

The Promise of Comparative Genomics in Mammals

Stephen J. O'Brien,¹ Marilyn Menotti-Raymond,¹ William J. Murphy,¹ William G. Nash,¹ Johannes Wienberg,¹ Roscoe Stanyon,¹ Neal G. Copeland,² Nancy A. Jenkins,² James E. Womack,³ Jennifer A. Marshall Graves⁴

Dense genetic maps of human, mouse, and rat genomes that are based on coding genes and on microsatellite and single-nucleotide polymorphism markers have been complemented by precise gene homolog alignment with moderate-resolution maps of livestock, companion animals, and additional mammal species. Comparative genetic assessment expands the utility of these maps in gene discovery, in functional genomics, and in tracking the evolutionary forces that sculpted the genome organization of modern mammalian species.

No one is sure exactly when, nor is there a strong consensus as to precisely where it happened. Yet it is almost certain that sometime around 165 million years ago, probably in Eurasia, a modest rat-sized creature with squared forelimbs adapted for travel, sprawling hind legs reminiscent of lizards or turtles, and a genome of considerable potential began an evolutionary divergence from reptiles culminating in a panoply of mammalian descendants who would one day dominate the planet. The mammals' earliest ancestors, documented this year by a spectacular near-complete skeletal fossil of *Jeholodens jenkinsi* from northern China (1), remained diminutive, gradually evolving for some 100 million years at the feet of the dinosaurs. An abrupt extinction of the dinosaurs 63 to 66 million years ago created a worldwide ecological vacuum that was backfilled by the mammalian radiations.

Tens of thousands of mammalian species have emerged, diverged, and disappeared in this interval, and the 4600 to 4800 species living today comprise approximately 28 orders, including the primitive egg-laying mammals (Monotremata: platypus and echidna), 7 marsupial orders, and 20 placental (eutherian) orders (2). Encrypted in the genomes of surviving species are novel genes, lost genes, modified genes, and reordered genes. These blueprints for species adaptation and distinction are vestiges of pivotal changes that discriminated a whale from a bat, a dog from a cat, or a chimpanzee from human. Today's molecular deciphering of the genomes of living species, whether focusing on homologous gene se-

quences, gene segments, chromosomes, or entire genomes, provides a new vision of important evolutionary questions about natural history, species origins and survival, and adaptation to occupy ecological niches. The comparative genomics approach is already revealing valuable insights into developmental functions, reproductive enhancements, inborn errors, and disease defense mechanisms that have protected our ancestors (and ourselves) from extinction.

In the 20th century, genetic science has moved from deducing how visible hereditary phenotypes are transmitted to anticipation of online, full-length DNA sequences of genomes from human and nominated model organisms (mouse, fly, worm, yeast, and *Escherichia coli*) within the next few years (3). Advancing technologies of the Human Genome Project are now being harnessed to describe the complexities of genome organization not only in the "gene-rich" mammal species (that is, human, mouse, and rat, which are species with high-density gene maps) but also in additional mammals that are representative of distant evolutionary lineages (4–7). The promise is to detail distinctive parallels in genome assemblages as a prelude to interpreting species and individual variation in a functional and evolutionary context. After centuries of study of comparative anatomy, behavior, and physiology to better understand human medicine, genomic information is reversing the course of information flow. Our knowledge of human genetics is leading the genomics era, so much so that human gene regulation and orientation inform us as to animal gene action. The comparative cycle closes when human functional genomics—the science of connecting gene to gene action—is inferred by comparative animal genetics.

Developing Mammalian Gene Maps

All mammals contain between 70,000 and 100,000 genes arranged in linear order along their chromosomes, with a total length of

about 3.2 billion nucleotide pairs. Chromosome numbers range from a low of three pairs ($2N = 6$ in the Indian muntjac, *Muntiacus muntjak*) to a high of 67 pairs ($2N = 134$ in the black rhinoceros, *Diceros bicornis*). Gene maps have been constructed in human, mouse, and about 30 other mammal species for two general reasons: first, as a resource for locating the genetic determinants of heritable characteristics, behaviors, and phenotypes; and second, as a template for resolving and interpreting patterns of evolving genome organization in their ancestry. Progress on more advanced mapping projects is summarized in Table 1.

The most efficacious mammalian gene maps include an integration of three categories of markers (8). Type I markers are coding genes that through DNA sequence comparison and comparative mapping are essential for identification of gene orthologs in distantly related species (that is, genes in different species that are descended from a single gene of a common ancestor). However, as a result of low polymorphism, type I markers offer little power in assessments of pedigree or population diversity. Type II markers [hypervariable microsatellites, also called short tandem repeats (STRs)] are highly informative in pedigree, forensic, and population assessment, because there are over 100,000 near-randomly dispersed STRs throughout mammal genomes, and because each carries multiple alleles. Type II STRs are less useful for orthologous locus recognition between species of different mammal orders, because the lifetime of type II markers and their distinctive flanking DNA sequence, which is required for polymerase chain reaction, arose after the divergence of mammal orders.

Type III markers are common bi-allelic single-nucleotide polymorphisms (SNPs) within coding regions, or more often in non-coding intron or intergenic regions (9). SNPs are also valuable for pedigree, family, or population screens within species, particularly with automated array-based genotyping technologies, but are usually uninformative when used for comparative ortholog identification between orders. Type III SNP markers occur once every 500 to 1000 base pairs (bp) in the human genome, totaling an estimated 3 million SNPs in the genomes of human and other mammals of comparable within-species genetic diversity. Approximately 8000 hu-

¹Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702–1201, USA. ²Mammalian Genetics Laboratory, Advanced BioScience Laboratories, National Cancer Institute, Frederick, MD 21702–1201, USA. ³Department of Veterinary Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843, USA. ⁴Department of Genetics and Evolution, LaTrobe University, Melbourne, Victoria 3083, Australia.

man SNPs have been described (9), and a pharmaceutical consortium has mounted an effort to place 300,000 human SNPs on the map by 2002 because of their promise in pharmacogenomic uses (10). The ideal goal of human, mouse, and other mapping projects has been to create a dense ordered map of each chromosome, integrating at least 500 type I markers, 1000 type II markers, and 1000 to 3000 type III SNPs (Table 1).

Traditional mapping technologies that taught us the power of dense gene maps a decade ago (somatic cell hybrid panels, fluorescent in situ hybridization, and pedigree analysis) have been supplemented by powerful new approaches that increase the precision of ordered gene-marker chromosome maps and comparative assessment in mammals. Interspecies hybrid backcrosses, first developed in mice to exploit accumulated sequence divergence around type I (coding) genes between related species that can interbreed, have produced dense meiotic linkage maps of mice, cat, cattle, deer, and pigs (11–14). Radiation hybrid panels, whereby chromosome fragments (produced by x irradiation) from mapped species are fused to Chinese hamster chromosomes in random combinations (after x-ray irradiation) have been used to construct dense physical orders of type I and II markers in human, mouse, rat, cattle, pig, cat, dog, and zebrafish (Table 1) (15–21). Even higher marker resolution is provided by DNA segment cloning and contig alignment from arrayed bacterial artificial chromosome (BAC), P1 phage artificial chromosome (PAC), and yeast artificial chromosome (YAC) libraries in a number of mammals (22). Finally, the application of interspecies chromosome painting (also called Zoo-FISH), whereby DNA from fluorescently labeled flow-sorted individual chromosomes of one species is hybridized in situ to metaphase spreads of a compared species, has allowed identification of evolutionarily con-

served chromosomes, chromosome arms, and segments virtually by direct observation (23).

The mammal gene maps listed in Table 1 have grown extensively through application of these methodologies (see foldout), placing mammals at last in a position to apply comparative genomic inference across biological disciplines. The exercise is new for mammals, but both plants and prokaryotes have already advanced considerably. Comparative maps of a dozen plant species have been analyzed to estimate rates of chromosome exchange and even to reconstruct ancestral genome organization predating the divergence of monocots and dicots (24). Among prokaryotes, full genome sequence comparisons of 13 bacterial species has allowed for the first time a chance to see the addition and loss of all the genes in compared species, showing that in many comparisons, 20 to 50% of the genes are gained or lost (25).

Biomedical Applications of Comparative Genomics with Rodent Models

Rodent gene orthologs of heritable human disorders have identified many venues for comparative insight, although it now seems that our ability to map and sequence human and mouse genes is outpacing attempts to discern their functions. Consider that over half of the 70,000 to 100,000 human expressed sequence tags (ESTs) for RNA transcripts are already mapped, and even more gene products have sequence representation in mouse and human EST databases (16), but less than 6000 genes have names and known functions. Further, it now appears that monogenic diseases, which were the great successes of the early years of the Human Genome Project, represent a simplification of reality, because most phenotypes are both polygenic and multifactorial (modified by environmental influences).

The understanding of complex disorders

such as diabetes, hypertension, and obesity has advanced considerably through the use of mouse and rat models (26–28). For example, human obesity is a common malady, particularly in Western societies, that has enormous public health impact. Yet until recently, the metabolic pathways associated with obesity remained obscure. Seminal advances in our understanding of obesity have come from the positional cloning of a number of mouse gene mutations (such as *fat*, *tubby*, *obese*, and *diabetes*) that cause obesity, a well as subsequent studies showing that some of these gene homologs are mutated in morbidly obese humans (26). One mouse mutation that suppresses diet-induced obesity (*mahogany*) was shown to be homologous to the human gene *attractin*, which encodes a serum glycoprotein secreted by activated T lymphocytes, which modulates immune cell interactions (27). Recent human clinical trials extending these inferences for therapeutic intervention to treat these devastating diseases are particularly promising (29).

Studies of hypertension in rats have uncovered multiple potential candidate genes for the same disease in humans, identified by comparative mapping (28). Aneuploidy for a small segment of mouse chromosome 16, homologous to human chromosome 21, has implicated not just one gene but the cluster of genes that together contribute to the developmental consequences of trisomy 21, Down syndrome (30). There are many additional examples of similar interactive reasoning in human-rodent genomic considerations. Most workers take for granted the occurrence and utility of parallel genome organization between humans and rodents. The Jackson Laboratory's genome database lists more than 1000 spontaneous mouse mutations, of which 128 have been characterized at the DNA level (31). Fifty-eight of these (45%) have homologous gene mutations discovered in humans with an associated genetic disease. Thirty-

Table 1. Advanced genetic maps in vertebrate species as compared to human genome organization.

Species	Order	Haploid chromosome number	No. mapped type I coding genes	No. mapped type II microsatellites	Genome length (cM)	No. of conserved synthetic segments with human	Min. number rearrangements versus humans*	Reference no.
Human	Primate	23	>30,000	~8,000	3,300	—	—	(15)
Mouse	Rodentia	20	6,992	7,377	1,450	180	160	(11, 16, 18)
Rat	Rodentia	21	552	~8,000	1,500	109	88	(17)
Cat	Carnivora	19	~500	254	3,300	32	13	(19, 51)
Dog	Carnivora	39	218	276	2,700	68	45	(52)
Mink	Carnivora	15	77	0	—	33	18	(80)
Cow	Cetartiodactyla	30	400–500	1,236	2,990–3,532	50	27	(12, 41, 42)
Goat	Cetartiodactyla	30	257	307	3,100	107	85	(73)
Sheep	Cetartiodactyla	27	254	504	3,063	40	17	(81)
Pig	Cetartiodactyla	19	369	~1,000	2,300–2,500	47	28	(82)
Horse	Perissodactyla	32	53	309	2,000	47	24	(83)
Shrew	Eulipotyphyla	10	31	0	—	33	23	(57)
Chicken	Galliformes	39	220	677	4,000–4,200	79–116	48	(58)
Zebrafish	Cypriniformes	25	275	2,000	2,145	252	227	(21)

*The minimum number of rearrangements equals the number of conserved segments minus the lower haploid number (N) of the compared species.

five of these gene variants (60%) were first discovered in mice, leading to characterization of the human homolog, and the rest (40%) were first reported in humans, stimulating mouse mutation detections (31).

Combining mouse gene knockout technologies with comparative gene mapping inferences has also led to some extremely valuable advances in assessing human gene function. One vivid example of interactive comparative insight involved the role of chemokine receptors in the pathogenesis of human immunodeficiency virus-acquired immunodeficiency syndrome (HIV-AIDS). Originally characterized in mice, gene families encoding chemokines ($n = 60$ loci) and chemokine receptors ($n = 16$ loci) were shown to play a pivotal role in healing abrasions and inflammations (32). In 1996, a series of cell biology and virology advances demonstrated that two human chemokine receptors, CCR5 and CXCR4, served as requisite entry portals for HIV-1 infection of T lymphocytes (33). Only

very rare and innocuous polymorphisms were discovered in the human *CXCR4* gene (34), but a large and rather common deletion mutation (32 bp) in the human *CCR5* receptor gene (*CCR5-Δ32*) effectively blocks HIV-1 infection in exposed individuals who are homozygous for the variant by eliminating the required HIV-1 receptor (35). *CCR5-Δ32/Δ32* homozygotes have no apparent genetic disease in spite of their complete loss of CCR5 receptors on lymphocyte cell surfaces (35). Why the difference in mutational occurrence in the two genes? Mouse knockouts of *CXCR4* are embryonic lethals, because the *CXCR4* chemokine binding functionality is unique and essential (36). The essential aspect of the human *CXCR4* gene function would explain the absence of disruptive *CXCR4* mutations. In contrast, mice with *CCR5* knockouts are rather healthy, because like the human homolog, the signaling function of *CCR5* (in response to chemokine ligands RANTES, MIP1 α , and MIP1 β) is re-

dundant in the mammalian genome, so that human and mouse *CCR5* knockouts are functionally compensated and healthy (32, 35, 37). The dispensability of both human and mouse *CCR5* make it an attractive object for AIDS therapy development (38).

Within the next few years, the entire sequence of both the mouse and human genomes will be determined, appreciably expanding the potential for comparative inference (3, 39). As this occurs, an era of functional genomics will use rodent models extensively, both to identify candidate genes for analogous functions and to define their interaction with other genes in the context of mutation, environment, infectious disease, toxins, age, sex, and other cofactors that contribute to human phenotypes. Nearly every human gene has a mouse homolog (known exceptions are some Y-chromosome analogs that are present in either human or mouse but not in both species) (40). This means that mouse homologs of virtually all human genes will one day be amenable to polymorphism discovery, to mutation by knockout, to transgenesis, and to medical intervention trials.

Table 2. Human hereditary disorders with identified mutations and associated phenotype in nonrodent species. See (52) for dog mapping and (19, 51) for cat mapping. For disease genes, see (49, 50) and <http://lgd.nci.nih.gov>.

Species	Disease or phenotype	Human gene symbol	Human chromosome	Species chromosome
Cow	Uridine monophosphate synthetase deficiency	<i>UMPS</i>	3q13	Bta1
	Leucocyte adhesion deficiency (LAD)	<i>ITGB2</i>	21q22.3	Bta1
	Double muscling	<i>GDF8</i>	2q32.1	Bta2
Sheep	Chondrodysplasia (Spider Lamb)	<i>FGFR3</i>	4p16.3	Oar
Pig	Porcine stress syndrome on malignant hyperthermia	<i>RYR1</i>	19q13.1	Ssc
	Hypercholesterolemia	<i>LDLR</i>	19p13.2-13.1	Ssc
	Oedema disease	<i>FYT1</i>	19q13.3	Ssc
	Coat color (dominant white)	<i>KIT</i>	4p11-q12	Ssc
	Coat color (red/black)	<i>MC1R</i>	16q24.3	Ssc
Cat	GM2 gangliosidosis (Sandhoff's disease)	<i>HEXB</i>	5q13	FcaA1
	Chylomicronemia	<i>LPL</i>	8p22	FcaB1
	Mannosidosis	<i>MANB</i>	19cen-q13.1	FcaA2
	MPSVI	<i>ARSB</i>	5q11-q13	FcaA1
	Glycogen storage disease, type IV	<i>GBE1</i>	3p21	FcaC2
	MPSVII	<i>GUSB</i>	7q21.11	FcaE3
	Duchenne/Becker muscular dystrophy	<i>DMD</i>	Xp21.2	FcaX
	Dog	Complement 3 deficiency	<i>C3</i>	19p13.3-p13.2
Fucosidosis	<i>FUCA1</i>	1p34-p36	Cfa2	
Glycogen storage disease, type 1a	<i>G6PC</i>	17q21	Cfa9	
Glycogen storage disease VII	<i>PFKM</i>	12q13.3		
Hemophilia B	<i>F9</i>	Xq27.1-q27.2	CfaX	
Krabbe disease	<i>GALC</i>	14q31		
MPSI	<i>IDUA</i>	4p16.3		
MPSIIIA	<i>SGSH</i>	17q25.3		
Duchenne muscular dystrophy	<i>DMD</i>	Xp21.2	CfaX	
Hereditary nephritis	<i>COL4A5,6</i>	Xq22	CfaX	
Pyruvate kinase deficiency	<i>PK1</i>	1q21-q22		
Progressive retinal atrophy	<i>PDE6B</i>	1p31		
von Willebrand disease	<i>VWF</i>	12p13.3-p13.2	CfaL11	
X-linked severe combined immunodeficiency disorder	<i>IL2RG</i>	Xq13.1	CfaX	
Congenital stationary night blindness	<i>RPE65</i>	1p31		
MPSVII	<i>GUSB</i>	7q21.11		
Leucocyte adhesion deficiency (LAD)	<i>ITGB2</i>	21q22.3		
Narcolepsy	<i>HCRTR2</i>	6p21	Cfa12	

Mapping Agricultural Mammals

Aggressive genome projects on agriculturally important animal species [cattle, pig, sheep, horse, and chicken (Table 1)] have already yielded powerful tools for assessing genes that specify hereditary disorders; infectious disease resistance; breed-specific quantitative trait loci (QTLs); and phenotypes of agricultural relevance, termed economic trait loci (ETLs) (41). Genetic identification and tracking of ETLs in animal pedigrees have considerable import for livestock improvement and production. Early maps for farm animals emphasized type II STR markers, because they were informative in pedigrees segregating disease or ETLs (12, 42). It soon became clear that the addition of type I comparative anchor loci that provided cross-reference to the gene-rich mouse and human maps was not only helpful but critical (6, 8). The benefit comes because type I-based anchor maps allow "comparative candidate positional cloning" (41), a three-step gene identification strategy that (i) assesses the linkage of an animal variant phenotype to a specific chromosomal position; (ii) identifies responsible candidate genes in that region by inspecting the homologous region of the human and mouse gene maps using comparative anchor (type I) loci as landmarks to demarcate chromosomal regions; (iii) identifies and genotypes type III SNP markers in or around the candidate loci and tests for association (or not) of these with the phenotype.

Powerful demonstrations of this approach that prove the principle have been achieved for ETLs and QTLs in cattle, sheep, and pigs [(43) and Table 2]. The muscular hypertro-

phy (double muscling) trait was mapped with type II markers to bovine chromosome 2 (44). Comparative candidate positional cloning suggested *myostatin* as a candidate gene to Grobet *et al.* (45), who identified an 11-bp deletion responsible for the trait in Belgian Blue cattle. In sheep, the Spider Lamb syndrome, or ovine hereditary chondrodysplasia, is a Mendelian recessive trait common to several breeds in North America, Australia, and New Zealand. Mapped to the distal end of ovine chromosome 6, comparative map inspection showed that homologous segments on human chromosome 4p16.3 and mouse chromosome 5.15 included the fibroblast growth factor receptor 3 (*FGFR3*) locus (46). Reports that human mutations and mouse knockouts of *FGFR3* showed skeletal deformities similar to Spider Lamb syndrome prompted a mutation search in over 1000 affected sheep that implicated causative mutations for Spider Lamb syndrome in the ovine *FGFR3* gene (46). In pigs, the ryano-dine receptor gene (*RYR1*) carries alleles that confer a stress syndrome analogous to malignant hypothermia in humans, caused by a variant of the homologous gene (47). Clearly, the strength of the comparative genetic approach is derived from connecting the dense gene-rich maps of human and mouse with the agricultural species' maps, in spite of the fact that livestock map densities are 100 to 1000 times lower than human or mouse map densities (see Table 1).

Mapping Cats and Dogs and Outgroups

Genome projects targetting the domestic dog and cat benefit from special genomic advantages for companion animals, including thousands of years of domestication (estimated to be at least 15,000 years for dogs and 7000 years for cats) driven by artificial selection (48). Over 33 cat and 400 dog breeds contribute enormous morphological, behavioral, and phenotypic variation that is documented and segregated among established purebred lineages. Furthermore, the veterinary profession provides trained clinical observers who are documenting breed-specific biomedical conditions (heredity and infectious and degenerative disease)—a medical surveillance second only to that of humans. Some 364 genetic diseases have been described in dogs (49) and over 200 in cats (50). Cancer registries exist for both cats and dogs, arising from genetic, viral (such as feline leukemia virus), and environmental origins (49, 50).

Moderate-level gene maps of both cat and dog have been developed this year, including integrated type I and II markers (19, 51, 52). Several mapping pedigrees for dogs, an interspecies backcross for cats (with an Asian leopard cat, *Prionailurus bengalensis*), and radiation hybrid and arrayed BAC panels her-

ald the onset of comparative genomic applications for both species. Eighteen heritable canine maladies and seven feline genetic diseases have been attributed to genes homologous to human disease gene mutations (Table 2).

A compelling illustration of the potential emerged from a long-term search for the gene for narcolepsy, a debilitating sleep disorder that causes afflicted dogs and people to irresistibly fall asleep as the sufferer is walking, talking, or simply excited (53). The Doberman Pinscher dog breed, artificially selected for guardian attributes over 100 years ago, segregates the malady as an autosomal recessive trait that was linked by type II markers to a region of canine chromosome 12 that is homologous with human chromosome 6p21. A BAC contig (overlapping ordered clones) of the canine genomic region was used to narrow the responsible locus by pedigree analysis of BAC-derived type II markers. A large (226 bp) SINE retroelement insertion/disruption caused abnormal splicing within the *hypocretin (orexin) receptor-2 (Hcrtr2)* locus in the region and was invariably associated with the Doberman's narcolepsy. It was discovered that a different mutation in the same gene (122-bp deletion) caused narcolepsy in Labrador retrievers. Mouse knockouts deficient in hypocretin, the ligand of *Hcrtr2* gene products, also have sleeping disorders, providing strong affirmation for the etiology of narcolepsy via hypocretin G-protein signaling in the brain stem and basal forebrain. At this writing, no human variants of the hypocretin receptor gene have been reported, but the implication of the existence of a hypocretin pathway by comparative inference is provocative.

Gene mapping in species other than rodents, livestock, and pets has been limited, largely because genome projects are costly. A chimpanzee genome project is now developing to resolve the differences between humans and their closest nonhuman relatives (54), whereas other primate gene maps (baboon and macaque) have been initiated to apply genetic assessment to these valuable primate animal models for behavior, vaccine development, and numerous genetic diseases (55). Some species' gene maps (for example, shrews, marsupials, zebrafish, and chicken) have a large potential for phylogenetic informativeness because of their great genetic distance from that of humans (5, 6, 21, 56–58). For instance, comparative mapping of human X- and Y-borne genes in marsupials and monotremes has shed light on the origin and divergence of mammalian sex chromosomes (56, 59). Comparative inference played a key role in implicating Y chromosome genes controlling sex determination and spermatogenesis (60, 61). The autosomal location of the *ZFY* orthologs in marsupials and monotremes was the first evidence that this was not the sex-determining gene, whereas the Y

location of marsupial *SRY* was consistent with its primary role in stimulating male-specific differentiation (60). More recent analyses of marsupial homologs of *SRY* and the candidate spermatogenesis gene *RBMY* demonstrated that both genes evolved from more widely expressed genes located on the X chromosome (59, 61). The construction of dense type I maps of chicken and zebrafish allows an assessment of the extent of conserved synteny (the linkage or chromosomal association of two or more gene homologs in maps of compared species) over a 450-million-year interval (21, 58). Although most of these homology blocks are usually short, a number of longer conserved segments surely indicates powerful selective constraints on clustered gene reorganization [for example, zebrafish LG9 and LG19 have long stretches of conserved synteny homologous to human chromosomes 2q and 7p, respectively (see foldout)].

The Mammalian Genome Radiations

A still-unfulfilled promise of comparative biology is a unified view of the evolutionary divergence and origin of mammalian species. In a time when collective syntheses of paleontologic, morphologic, and molecular sequence data struggle to identify ancient splits (62, 63), genome comparisons among mammal taxa provide unusually powerful phylogenetic characters in gene marker-defined chromosome segment exchanges. Because they are nearly always unique in chromosomal position and in most cases exceedingly rare, chromosomal rearrangements offer a large cadre of cladistic characteristics (that is, changes so unusual that they are likely to have occurred only once), which combine the advantages of previous molecular and morphological evolutionary tacks.

To appreciate this potential, it is useful first to describe the patterns of genome evolution in mammals that have been encountered. By comparing the conserved syntenies revealed by gene maps and chromosome painting, two very different rates of genome rearrangements have been observed. A high degree of genomic conservation is the predominant mode for the mammalian genome. In primates, only a handful of differences are apparent between the genomes of humans, great apes, the Old and New World primates, and lemurs (5–7, 23). Recognizable chromosomal exchanges are so infrequent as to allow reconstruction of the genomic arrangement of a common primate ancestor and the steps leading to modern species' genome disposition (23, 64). As few as seven translocation steps discriminate the hypothetical primate ancestor (estimated to have existed over 60 million years ago) from human genome organization. The order Carnivora also displays extensive genome conservation among families (cats, seals, weasels, and racoons) as

compared with primates. The cat's genome organization can be reorganized to the human status by as few as 13 translocation steps (Table 1) (51). Such extreme genome conservation results from an exceedingly slow rate of exchange that is also observed in genome comparisons between human and other Carnivora families [seal (Pinnipedia) and mink (Mustelidae)] (65) or between human, Carnivora, and Artiodactyl families [cow, sheep (Bovidae), and pig (Suidae)] (66). This computes to a very slow rate of chromosomal change among eutherian mammals: about one to two exchanges every 10 million years. This slow rate explains why multiple chromosomes or chromosome arms are preserved intact across the long divergence times separating orders of mammals (see foldout).

Dramatic exceptions to the slow rate of exchange occur in nearly every mammalian order, where abrupt global genome rearrangement episodes have led to 3- to 10-fold shuffles in genome organizational structure. Among primates, gibbons and siamangs have genomes that are rearranged three to four times more extensively than those of human or great apes (67), as do certain New World primates (owl and spider monkeys) and lemurs (68). Among carnivores, members of the dog family Canidae have appreciably rearranged genomes relative to the ancestral carnivore organization, with chromosome numbers ranging from 2N = 36 in the red fox to a high of 2N = 78 in the wolf and domestic dog (69). The Carnivora family Ursidae (bears) and the Cervidae (deer) family of Artiodactyla also display global genomic shuffles, whereas other families in these orders show the conserved slow rate of exchange (5-7, 66-70). Rodent species, particularly mouse and rat, show the more rapid pattern of chromosome change, with some 180 conserved segments shared between human and mouse and 109 shared between human and rat (17, 71) (Table 1). In sum, the mammal radiations generally display a slow rate of chromosome exchange (one to two exchanges per 10 million years) that is infrequently punctuated in certain lineages by episodes of global genome reorganization. The reason for these periodic abrupt global shuffles is an unsolved puzzle of this field.

Applying genomic exchanges as informative phylogenetic characters requires an understanding of the resolving power of different map methods to reveal conserved segments that occur and how they were derived. Reciprocal chromosome painting can reveal conserved homology segments but does not reveal interstitial inversions that can alter the order of genes within conserved segments. For example, the human and mouse X chromosomes retain the same genes, but the relative orders of gene homologs have been rearranged by inversions into at least six ho-

mology segments (Fig. 1). In contrast, alignment of gene orders discerned from parallel RH mapping of cat and human (72) shows that the feline and human gene order are identical (Fig. 1). These observations reinforce the impression that cat and human genome organizations are close to the ancestral version for their respective orders and perhaps for mammals in general, because similar genome-wide conservation is also apparent in whole eutherian genome comparisons of human/cat with Cetartiodactyla (cow and sheep), Perissodactyla (horse), Chiroptera (bat), and Eulipotyphla (shrew) (Table 1 and foldout). However, cryptic intrachromosomal inversions are also common in autosomes of compared mammal genomes (73) and need to be factored into more definitive phylogenomic reconstructions.

Once we can discriminate between slow and rapid patterns, it becomes feasible to assess different types of genomic exchange (such as fission, translocation, inversion, and transposition) to estimate their frequencies and to apply phylogenetic principles to their genomic reorganizations. Several methods for identifying conserved segments and as-

sessing them have been attempted, but this process is just beginning (71). A preliminary example of the analytical process combining data from gene maps, chromosome painting, cytogenetic banding homology, and molecular phylogeny is illustrated for the carnivores in Fig. 2. The previously determined topology is supplemented with demonstrable genomic exchanges (fusions, fissions, and inversions) that are postulated to have occurred from genomic comparisons (70). Briefly, the ancestral genome of primates and carnivores (and perhaps of eutherian mammals) was a low-numbered, largely metacentric genome (2N ~ 40 to 50) that evolved at the slow rate to human (11 steps), to cat (6 steps), to mink (10 steps), and to seal (8 steps). The modern Ursidae family includes eight species whose genomes are highly rearranged, mostly through 19 fissions and seven inversions of the ancestral genome that persist today as a (2N = 74) largely acrocentric karyotype in six Ursinae bear species (Fig. 2). The early global fissioning was subsequently followed by reorganization (through centromere fusions), once in the recent ancestry of *Ailuropoda melanoleuca*, the giant panda, and once more several

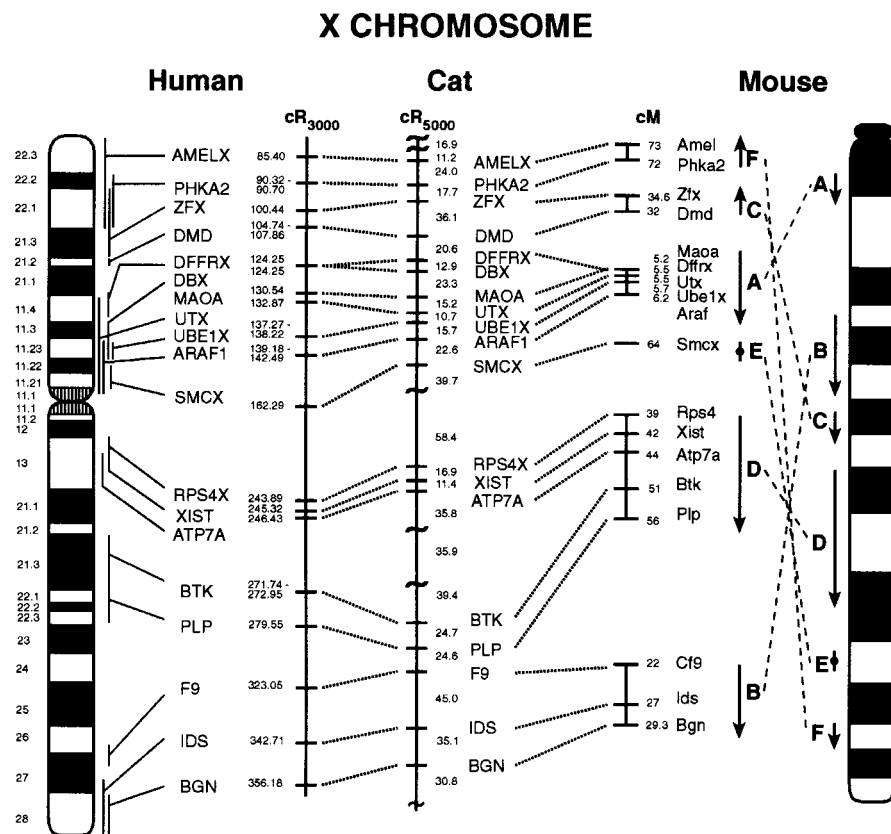


Fig. 1. Comparison of the relative order of X chromosome type 1 coding gene homologs between human, feline, and mouse X chromosomes shows six rearranged (by intrachromosomal inversion) segments conserved between mouse and human or between mouse and cat. The same genes have an identical order across the entire feline and human X chromosomes (72). ~ indicates type II STRs used to build the integrated cat map (19, 51, 72). Arrows indicate the polarity of mouse gene order; for example, toward the chromosome terminus in mouse.

million years later on a lineage leading to *Tremarctos ornatus*, the South American spectacled bear. The global shuffling episodes produced about 56 steps (including 12 intrachromosomal inversions revealed by G-band alignments) for the carnivore ancestor to reach the giant panda, 45 to reach the spectacled bear, and 34 to the ursine bears. These events are unique, cumulative, and useful in recapitulating the molecular basis of species divergence and the steps taken to assemble the genomes of modern species.

The hierarchal phylogeny among the mammal orders, dating back to at least 60 to 100 million years ago, is not yet resolved. The foldout included with this issue of *Science* presents an amalgam of opinions about placental-eutherian orders that are supported by some but not all data examined to date. For example, the superordinal group Glires, which associates rodents and hares, is strongly supported by a number of unusual morphological traits and fossils (63, 74, 75), but numerous molecular studies of nuclear and mitochondrial genes do not support the association (62, 76). In contrast, multiple molecular studies have aligned the superorder Afrotheria (six orders derived from Africa, including hyraxes, elephants, elephant shrews, sea cows, armadillos, golden moles, and tenrecs), but strong morphological association of these orders is not apparent (63, 75, 76). The fossil record of only one group (Lipotyphyla—a polyphyletic Insectivora suborder, which includes Eulipotyphyla and Afrosoricida) extends to before the Cretaceous-Tertiary boundary (65 million years ago) (63), a

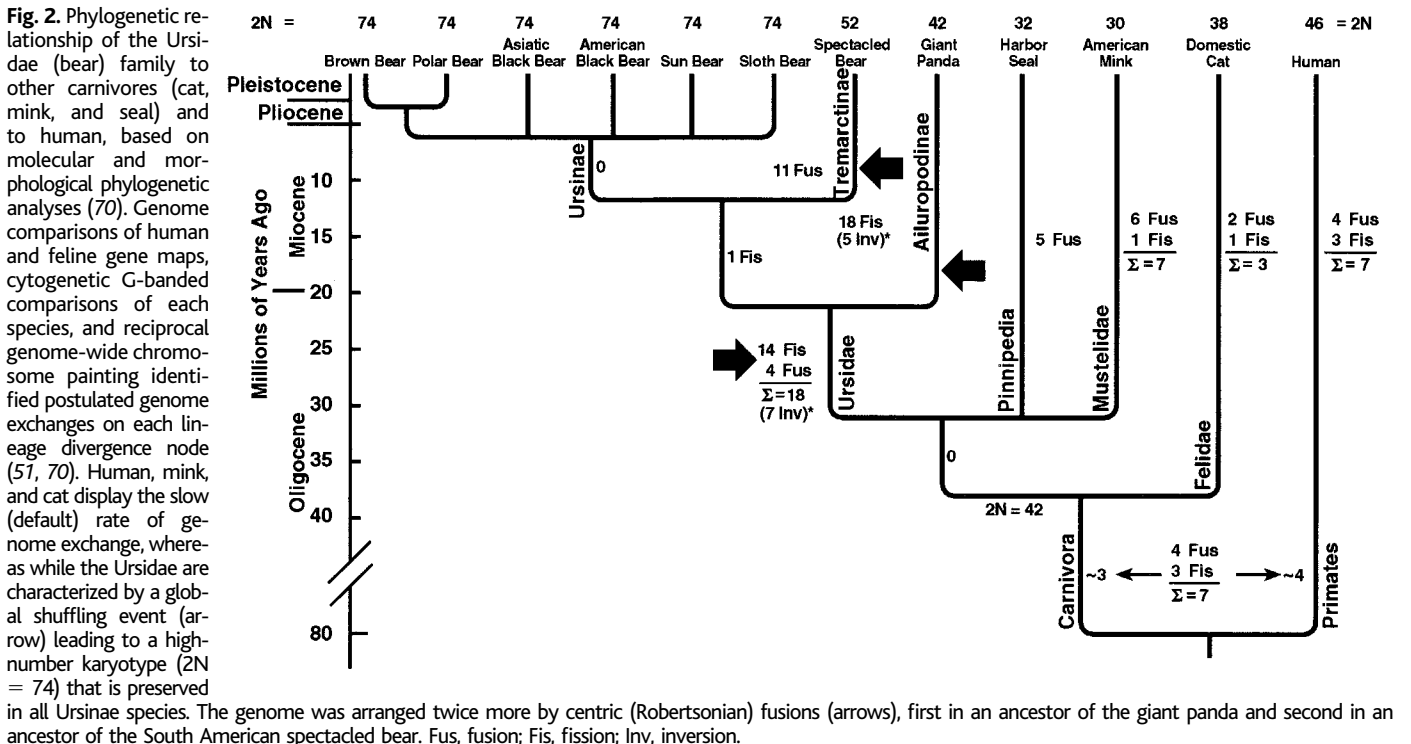
finding that persuaded some paleontologists that the common ancestor of modern mammals is not much older (75); however, molecular analyses prescribe a somewhat earlier ancestry, between 100 and 120 million years ago (76, 77). The conflicts in resolving relationships among mammal orders arise from both their contemporaneous divergence and the great age of these events. Finding phylogenetic characters that date back this far is not trivial; most dramatic morphological adaptations and molecular gene sequence changes are too recent to be informative for associating mammal orders. The default slow chromosome exchange events may be useful here precisely because they are remarkably slow, distinctive, heritable, and ancient.

There are cautionary notes that should be mentioned. First, comparative genomic data are only available for 11 of 28 mammalian orders, although attempts to map several unrepresented orders are beginning. There is a strong imperative for developing moderate-resolution comparative gene maps for representative species for each of these orders to fulfill the promise outlined here. Second, the least ambiguous species to choose for comparative analyses would be ones that display the slower exchange rate (for example, the common shrew; order Eulipotyphla), but for most species this rate is unknown and is not so obvious from chromosome number. Third, available estimators for quantifying genome exchange ignore inversions, and chromosome painting may not reveal it. To date, only a few species (human, mouse, rat, pig, sheep, cat, and bovid) have meiotic or RH

gene ordered maps, which are required for confident “phylogenomic” reconstruction. Finally, the causes of genome exchanges, their dichotomous rates, and the reasons for species-level fixation are not well understood, a caveat that would affect the assumptions of phylogenetic analysis. None of these aspects are fatal, but considering the limitations will be critical in applying genome differences to the evolutionary history of mammalian orders.

Conclusions

Until recently, comparative genomics was a cottage industry overshadowed by genetic advances in human and model organisms. Improved technologies and the potential for valuable applications have put the prospect of dense gene maps of domesticated livestock and companion animal species within our reach (Table 1). Some immediate practical applications of these maps that we envision include: (i) supplying animal models for human genetic diseases based on explicit gene homology as monitors for pathogenesis and therapy; (ii) an opportunity to identify candidate polygenes that affect human and veterinary disease; (iii) assessing multifactorial characters and pathologies; (iv) the discovery of evolved adaptations in mammal species that ameliorate maladies homologous to human hereditary and infectious diseases as a prelude to gene therapy, a concept termed genomic prospecting (78); (v) developing treatments for veterinary pathologies based on human trials for homologous gene defects; and (vi) the prospect of building fatter pigs,



finer wool, leaner beef, more tasty chickens, or faster racehorses. Hope for each of these potential applications is growing in our community every day.

There are additional ambitious expectations with regard to basic biology. Among them are (i) the hope of explaining the physical clustering of gene families (such as the major histocompatibility complex, immunoglobulin genes, *Hox* genes, the T cell receptor cluster, and chemokine receptors) as adaptive combinations of coordinate *cis* regulation, gene editing, or selective retention; (ii) the chance to understand whether even longer linkage associations preserved for tens of millions of years through billions of individuals in thousands of species are merely "frozen accidents" or were selectively retained by developmental or functional dependence (79); (iii) the opportunity to resolve the 100- to 150-million-year-old phylogeny of mammal orders using genomic segment exchanges as phylogenetic characters; (iv) the discovery of precipitous genomic events, such as the invasion of endogenous retrovirus families, preserved in modern genomics as molecular fossils of ancient epidemics; (v) the application of gene maps to nondomestic species, offering biologists the mapping tools to identify genetic determinants of reproductive isolation, adaptation, survival, and species formation.

One day soon, sequencing centers will begin to target mammals beyond humans and mice. The comparison of full genome sequences offers opportunities to identify gene birth and death in mammal lineages (for example, chimpanzee versus human) (54) as has already been approached with compared prokaryote genomes (25). The promise of comparative genomics for mammals extends further than we can imagine, as few biological disciplines will not be enhanced by knowledge of the natural history of the genes that make up living forms.

References and Notes

1. J. Qiang et al., *Nature* **398**, 326 (1999).
2. D. E. Wilson and D. M. Reeder, Eds., *Mammal Species of the World: A Taxonomic and Geographic Reference* (Smithsonian Institution Press, Washington, DC, ed. 2, 1993); R. M. Nowak, *Walker's Mammals of the World* (Johns Hopkins Univ. Press, Baltimore, MD, ed. 5, 1999).
3. F. S. Collins et al., *Science* **282**, 682 (1998); E. Marshall, *ibid.* **284**, 1439 (1999).
4. J. A. Marshall Graves and J. L. Vandeberg, *ILAR J. Comp. Gene Mapping* **39**, 47 (1999); S. J. O'Brien, *Genetic Maps: Locus Maps of Complex Genomes* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 6, 1993).
5. S. J. O'Brien, H. N. Seunaz, J. C. Womack, *Ann. Rev. Genet.* **22**, 323 (1988).
6. Comparative Genome Organization: First International Workshop, Fraser Island, Australia, *Mamm. Genome* **7**, 717 (1996); M. J. Wakefield and J. A. M. Graves, *ILAR J.* **39**, poster (1998); R. Glas et al., *ibid.*
7. J. A. Marshall Graves, *ILAR J.* **39**, 48 (1998);
8. S. J. O'Brien, *Curr. Opin. Genet. Dev.* **1**, 105 (1991); _____ and J. A. M. Graves, *Cytogenet. Cell Genet.* **58**, 1124 (1991).
9. D. G. Wang et al., *Science* **280**, 1077 (1998); M. Cargill et al., *Nature Genet.* **22**, 231 (1999); M. K. Halushka et al., *ibid.*, p. 239; F. C. Collins et al., *Science* **278**, 1580 (1999); R. J. Lipshutz et al., *Nature Genet.* **21**, 20 (1999); K. H. Buetow et al., *ibid.*, p. 323.
10. E. Masood, *Nature* **398**, 545 (1999); D. Housman and F. D. Ledley, *Nature Biotechnol.* **16**, 492 (1998); P. W. Kleyn and E. S. Vesell, *Science* **281**, 1820 (1998).
11. N. G. Copeland et al., *Science* **262**, 57 (1993); V. M. Chapman, in *Origins of Inbred Mice*, H. C. Morse III, Ed. (Academic Press, San Diego, CA, 1978), pp. 555-568; M. T. Davison et al., *ILAR J.* **39**, 96 (1998).
12. A. M. Crawford et al., *Genetics* **140**, 703 (1995).
13. M. J. DeGortari et al., *Mamm. Genome* **9**, 204 (1998).
14. S. M. Galloway et al., *Genome Res.* **6**, 667 (1996); M. L. Tate et al., *J. Hered.* **88**, 36 (1997); P. D. Muir et al., *ibid.*, p. 366; P. K. Riggs et al., *ibid.*, p. 373; L. Andersson, *ibid.*, p. 380.
15. P. Deloukas et al., *Science* **282**, 744 (1998); J. Weissenbach, *Clin. Chem. Lab. Med.* **36**, 511 (1998); C. Dib et al., *Nature* **380**, 152 (1996).
16. M. Marra et al., *Nature Genet.* **21**, 191 (1999); M. D. Adams et al., *Nature* **377**, 3 (1995); L. Hiller et al., *Genome Res.* **6**, 807 (1996); W. F. Dietrich et al., *Nature* **380**, 149 (1996).
17. T. K. Watanabe et al., *Nature Genet.* **22**, 27 (1999); R. G. Steen et al., *Genome Res.* **9**, 1 (1999); M. R. James and K. Lindpainter, *Trends Genet.* **13**, 171 (1997).
18. W. J. Van Etten et al., *Nature Genet.* **22**, 385 (1999).
19. W. J. Murphy et al., *Genomics* **57**, 1 (1999).
20. Y.-P. Yang and J. E. Womack, *Genome Res.* **8**, 731 (1998); J. E. Womack et al., *Mamm. Genome* **8**, 854 (1997).
21. J. H. Postlewait et al., *Nature Genet.* **18**, 345 (1998); N. Shimoda et al., *Genomics* **58**, 219 (1999); M. A. Gates et al., *Genome Res.* **9**, 334 (1999).
22. P. Y. Woon et al., *Genomics* **50**, 306 (1998); R. Li et al., *ibid.* **58**, 9 (1999); B. Zhu et al., *Mamm. Genome* **10**, 706 (1999); K. Osoegawa et al., *Genomics* **52**, 1 (1998); H. K. Al-Bayati et al., *Mamm. Genome* **10**, 569 (1999); C. T. Amemiya et al., *Methods Cell Biol.* **60**, 235 (1999); C. Nusbaun et al., *Nature Genet.* **22**, 388 (1999). See also <http://www.resgen.com>
23. J. Wienberg and R. Stanyon, *Curr. Opin. Genet. Dev.* **7**, 784 (1997); M. A. Ferguson-Smith et al., *ILAR J.* **39**, 68 (1998); J. Wienberg et al., *ibid.*, p. 77.
24. A. Paterson, et al., *Nature Genet.* **14**, 380 (1996); M. D. Gale and K. M. Devos, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1971 (1998); J. L. Bennetzen et al., *ibid.*, p. 1975.
25. B. Snel et al., *Nature Genet.* **21**, 108 (1999).
26. Y. Zhang et al., *Nature* **372**, 424 (1994); J. M. Friedman and J. L. Helaas, *ibid.* **395**, 763 (1998); J. M. Friedman and R. L. Leibell, *Cell* **69**, 217 (1992); J. K. Naggert et al., *Nature Genet.* **10**, 135 (1995); P. W. Kleyn et al., *Cell* **85**, 281 (1996); K. Noben-Trauth et al., *Nature* **380**, 534 (1996); C. T. Montague et al., *ibid.* **387**, 903 (1997); K. Clement et al., *ibid.* **392**, 398 (1998).
27. D. L. Nagle et al., *Nature* **398**, 148 (1999); T. M. Gunn et al., *ibid.*, p. 152.
28. J. P. Rapp et al., *Science* **243**, 542 (1989); P. Hilbert et al., *Nature* **353**, 521 (1991); H. J. Jacob et al., *Cell* **67**, 213 (1991); A. Bottger et al., *J. Clin. Invest.* **98**, 856 (1996).
29. I. S. Feroogi et al., *N. Engl. J. Med.* **341**, 879 (1999); M. Rosenbaum and R. L. Leibell, *ibid.*, p. 913.
30. R. Reeves et al., *Nature Genet.* **11**, 177 (1995).
31. Mouse Genome Database (MGD), Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine (available at www.informatics.jax.org/); J. A. Blake et al., *Nucleic Acids Res.* **27**, 95 (1999); M. F. Lyon and A. G. Searle, Eds., *Genetic Variants and Strains of the Laboratory Mouse* (Oxford Univ. Press, Oxford, 1990).
32. B. A. Premack and T. J. Schall, *Nature Med.* **2**, 1174 (1996); P. M. Murphy, *Cytokine Growth Bact. Rev.* **7**, 147 (1996); *Annu. Rev. Immunol.* **12**, 593 (1994).
33. M. P. D'Souza and V. A. Harden, *Nature Med.* **2**, 1293 (1996); D. R. Littman, *Cell* **93**, 677 (1998); E. A. Berger et al., *Annu. Rev. Immunol.* **17**, 657 (1999).
34. M. P. Martin et al., *J. Acquired Immune Defic. Syndr.* **19**, 430 (1998); V. Alvarez et al., *Hum. Genet.* **102**, 483 (1998); O. J. Cohen et al., *J. Virol.* **72**, 6215 (1998).
35. M. Dean et al., *Science* **273**, 1856 (1996); Y. Huang et al., *Nature Med.* **2**, 1240 (1996); S. Mummidi et al., *ibid.* **4**, 786 (1998); P. A. Zimmerman et al., *Mol. Med.* **3**, 23 (1997); S. J. O'Brien and M. Dean, *Sci. Am.* **277**, 44 (September 1997); G. T. Nguyen et al., *J. Infect. Dis.*, in press.
36. Y.-R. Zou et al., *Nature* **393**, 595 (1998); K. Tachibana et al., *ibid.*, p. 591.
37. Y. Zhou et al., *J. Immunol.* **160**, 4018 (1998).
38. J. S. Cairns and M. P. D'Souza, *Nature Med.* **4**, 563 (1998); J. P. Moore, *Science* **276**, 51 (1997).
39. J. Battey et al., *Nature Genet.* **21**, 73 (1999); W. Dove and D. Cox URL: <http://www.nih.gov/welcome/director/reports/mgenome.htm> (1998).
40. S. Mazeyrat et al., *Hum. Mol. Genet.* **7**, 1713 (1998).
41. J. E. Womack, *ILAR J.* **39**, 153 (1998); J. Womack and S. R. Kata, *Curr. Opin. Genet. Dev.* **5**, 725 (1995).
42. W. Barendse et al., *Nature Genet.* **6**, 227 (1994); M. D. Bishop et al., *Genetics* **136**, 619 (1994); G. Rohrer et al., *ibid.* **136**, 231 (1994); H. Ellegren et al., *Genomics* **17**, 599 (1993).
43. L. Anderson et al., *Science* **236**, 1771 (1994); M. Georges et al., *Nature Genet.* **4**, 206 (1993); M. Georges et al., *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1058 (1993); M. Georges et al., *Genetics* **139**, 907 (1995).
44. C. Charlier et al., *Mamm. Genome* **6**, 788 (1995).
45. L. Grobet et al., *Nature Genet.* **17**, 71 (1997).
46. N. E. Cockett et al., *Mamm. Genome* **10**, 35 (1999); M. K. Webster and D. J. Donoghue, *Trends Genet.* **13**, 178 (1997).
47. J. Fujii et al., *Science* **253**, 448 (1991); H. Geldermann et al., *Genetics* **113**, 381 (1996).
48. E. A. Ostrander et al., *Trends Genet.*, in press; R. K. Wayne and E. A. Ostrander, *Bioessays* **21**, 247 (1999); R. Robinson, *Genetics for Cat Breeders* (Pergamon, Oxford, UK, ed. 3, 1991); S. J. M. Davis, in *Foilles Récentes à Khirkotia (Chypre)*, A. Le Brun, Ed. (Editions Recherche sur les Civilisations, Paris, 1983-1986), pp. 189.
49. D. F. Patterson, *Canine Genetic Disease Information System—A Computerized Knowledgebase of Genetic Diseases of the Dog* (Mosby Yearbook, St. Louis, MO, 1999).
50. F. W. Nicholas, *Veterinary Genetics* (Clarendon, Oxford, UK, 1987); _____ and P. R. Le Tissier, Online Mendelian Inheritance in Animals (OMIA) Database, available at www.angis.oz.au/Databases/BIRX/omia/; W. D. Hardy, M. Essex, A. J. McClelland, *Feline Leukemia Virus* (Elsevier/North Holland, New York, 1980); A. Hochhaus et al., *Eur. J. Med. Res.* **3**, 216 (1999); C. E. Green, Ed., *Infectious Diseases in the Dog and Cat* (Saunders, Philadelphia, ed. 2, 1998).
51. S. J. O'Brien et al., *Trends Genet.* **13**, 393 (1997); J. Wienberg et al., *Cytogenet. Cell Genet.* **77**, 211 (1997); M. Menotti-Raymond et al., *Genomics* **57**, 9 (1999); W. J. Murphy et al., in preparation; S. J. O'Brien et al., *J. Hered.* **88**, 408 (1997); G. Rettenberger et al., *Chromosome Res.* **3**, 479 (1995).
52. M. W. Neff et al., *Genetics* **151**, 803 (1999); C. Priat et al., *Genomics* **54**, 361 (1998); P. Werner et al., *Mamm. Genome* **10**, 814 (1999).
53. L. Lin et al., *Cell* **98**, 365 (1999); R. M. Chemelli et al., *ibid.*, p. 437.
54. A. Gibbons, *Science* **281**, 1432 (1998); E. H. McConkey and M. Goodman, *Trends Genet.* **13**, 350 (1997).
55. P. A. Morin et al., *Genomics* **53**, 12 (1998).
56. J. A. M. Graves, *Annu. Rev. Genet.* **30**, 233 (1996).
57. O. L. Serov et al., *ILAR J.* **39**, 195 (1998).
58. D. W. Burt and H. H. Cheng, *ibid.*, p. 229.
59. A. Pask and J. A. M. Graves, *Cell Mol. Life Sci.* **55**, 864 (1999); J. A. M. Graves, *Bioessays* **17**, 311 (1995); *ibid.* **20**, 264 (1998).
60. A. H. Sinclair et al., *Nature* **336**, 780 (1988); J. W. Foster et al., *ibid.* **359**, 531 (1992).
61. M. L. Delbridge et al., *Nature Genet.* **22**, 223 (1999).
62. P. J. Waddell et al., *Syst. Biol.* **48**, 1 (1999); *ibid.*, p. 31.
63. M. C. McKenna and S. K. Bell, *Classification of Mammals above the Species Level* (Columbia Univ. Press, New York, 1997); M. J. Novacek, *Nature* **356**, 121 (1992).
64. B. Dutrillaux, *Hum. Genet.* **48**, 251 (1979); S. Müller et al., *Chromosoma*, in press.
65. H. Hameister et al., *Chromosome Res.* **5**, 5 (1996); L. Froncice et al., *Chromosoma* **106**, 108 (1997).
66. B. P. Chowdhary et al., *Mamm. Genome* **7**, 297

- (1996); H. Hayes, *Cytogenet. Cell Genet.* **71**, 168 (1995); M. Johansson *et al.*, *Genomics* **25**, 682 (1995); L. Fronicke *et al.*, *Mamm. Genome* **7**, 285 (1996); A. Goureau, *Genomics* **36**, 252 (1996); G. Rettenberger *et al.*, *ibid.* **26**, 372 (1995).
67. A. Jauch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8611 (1992); U. Koehler *et al.*, *Am. J. Phys. Anthropol.* **97**, 37 (1995); U. Koehler *et al.*, *Genomics* **30**, 287 (1995).
68. M. A. Moreschalchi *et al.*, *Chromosome Res.* **5**, 527 (1997); S. Muller *et al.*, *Cytogenet. Cell Genet.* **78**, 260 (1997); S. Consigliere *et al.*, *Chromosome Res.* **3**, 1 (1996).
69. R. K. Wayne *et al.*, *Cytogenet. Cell Genet.* **44**, 123 (1987); *ibid.*, p. 134.
70. W. Nash *et al.*, *ibid.* **83**, 182 (1998); W. G. Nash and S. J. O'Brien, *ibid.* **45**, 206 (1987); S. J. O'Brien *et al.*, *Nature* **317**, 140 (1985); D. Goldman, P. Rathna Giri, S. J. O'Brien, *Evolution* **43**, 282 (1989); J. Pecón Slattery and S. J. O'Brien, *J. Hered.* **86**, 413 (1995).
71. J. H. Nadeau and D. Sankoff, *Trends Genet.* **14**, 495 (1998); R. W. DeBry and M. F. Seldin, *Genomics* **33**, 337 (1996); J. Ehrlich, D. Sankoff, J. H. Nadeau, *Genetics* **147**, 289 (1997).
72. W. J. Murphy *et al.*, *Genome Res.*, in press.
73. L. Schibler, *ibid.* **8**, 901 (1999).
74. M. M. Miyamoto, *Mol. Phylogenet. Evol.* **6**, 373 (1996).
75. J. Shoshani and M. C. McKenna, *ibid.* **9**, 572 (1998); M. Foote *et al.*, *Science* **283**, 1310 (1999).
76. J. Gately *et al.*, *Syst. Biol.* **48**, 6 (1999); M. S. Springer *et al.*, *ibid.*, p. 65; F. R. Liu and M. M. Miyamoto, *ibid.*, p. 54; M. S. Springer *et al.*, *Nature* **388**, 61 (1997); M. J. Stanhope *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9967 (1998).
77. S. Kumar and S. B. Hedges, *Nature* **392**, 917 (1998); L. Bromham, M. J. Phillips, D. Penny, *Trends Ecol. Evol.* **14**, 113 (1999).
78. S. J. O'Brien, *Nature Med.* **1**, 742 (1995).
79. S. Ohno, *Nature* **244**, 259 (1973).
80. O. L. Serov *et al.*, *ILAR J.* **39**, 189 (1998).
81. T. E. Broad *et al.*, *ibid.*, p. 160.
82. L. Marklund *et al.*, *Anim. Genet.* **27**, 255 (1996); G. A. Rohrer *et al.*, *Genome Res.* **6**, 371 (1996); Comparative map of the pig, available at www.toulouse.inra.fr/lgc/pig/compare/compare.html; <http://www.ri.bbsrc.ac.uk/pigmap/pigbase/loclist.html>.
83. G. Lindgren, *Genome Res.* **8**, 951 (1998); E. Bailey and M. M. Binns, *ILAR J.* **39**, 171 (1998).

REVIEW

Epigenetics: Regulation Through Repression

Alan P. Wolffe^{1*} and Marjori A. Matzke^{2*}

Epigenetics is the study of heritable changes in gene expression that occur without a change in DNA sequence. Epigenetic phenomena have major economic and medical relevance, and several, such as imprinting and paramutation, violate Mendelian principles. Recent discoveries link the recognition of nucleic acid sequence homology to the targeting of DNA methylation, chromosome remodeling, and RNA turnover. Although epigenetic mechanisms help to protect cells from parasitic elements, this defense can complicate the genetic manipulation of plants and animals. Essential for normal development, epigenetic controls become misdirected in cancer cells and other human disease syndromes.

Epigenetics has an impact on many seemingly disparate areas of scientific enterprise (1). Even a partial listing would include areas of applied science such as somatic gene therapy (2), cloning and transgenic technologies in plants and animals (3), aspects of cancer biology (4), the study of viral latency (5), the activity of mobile elements (6), genomic imprinting (7), and developmental abnormalities in plants (8) and animals (9). The excitement in this research area follows from the realization that diverse organisms have common molecular mechanisms that contribute to the epigenetic control of gene expression (10). A key element in many epigenetic effects concerns the recognition of nucleic acid sequence homologies at both the DNA and RNA levels. Exactly how this recognition occurs is presently unknown; however, DNA-DNA and RNA-RNA interactions can trigger gene silencing by inducing chromatin modifications and RNA degradation, respectively.

These homology-dependent gene-silencing events appear to be closely connected to genomic and cytoplasmic defense systems that protect cells against infiltration by invasive DNA and by RNA pathogens. We discuss the possible origins and the molecular mechanisms underlying both pathways of epigenetic silencing. It is important to recognize that these apparently distinct regulatory processes are linked not only in the end result of inactivating foreign nucleic acids but also potentially through interconnected mechanisms.

Epigenetic Control

Epigenetic control of gene expression can be considered from the standpoint of normal development, which requires stable repression of genes not required in specific cell types. Many epigenetic effects, however, are observed in unusual circumstances, and these have recently provided new insights into mechanisms. The discovery of epigenetic processes based on nucleic acid sequence recognition followed from the development of methods to introduce genes into the genomes of fungi, plants, and animals. Introduced transgenes often integrated as multiple copies or were identical to endogenous sequences. Contrary to expectations, the increased gene dosage did not result in enhanced expression but in gene silencing. Subsequent

work distinguished distinct nucleic sequence homology-based mechanisms that lead to transcriptional or posttranscriptional gene silencing (TGS and PTGS, respectively).

DNA-Based Mechanisms—The Power of Repeats

An important advance in epigenetics research has been the realization that interactions between repeated DNA sequences can trigger the formation and the transmission of inactive genetic states and DNA modifications. The source of this concept was influential work with two filamentous fungi, which provided precedents for how eukaryotes can treat redundant sequences by mechanisms involving the recognition of DNA repeats. The RIP (repeat-induced point mutation) phenomenon in *Neurospora crassa* and MIP (methylation induced premeiotically) in *Ascobolus immersus* result in the pairing-dependent modification of DNA sequence duplications during the sexual cycle of these organisms (11). These modifications protect the streamlined haploid genome from potentially deleterious recombination events and from the activity of endogenous transposable elements. During RIP, both linked and unlinked duplicated DNA sequences ranging in size from a few hundred to several thousand base pairs incur G-C to A-T transition mutations. For reasons that are not yet clear, RIP-modified sequences become substrates for de novo postreplicative enzymatic modification of DNA in which any remaining cytosines are converted to 5-methylcytosine. This extensive mutagenesis generates missense and nonsense codons that inactivate gene expression and also creates sequence divergence that can prevent homologous recombination. In MIP, sequence duplications become heavily methylated and silenced without mutation. Moreover, DNA methylation

¹Laboratory of Molecular Embryology, National Institute of Child Health and Human Development, NIH, Building 18T, Room 106, Bethesda, MD 20892-5431, USA. ²Austrian Academy of Sciences, Institute of Molecular Biology, Billrothstrasse 11, A-5020 Salzburg, Austria.

*To whom correspondence should be addressed: E-mail: awlme@helix.nih.gov (A.P.W.) and amatzke@server1.imolbio.oew.ac.at (M.A.M.)