological advances driven by human and mouse gene mapping projects have clearly facilitated

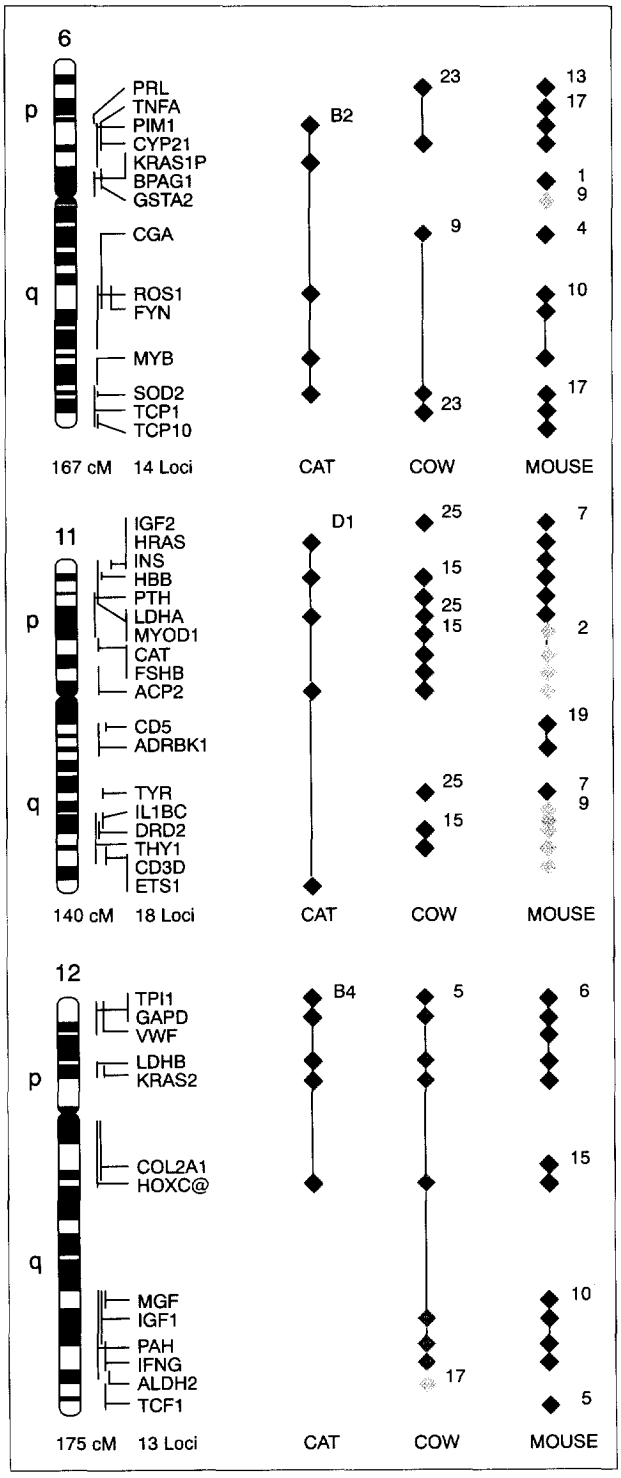

FIGURE 5. Comparative anchor tagged sequence (CATS with tested universal mammalian PCR primers) maps based upon human cytogenetic position. CATS assigned in cow, mouse and cat gene maps are indicated by diamonds and when adjacent homologs are syntenic (located on the same chromosome), diamonds are connected. Different chromosomes are shaded in different colors for the three species.

progress in other species. If the Human Genome Project sets a principal goal of an ordered one-dimensional listing of 100 000 loci and their sequences, comparative map development promises to add two additional dimensions. The first is to include the genomes of represented modern species of mammals; the second is to reconstruct, by phylogenetic inference, the ancestral genome organization for all modern mammals. The jackpot is a wealth of evolutionary context by which gene action and adaptation might be resolved. A byproduct would be revealing the phylogenetic 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 Xenarthia (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

1 Lander, E.S. (1996) *Science* 274, 536–540
2 Hudson, T.J. *et al.* (1995) *Science* 270, 1945–1954
3 Schuler, G.D. *et al.* (1996) *Science* 274, 540
4 Dib, C. *et al.* (1996) *Nature* 380, 152–154
5 Collins, F. and Galas, D. (1993) *Science* 262, 43–44
6 Dietrich, W.F. *et al.* (1996) *Nature* 380, 149–153
7 Copeland, N.G. *et al.* (1993) *Science* 262, 57–66
8 Dujon, B. (1996) *Trends Genet.* 12, 263–270
9 *C. elegans* database Genome Informatics Group, University of Maryland, College Park, MD (http://probe.nalusda.gov)
10 Bellen, H.J. and Smith, R.F. (1995) *Trends Genet.* 11, 456–457
11 O'Brien, S.J., ed. (1993) *Genetic Maps, Locus Maps of Complex Genomes* (6th edn), Cold Spring Harbor Laboratory Press
12 O'Brien, S.J. *et al.* (1988) *Annu. Rev. Genet.* 22, 323–351
13 Marshall-Graves, J.A. (1996) *Annu. Rev. Genet.* 30, 233–260
14 Womack, J.E. and Kata, S.R. (1995) *Curr. Opin. Genet. Dev.* 5, 724–733
15 Bishop, M.D. *et al.* (1994) *Genetics* 136, 619–639
16 Barendse, W. *et al.* (1994) *Nat. Genet.* 6, 227–235
17 Crawford, A.M. *et al.* (1995) *Genetics* 140, 703–724
18 Tate, M.L. *et al.* (1995) *Genetics* 139, 1383–1391
19 Archibald, A.L. *et al.* (1995) *Mamm. Genome* 6, 157–175
20 Burt, D.W. *et al.* (1995) *Trends Genet.* 11, 190–194
21 Adams, M.D. *et al.* (1995) *Nature* (Suppl. 6547S) 377, 3–174
22 http://www.informatics.jax.org/
23 http://www.ncbi.nlm.nih.gov/dbEST/index.html
24 Lindsley, D.L. and Grell, E.H. (1968) *Genetic Variations of Drosophila melanogaster*, Carnegie Institute
25 Lindsley, D.L. and Zimm, G.G. (1992) *The Genome of Drosophila*, Academic Press
26 Comparative Genome Organization, First International Workshop (1996) *Mamm. Genome* 7, 717–734
27 Lyons, L.A. *et al.* (1997) *Nat. Genet.* 15, 47–56
28 Desnick, R.J., Patterson, D.F. and Scarpelli, D.G., eds (1982) *Animal Models of Inborn Errors of Metabolism*, Alan R. Liss

29 Nicholas, F.W. and Harper, P.A.W. (1996) *Aust. Vet. J.* 73, 64–66
30 O'Brien, S.J. (1995) *Nat. Med.* 1, 742–744
31 O'Brien, S.J. and Dean, M. *Sci. Am.* (in press)
32 O'Brien, S.J. and Nash, W.G. (1982) *Science* 216, 257–265
33 Modi, W.S. and O'Brien, S.J. (1988) in *Chromosome Structure and Function* (Gustafson, J.P. and Appels, R., eds), pp. 215–242, Plenum Press
34 Sankoff, D. and Nadeau, J.H. (1995) *Centre de Recherches Mathematiques*, Université de Montreal, 1–10
35 O'Brien, S.J. *et al. J. Hered.* (in press)
36 Lopez, J.V. *et al.* (1996) *Genomics* 33, 229–246
37 Yuhki, N. and O'Brien, S.J. (1988) *Immunogenetics* 27, 414–425
38 Wienberg, J. and Stanyon, R. (1995) *Curr. Opin. Genet. Dev.* 5, 792–797
39 Wienberg, J. *et al. Cytogenet. Cell Genet.* (in press)
40 Rettenberger, G. *et al.* (1995) *Chromosome Res.* 3, 479–486
41 Solinas-Toldo, S., Lengauer, C. and Fries, R. (1995) *Genomics* 27, 489–496
42 Fronicke, I., Chowdhary, B.P., Scherthan, H. and Gustavsson, L. (1996) *Mamm. Genome* 7, 285–290
43 Rettenberger, G. *et al.* (1995) *Genomics* 26, 372–378
44 O'Brien, S.J. (1993) *Curr. Biol.* 3, 395–397
45 Goureau, A. *et al.* (1996) *Genomics* 36, 252–262
46 O'Brien, S.J. *et al.* (1993) *Nat. Genet.* 3, 103–112
47 Nadeau, J.H. and Taylor, B.A. (1984) *Proc. Natl. Acad. Sci. U. S. A.* 81, 814–818
48 DeBry, R.W. and Seldin, M.F. (1996) *Genomics* 33, 337–351
49 Shows, T.B. *et al.* (1996) *Cytogenet. Cell Genet.* 74, 1–56
50 O'Brien, S.J. (1991) *Curr. Opin. Genet. Dev.* 1, 105–111
51 Menotti-Raymond, M., David, V.A., Stephens, J.C. and O'Brien, S.J. *J. Forensic Sci.* (in press)
52 Lyons, L.A. *et al.* (1994) *Anim. Biotech.* 5, 103–111
53 Menotti-Raymond, M.A. and O'Brien, S.J. (1995) *J. Hered.* 86, 319–322
54 Johansson, M., Ellegren, H.J. and Andersson, L. (1995) *Genomics* 25, 682–690

S.J. O'Brien *and* ***L.A. Lyon*** *are in the Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702-1201, USA.*
J. Wienberg *is in the Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, UK CB2 1QR.*

Haemophilus influence: the impact of whole genome sequencing on microbiology

CHRISTOPH M. TANG (christoph.tang@paediatrics.oxford.ac.uk)

DEREK W. HOOD (dhood@worf.molbiol.ox.ac.uk)

E. RICHARD MOXON (richard.moxon@paediatrics.oxford.ac.uk)

The publication of the Haemophilus influenzae genome sequence in 1995 was a landmark in microbiological research. It has changed our understanding of the prokaryotic world, and will influence the approach and focus of research on microorganisms 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'.

The first microbial genome to be sequenced in its entirety was the 5386 bp of bacteriophage ΦX174 in 1978 (Ref. 1). Between that time and 1995, the only other genomes sequenced were viral; the *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 *Haemophilus influenzae* strain Rd sequence was completed and published[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 taken by Venter and his colleagues was that the sequence was determined from shotgun-cloned small (1.6–2.0 kb) fragments of DNA, and then these short stretches were assembled into large contiguous segments (and eventually the entire chromosome) by computer. This obviated the need for the laborious time-consuming step of obtaining a physical map of a bacterial genome by ordering a library (cosmid or phage) containing large fragments of DNA to assemble the sequence; many believed that this step was a prerequisite for completing the sequence of a bacterial genome. The shotgun method also largely circumvented the problem of 'unclonable' DNA, which delayed or prevented the completion of physical maps; any gaps evident from assembling small shotgun cloned fragments were closed using PCR and inserts from two unmapped lambda libraries. The *H. influenzae* genome project was proof of the principle that physical mapping is unnecessary, and immediately demonstrated that microbial genomes can be sequenced and the information assembled far faster than had been thought possible. Once the *Haemophilus* sequence became available, the utility of whole genome sequencing was readily apparent. It contains a vast body of information in terms of the genes present, and the organization of the coding and noncoding regions; nothing except the whole genome sequence can show this. The publication of the *Haemophilus* sequence initiated a surge of interest in whole bacterial genome sequencing, and has led to an impetus to determine the genome sequences of many microorganisms. Until the release of the *Haemophilus* sequence, microbial genome sequencing had been an under-resourced area of research with little interest shown by either funding agencies or scientists themselves. Now there is the very real

Copyright © 1997 Elsevier Science Ltd. All rights reserved. 0168-9525/97/$17.00
PII: S0168-9525(97)01245-6