29. E. W. Brown, J. E. LeClerc, B. Li, W. L. Payne, T. A. Cebula, *J. Bacteriol.* **183**, 1631 (2001).
30. E. Denamur *et al.*, *Cell* **103**, 711 (2000).
31. A. Oliver, R. Canton, P. Campo, F. Baquero, J. Blazquez, *Science* **288**, 1251 (2000).
32. A. L. Prunier *et al.*, *J. Infect. Dis.* **187**, 1709 (2003).
33. A. R. Richardson, I. Stojiljkovic, *Mol. Microbiol.* **40**, 645 (2001).
34. S. J. Weissman, S. L. Moseley, D. E. Dykhuizen, E. V. Sokurenko, *Trends Microbiol.* **11**, 115 (2003).
35. J. D. Boddicker, N. A. Ledeboer, J. Jagnow, B. D. Jones, S. Clegg, *Mol. Microbiol.* **45**, 1255 (2002).
36. We thank H. Merkert for excellent graphical assistance and R. Gross, W. Goebel, and T. Ölschläger for providing electron micrographs. Our work related to the topic is supported by the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie, and the Bavarian Research Foundation.

REVIEW

Structural Dynamics of Eukaryotic Chromosome Evolution

Evan E. Eichler[1]* and David Sankoff[2]

Large-scale genome sequencing is providing a comprehensive view of the complex evolutionary forces that have shaped the structure of eukaryotic chromosomes. Comparative sequence analyses reveal patterns of apparently random rearrangement interspersed with regions of extraordinarily rapid, localized genome evolution. Numerous subtle rearrangements near centromeres, telomeres, duplications, and interspersed repeats suggest hotspots for eukaryotic chromosome evolution. This localized chromosomal instability may play a role in rapidly evolving lineage-specific gene families and in fostering large-scale changes in gene order. Computational algorithms that take into account these dynamic forces along with traditional models of chromosomal rearrangement show promise for reconstructing the natural history of eukaryotic chromosomes.

Chromosomes evolve by the modification, acquisition, deletion, and/or rearrangement of genetic material. Defining the forces that have affected the eukaryotic genome is fundamental to our understanding of biology and evolution (species origin, survival, and adaptation). Chromosomal evolution includes a continuum of molecular-based events of greatly varied scope. For historical and methodological reasons, complete integration of these different levels of chromosomal structural change has not been practical. Evolutionary biologists have approached genome evolution from two different perspectives. The holistic view compared the number of chromosomes and the order of fragments (homologous segments) among closely and distantly related species by using genetic mapping tools and in situ methods (*1*). These studies provided a framework for understanding the nature and pattern of chromosomal rearrangement among eukaryotic species. However, because of limitations in resolution, these studies provided little insight into the underlying mechanisms responsible for such

changes, and they were not adequate for assessing less conserved regions. The alternate, reductionist perspective has focused on analysis corresponding to small blocks of DNA sequence. Through comparative sequencing among closely related species, considerable diversity of mutational events has been inferred. Such inferences, however, are restricted to regional analyses of DNA and, by their very nature, are limited.

With the advent of large-scale sequencing of eukaryotic genomes, a bridge connecting these two perspectives is emerging. Comparative analyses of complete genomes can provide a comprehensive view of large-scale changes in synteny, gene order, and regions of nonconservation while simultaneously affording exquisite molecular resolution at the level of single–base pair differences. Knowing the precise sequence at regions of rearrangement gives insight into underlying molecular mechanisms. New computational methods can be developed to effectively digest and model these vast quantities of data. As a result of this genomic revolution, novel approaches and insights into the patterns and mechanisms of both small- and large-scale chromosomal rearrangement are beginning to emerge.

To date, whole-genome sequence data are available for ~20 different eukaryotic genomes and an additional 50 are to be sequenced within the next 4 years (Table 1). The selected organisms (~20 fungal, 7 plant, and 35 animal genomes) represent

considerable breadth of eukaryotic evolutionary diversity but can hardly be viewed as representative. The primary motivation for the initial phase of complete-genome sequencing was not evolutionary biology, but rather medical, agricultural, and/or commercial relevance. Furthermore, small genomes (*Arabidopsis*, *Fugu*, *Tetraodon*) (*2*, *3*) have been favored over larger ones because of the still relatively prohibitive costs of whole-genome shotgun sequencing at $50 million to $100 million per 3-Gb genome. Despite this ascertainment bias, the available sequence has provided an unparalleled opportunity to investigate changes in the eukaryotic genome. Several important trends, as well as idiosyncrasies, regarding chromosomal evolution already have become apparent, particularly from comparisons of more closely related species.

Synteny: Fragile Versus Random Breakage Model?

In two eukaryotic genomes with a common ancestor, chromosome organization may be altered by intrachromosomal rearrangements (inversions) or reciprocal interchromosomal rearrangements (translocations) in one or the other lineage. In addition to these events, genetic material may become transposed into the DNA of one lineage or deleted, which disrupts the shared homologous segments. We denote by conserved synteny a number of sequence markers mapping to a single chromosome in each genome, irrespective of order. If the corresponding chromosomes also order these markers in the same way, they are said to constitute a conserved linkage group or a homologous segment. Nearly 20 years ago, Nadeau and Taylor argued that the distribution of breakpoints between homologous segments along the chromosomes of either species should be uniformly random (*4*). At a gross level of resolution, subsequent comparative mapping and sequencing studies among vertebrate species have, in general,

[1]Department of Genetics, Center for Human Genetics and Center for Computational Genomics, Case Western Reserve University School of Medicine and University Hospitals of Cleveland, Cleveland, OH 44106, USA. [2]Department of Mathematics and Statistics, University of Ottawa, 585 King Edward Avenue, Ottawa, ON Canada, K1N 6N5.

*To whom correspondence should be addressed. E-mail: eee@po.cwru.edu

upheld the apparent randomness of rearrangement (*3*, *5*, *6*).

The rates of chromosomal rearrangement vary radically among different lineages (*1*, *7*) and between sex chromosomes and autosomes. Among vertebrates, for example, rates of chromosomal rearrangement have been reported to range from two-tenths to one or two rearrangements per million years (*8*), whereas among invertebrate species, estimates rise precipitously, attaining seven and 50 rearrangements per million years (*9*, *10*). Differences in generation time and reproduction strategies between vertebrates and invertebrates readily account for much of the disparity between lineages. The sex chromosomes of eutherian mammals represent opposite extremes in the degree of conserved synteny. The X chromosome shows extensive conservation (Fig. 1), such that syntenic relationships can be readily defined between distant species (*11*). By contrast, the Y chromosome is a paradigm of rapid and unconstrained chromosomal evolution (*12*). The absence of recombination over most of the chromosome, rampant homology-mediated rearrangement, and the extraordinary degree of gene conversion among duplicated segments have led to a chromosome where gene order is scrambled rapidly and orthologous relationships quickly disintegrate between species (*12*, *13*).

Complete sequencing of genomes has confirmed the extensive levels of conserved synteny originally found by comparative mapping, but the high density of markers afforded by complete sequence also results in a more complicated view of chromosomal evolution, with remarkable levels of intrachromosomal rearrangement. Small local inversions appear to be preva-

lent within many eukaryotic lineages (*3*, *14–16*). Comparison of *Anopheles gambiae* and *Drosophila melanogaster* [species that diverged 250 million years ago (Ma)], as well as closer evolutionary comparisons within the genus *Anopheles*, show extensive reshuffling of gene order within chromosomes (*9*, *17*). Among ascomycetes yeast, small inversions have likewise contributed significantly to chromosomal evolution. Comparison of the finished genome of *Saccharomyces cerevisiae* and the shotgun sequence from *Candida albicans* provides an estimate of 1100 single-gene inversions since species divergence (140 to 330 Ma) (*15*).

The prevalence of short inversions represents a departure from the Nadeau-Taylor model in evidencing many pairs of closely spaced breakpoints. Another departure was inferred by Pevzner and Tesler in comparing human and mouse genomes (*18*, *19*). They found 281 synteny blocks (homology segments internally disrupted only by local micro-rearrangements) (Fig. 1) compared with the 180 known from comparative gene mapping (*4*). In trying to infer the evolutionary rearrangements responsible for this configuration, they found that the breakpoint regions between the synteny blocks would have had to have been disrupted an average of 1.9 times each, a high density of breakpoints over these small regions. This suggests an alternate model for chromosomal evolution, termed "fragile breakage" (*18*). When micro-rearrangements are considered (small inversions, deletions, or transpositions within what would otherwise be a conserved segment), the total number increases an order of magnitude to a few thousand, varying widely between chromosomal regions (*18*, *20*).

Centromeric and Telomeric Regions— Sites of Rapid Genomic Change

Centromeres and telomeres have long been recognized as peculiarly dynamic regions of chromosomal evolution. Both regions have posed particular problems during mapping and sequencing (*21*, *22*). Detailed sequencing and annotation of these areas remain limited even among finished genomes (*23–25*). Nevertheless, several important observations have been made regarding the evolutionary fluidity of these regions. For example, recent comparative fluorescent in situ hybridization studies among primates indicate that centromere position can change radically over short periods of evolutionary time without an obligatory alteration of closely flanking markers (*26*). The activation of evolutionary neocentromeres has been put forward as one possible explanation (*27*).

In most eukaryotes, centromeres and telomeres are composed of tandem arrays of repetitive sequence. The repetitive nature of these regions extends beyond the classically defined boundaries of centromeric and telomeric sequences to influence the frequency of structural rearrangements in surrounding regions. Transition regions, termed pericentromeric and subtelomeric DNA, are hotspots for the insertion or retention of repeat sequences. The nature of the repetitive DNA differs among organisms. In diverse eukaryotic genera, such as *Drosophila*, *Anopheles*, *Arabidopsis*, *Dictyostelium*, and rice, pericentromeric regions are reservoirs for the accumulation of a medley of lineage-specific transposable elements (see below) (*24*, *28–31*). Among primates, there is now overwhelming evidence that blocks of recently duplicated sequence populate subtelomeric and pericentromeric regions (*21*, *32–34*). Comparative studies of closely related pri-

Table 1. Census of sequence eukaryote genomes. A complete list of all finished and ongoing whole-genome sequencing projects is available (*63*). Repeat content and genome size are based on sequenced euchromatin. Haploid chromosome number reflects the number of bivalents. Gene annotation for most sequenced genomes is still ongoing. Therefore, gene number estimates are only a rough approximation. The table does not include organisms for which whole-genome sequence is available without an accompanying publication as of 23 April 2003. Asterisk indicates organisms for which only an entire chromosome sequence has been published.

Group	Species	Common	Size (Mb)	Chromosome (1N)	Gene no.	Repeat %
Mammal	*Homo sapiens*	Human	2900	23	30,000	46
Mammal	*Mus musculus*	House mouse	2500	20	30,000	38
Fish	*Takifugu rubripes*	Tiger pufferfish	400	22 (?)	30,000	<10
Urochordate	*Ciona intestinales*	Sea squirt	155	14	16,000	~10
Insect	*Anopheles gambiae*	Malaria mosquito	280	3	14,000	16
Insect	*Drosophila melanogaster*	Fruit fly	137	4	13,600	2
Nematode	*Caenorhabditis elegans*	Nematode worm	97	6	19,100	<1
Apicomplexa	*Plasmodium falciparum*	Human malaria parasite	23	14	5,300	<1
Apicomplexa	*Plasmodium yoelli*	Rodent malaria parasite	25	14	5,300	<1
Dictyosteliida	*Dictyostelium discoideum**	Social amoeba	34	6	2,800	<1
Protozoan	*Leishmania major**	Intracellular parasite	34	36	9,800	<1
Fungi	*Saccharomyces cerevisiae*	Brewer's yeast	12	16	5,700	2.4
Fungi	*Schizosaccharomyces pombe*	Fission yeast	13.8	3	4,900	0.35
Microsporidium	*Encephalitozoon cuniculi*	Intracellular parasite	2.5	11	2,000	<0.1
Angiosperm	*Arabidopsis thaliana*	Mustard weed	125	5	25,500	14
Angiosperm	*Oryza sativa*	Rice	400	12	32000–50000	?

mate species, as well as population studies, reveal dramatic quantitative and qualitative differences in the distribution and organization of these duplications (*34–36*). In light of this, it is perhaps not surprising that conserved synteny maps quickly deliquesce as centromere and telomere positions are approached (Figs. 1 and 2). Indeed, some of the "fragility" observed within human and mouse synteny maps corresponds to these evolutionarily dynamic areas of the genome.

These studies of human genome organization suggest frequent promiscuous nonhomologous exchanges during the course of chromosomal evolution. Nonreciprocal exchanges and duplications among subtelomeric regions appear to be widespread among eukaroytes. Many ambiguities in the mapping of orthologous yeast genes occur specifically among expanding gene families near the telomere (*16*). These areas show radical changes in gene order, harbor novel sequences, show extensive genomic rearrangement, and are the preferential sites of reciprocal translocations. Subtelomeric regions among *Plasmodium* parasites reveal extensive sequence similarity among nonhomologous chromosomes indicative of frequent exchange (*37, 38*). In the case of *Plasmodium*, these exchanges include large multigene families (*var*, *rif*, and *stevor*) that help the organism to escape host immune response and to establish chronic malarial infections.

Duplications: Engines of Gene and Genome Evolution?

Duplication events have the potential to significantly alter genome structure and the tempo of chromosomal evolution (*35, 39*). Two types of duplications are distinguished by their mechanism of origin: whole-genome and segmental duplication. Whole-genome duplications are cataclysmic genomic events that require the formation of a tetraploid (4N) where all chromosomal material is effectively duplicated. After extensive chromosomal rearrangement and deletion, the disomic state is gradually reestablished with large blocks of conserved gene order evident between nonhomologous chromosomes (*40*). In contrast, segmental duplication involves the duplication of small portions of chromosomal material either in tandem or transposed to new locations within a genome. Both types of duplications may complicate the analysis of chromosomal evolution by obscuring orthologous relationships and promoting nonallelic homologous recombination.

Gene families produced as a consequence of whole-genome duplication are expected to show specific temporal and structural patterns, characterized by a disproportionate number of large paralogous segments that emerged at a specific time point during evolution (*39*). Although still controversial in their extent and

number (*41*), many analyses are consistent with at least one whole-genome eukaryotic duplication occurring independently in several eukaryotic lineages (*42–45*). To the extent that evidence of paralogy for a sufficient number of genes remains after genome duplication, algorithms have been devised to reconstruct the rearrangements that have affected the genome since the tetraploidization event (*46*). Such large-scale duplications have contributed significantly to the expansion of eukaryotic proteome content with estimates ranging from 15 to 50% for the number of genes that owe their existence to these large, often ancient, duplications (*42–44, 47*).

Segmental duplications are most easily recognized as tandem arrays of gene families. A common feature of many of these genes is their importance in the adaptive evolution of the organism. Although the mathematical problem of reconstructing the history of overlapping tandem segmental duplications has been extensively studied (*48*), the analysis of their conserved synteny has been problematic. These clusters tend to create gaps in conserved synteny between species because of extensive rearrangement, functional diversification, or concerted evolution of gene family members (*49–51*). For example, comparative
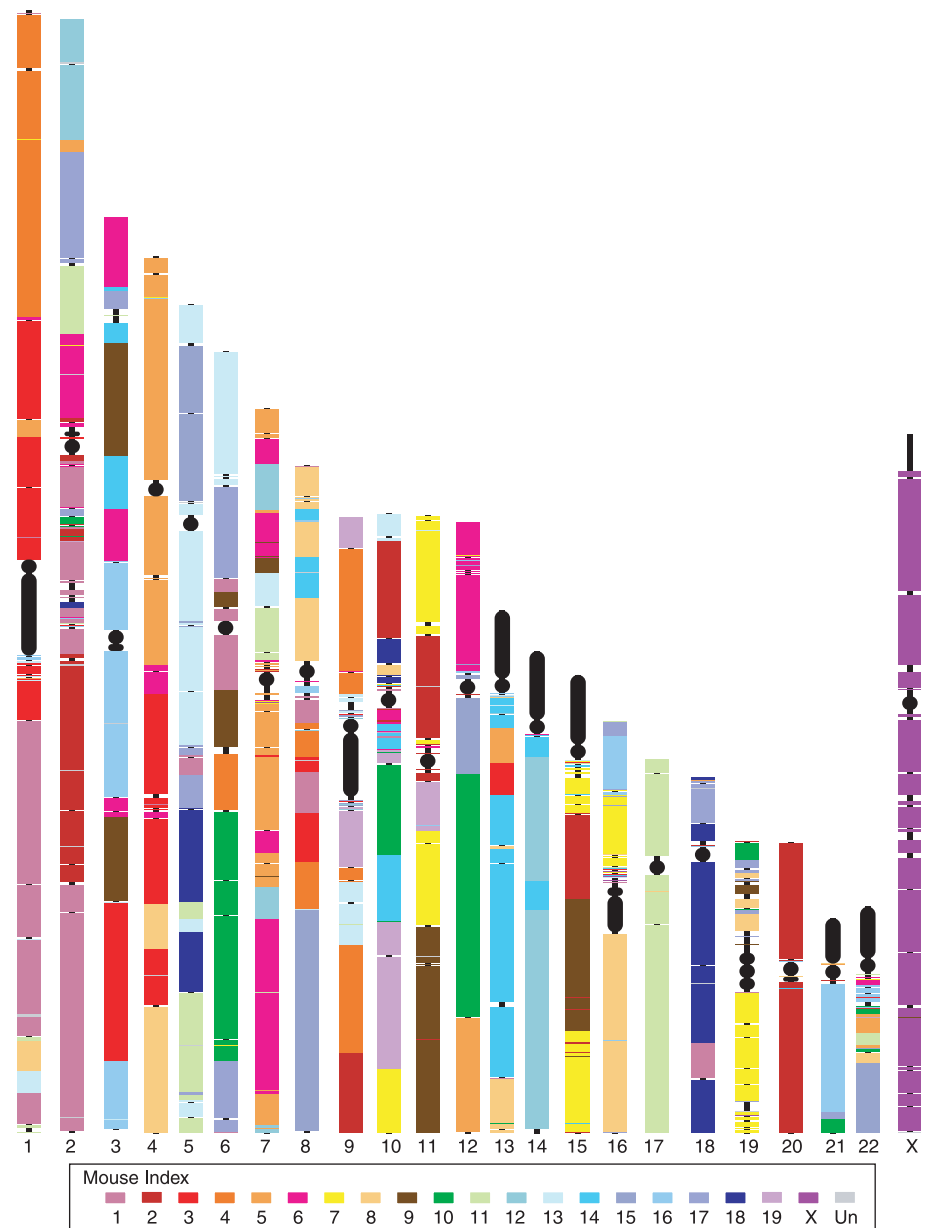

Fig. 1. Conserved synteny between human and mouse. Conserved synteny blocks from the mouse genome (MGSCv. 3.0) are overlaid on human chromosomes (April 2003, assembly). All conserved sytenic blocks >10 kb are shown. Heterochromatic regions (acrocentric arms and centromeres) are colored purple and are not represented by actual sequence. Data were adapted from (*61*), mouse synteny track.

mapping of many tandem gene-family clusters, such as ribosomal DNA, storage protein gene clusters, and disease-resistance genes among cereal genomes, show lack of colinearity among closely related species (*49*). Similarly, analyses of recent tandem expansion of gene families associated with mammalian olfaction or insecticide resistance in *Anopheles gambiae* (cytochrome P-450, glutathione transferases, and carboxylesterase) reveal that "secure orthologs" can be identified for only a small fraction of such genes (*5, 50, 51*).

The proportion of recent segmental duplication varies extensively among sequenced eukaryotic organisms. Between *Caenorhabditis elegans* and *C. brigssae*, for example, only 14 such events have been identified (10 interspersed and 4 tandem) since their divergence (100 Ma) (*10*). In other species, estimates of recent segmental duplication are considerably higher (*33*). Most of the evolutionarily recent events appear as tandem duplication events (*6, 41*). To date, a unique aspect of human genome architecture, however, is the disproportionately large fraction of recent (>90% sequence identity) segmental duplications that are interspersed. These are nonrandomly distributed, vary radically in content among closely related primate species, and are associated with recurrent chromosomal structural rearrangements and disease. It has been estimated that ~200 such regions exist within the genome wherein 5% of the human genome sequence is duplicated (*33*). Once initially formed, segmental duplications promote further rearrangement through their own misalignment and subsequent nonallelic homologous recombination (*52*). This has led to the formation of rapidly evolving pockets (100 kb to 1 Mb in size) of complex genomic architecture. The juxtaposition of genomic regions through segmental duplication has created the potential for the formation of novel genes through positive selection and domain accretion (*35*).

From these observations, it is not unreasonable to assume that rates of segmental duplications have been extremely variable during the course of evolution. Without making a direct estimate of the rate of segmental duplication, Lynch and Conery initially computed an overall rate of 0.01 duplications per gene per million years for the vertebrate lineage (*53*). Gu dated 1739 duplication events based on examination of vertebrate gene families and identified three peaks in the activity of duplication: a wave after the mammalian radiation (potentially primate specific); a major wave at 450 to 650 Ma, consistent with a whole-

genome duplication event at an early stage of vertebrate evolution; and a much earlier wave that occurred during metazoan evolution (*45*). Moreover, these results suggest that rates of gene duplication have been highly variable, ranging from about three to five events per genome per million years before mammalian radiation and increasing to about 10 or more gene duplication events per genome per million years (*33, 45, 53*). Concomitantly, it is unlikely that the impact of segmental duplication on proteome and chromosomal evolution has remained constant.

Transposable Elements Transform Chromosomal Landscapes

Eukaryotic genomes contain substantially differing amounts of repetitive DNA (Table 1) because of differential propagation and deletion of selfish genetic elements. These elements are distinguished by their mode of propagation. LINE (long-interspersed nucleotide elements),
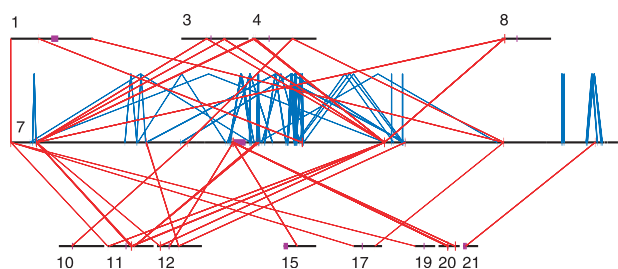

Fig. 2. Recent segmental duplications on human chromosome 7. The distribution of both interchromosomal (red) and intrachromosomal (blue) duplications is shown for human chromosome 7 (drawn ×50 to scale) (*62*). Duplications (>90% sequence identity and >40 kb in length) correspond to duplications/gene conversion events that occurred over the last ~30 million years of human genome evolution.

SINE (short-interspersed repeat elements), and retrovirus-like elements with long terminal repeats (LTR) propagate by reverse-transcription of an RNA intermediate. In contrast, DNA transposons move by a direct "cut-and-paste" mechanism of DNA sequence. These types of events are thought to lead to subtle restructuring of chromosomal landscapes through their integration and subsequent deletion.

Genome analyses have shown that closely related lineages may experience radically different rates of retrotransposition activity. Comparisons of large-scale primate data, for example, indicate that retrotransposition among great-ape species has slowed to a crawl when compared with that of Old World monkeys (*54*). Primate Alu SINE activity reached its zenith 30 to 40 Ma (*5, 55*). Differential rates of SINE/LINE retrotransposition and/or deletion are claimed to be responsible for the 14 to 15% increase in genome size observed among anthropoid primates when compared with

mouse and prosimian primates (*5, 6, 56*). Among many cereal genomes (sorghum, barley, wheat, maize), rapid retrotransposon amplifications have played a more dramatic role in genome size expansion. Sample sequencing of maize clones, for example, indicates that this genome has doubled in size over the last 10 million years almost exclusively because of multiple rounds of retrotransposon bombardment (*57, 58*). For many cereal genomes, this has created a nested-layered effect of retrotranposons inserting into previously integrated copies (*49*). So extensive is retrotransposon amplification that vigorous counterbalancing deletion mechanisms, including both illegitimate recombination and unequal homologous recombination, have been postulated to prevent "genome obesity" within these species (*59*).

Large-scale sequencing of eukaryotic genomes has confirmed unequivocally that such repeats are nonrandomly distributed. Within mouse and human, it is well known that L1 repeats preferentially associate with gene-poor AT-rich regions, whereas Alu SINE repeats accumulate within GC-rich gene-rich areas (*5, 6, 56*). Although many of these events are lineage-specific, similar accumulation biases have been noted in different species such as human and mouse (*6*). Similarly, among different cereal genomes, LTR retrotransposons concentrate within intergenic regions, whereas MITE (miniature inverted transposable elements) propagate within low-copy genic sequence (*29, 57*). These distribution biases become more compartmentalized among highly streamlined, repeat-poor genomes. Among other eukaryotes where repeats constitute <10% of the total content, such as *Arabidopsis* (*2*), *Dictyostelium* (*31*), and *Drosophila* (*28*), repeats are encountered infrequently within euchromatic regions but instead accumulate within heterochromatic areas. Differences in selective constraint and recombination are thought to underlie these biases (*60*).

Chromosomal Rearrangements and Repeats: Cause or Consequence?

One of the common themes to emerge from comparative sequence is that large-scale rearrangements are commonly found near or at regions enriched for repetitive DNA (duplications and transposons). Numerous examples in almost every sequenced eukaryotic group have now been documented. Coghlan and Wolfe considered the distribution of 33 dispersed repeat classes within *C. elegans* and found a significant association with translocation and transposition events but not

chromosomal inversions (*10*). A recent large-scale sequence analysis of four closely related species of yeast (*Saccharomyces cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. bayanus*) indicates that all inversion breakpoints were flanked by transfer RNA genes oriented in the opposite transcriptional orientation and usually of the same isoacceptor type (*16*). Among reciprocal translocation events, 9 out of 10 occurred between highly similar pairs of Ty elements or highly similar ribosomal RNA genes. Almost all disruptions in conserved synteny between *Plasmodium yoelli* and *P. falciparum* map to repetitive RNA genes (*37*).

In an analysis of breakpoint regions between human chromosome 19 and the mouse genome, 10 out of 15 were found to lie in the midst of clustered gene families or an unusually high concentration of L1 and LTR repeat DNA (*51*). Detailed mapping of five breakpoint regions associated with large-scale rearrangements in primate karyotype evolution showed that four out of five of these associated with segmental duplications. Of particular interest, many of these primate segmental duplications also function as breakpoints of recurrent chromosomal structural rearrangements associated with disease and polymorphism within the human population (*35*, *52*). Although these examples suggest that nonhomologous recombination plays a role in chromosomal rearrangements, the temporal order of these events and, therefore, the cause-consequence relations have not been unambiguously determined. It is apparent, however, that the nature and pattern of repetitive DNA is key to understanding the mechanism and dynamics of chromosomal rearrangement among eukaryotic genomes.

Conclusions and Future Directions

Complete-genome sequence of the first biomedical and commercially relevant eukaroytes has ushered in a new era of large-scale evolutionary genomics. The wide phylogenetic gulf, however, separating many of these index genomes means that reconstruction of the events marking their evolutionary divergence can only be summary approximations. The inferred rate of reuse of breakpoints in the human-mouse comparison (*18*, *19*) suggests that few of the actual historical inversions or translocations can be reconstructed with confidence. Only by sequencing related genomes can we hope to reconstruct the general shape of the ancestral genome. Many new genome project initiatives scheduled to be completed within the next 4 years

aim at expanding phylogenetic diversity near these index organisms. As such, we anticipate that this phylogenomic approach will foster advances in three major areas: (i) an understanding of the underlying molecular basis of evolutionary chromosomal rearrangement; (ii) its association with disease, as well as structural polymorphisms and adaptation within species; and (iii) development of computational algorithms to effectively model such changes.

It is important to realize that some of the most structurally dynamic regions of the genome remain, technically, the most challenging to characterize. Despite their central role in chromosomal evolution, complete-sequence characterization of centromeres, telomeres, and highly duplicated regions is still elusive for most organisms. In the short term, few species will be sequenced with the rigor and requisite redundancy of the initial model organisms. Whole-genome shotgun sequencing, although it is expedient and cost-effective, threatens to oversimplify our view of chromosomal evolution by excluding regions by virtue of their structural complexity. The greatest promise of genome sequence is that it is comprehensive. Advances in genome sequencing technology that allow such complex regions to be effectively tackled are currently the most significant technical hurdle to a complete understanding of the dynamics of eukaryotic chromosomal evolution. Relating these evolutionary changes to the functional biology of the chromosome remains an even grander challenge.

References and Notes

1. S. J. O'Brien *et al.*, *Science* **286**, 458 (1999).
2. *Arabidopsis* Genome Initiative, *Nature* **408**, 796 (2000).
3. S. Aparicio *et al.*, *Science* **297**, 1301 (2002).
4. J. H. Nadeau, B. A. Taylor, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 814 (1984).
5. International Human Genome Sequencing Consortium, *Nature* **409**, 860 (2001).
6. Mouse Genome Sequencing Consortium, *Nature* **420**, 520 (2002).
7. J. Ehrlich, D. Sankoff, J. H. Nadeau, *Genetics* **147**, 289 (1997).
8. D. W. Burt *et al.*, *Nature* **402**, 411 (1999).
9. I. V. Sharakhov *et al.*, *Science* **298**, 182 (2002).
10. A. Coghlan, K. H. Wolfe, *Genome Res.* **12**, 857 (2002).
11. F. Grutzner *et al.*, *Genome Res.* **12**, 1316 (2002).
12. H. Skaletsky *et al.*, *Nature* **423**, 825 (2003).
13. N. Archidiacono *et al.*, *Chromosoma* **107**, 241 (1998).
14. M. A. Huynen, B. Snel, P. Bork, *Trends Genet.* **17**, 304 (2001).
15. C. Seoighe *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14433 (2000).
16. M. Kellis, N. Patterson, M. Endrizzi, B. Birren, E. S. Lander, *Nature* **423**, 241 (2003).
17. E. M. Zdobnov *et al.*, *Science* **298**, 149 (2002).
18. P. Pevzner, G. Tesler, RECOMB 2003: The Seventh Annual International Conference on Research in Computational Molecular Biology, organized by the Max Planck Institute for Molecular Genetics and the Berlin Center for Genome-Based Bioinformatics, 10-13 April 2003, Berlin, Germany.
19. P. Pevzner, G. Tesler, *Genome Res.* **13**, 37 (2003).
20. S. Kumar, S. R. Gadagkar, A. Filipski, X. Gu, *Genetics* **157**, 1387 (2001).
21. J. E. Horvath, J. A. Bailey, D. P. Locke, E. E. Eichler, *Hum. Mol. Genet.* **10**, 2215 (2001).
22. R. A. Hoskins *et al.*, *Genome Biol.* **3**, 85 (2002).
23. M. G. Schueler, A. W. Higgins, M. K. Rudd, K. Gustashaw, H. F. Willard, *Science* **294**, 109 (2001).
24. G. P. Copenhaver *et al.*, *Science* **286**, 2468 (1999).
25. H. C. Riethman *et al.*, *Nature* **409**, 948 (2001).
26. M. Ventura, N. Archidiacono, M. Rocchi, *Genome Res.* **11**, 595 (2001).
27. D. J. Amor, K. H. Choo, *Am. J. Hum. Genet.* **71**, 695 (2002).
28. J. S. Kaminker *et al.*, *Genome Biol.* **3**, 84 (2002).
29. N. Jiang *et al.*, *Nature* **421**, 163 (2003).
30. R. A. Holt *et al.*, *Science* **298**, 129 (2002).
31. G. Glockner *et al.*, *Nature* **418**, 79 (2002).
32. J. Guy *et al.*, *Genome Res.* **13**, 159 (2003).
33. J. A. Bailey *et al.*, *Science* **297**, 1003 (2002).
34. H. C. Mefford, B. J. Trask, *Nature Rev. Genet.* **3**, 95 (2002).
35. E. E. Eichler, *Trends Genet.* **17**, 661 (2001).
36. B. Trask *et al.*, *Hum. Mol. Genet.* **7**, 13 (1998).
37. J. M. Carlton *et al.*, *Nature* **419**, 512 (2002).
38. M. J. Gardner *et al.*, *Nature* **419**, 498 (2002).
39. D. Durand, *Trends Genet.* **19**, 2 (2003).
40. S. P. Otto, J. Whitton, *Annu. Rev. Genet.* **34**, 401 (2000).
41. R. Friedman, A. L. Hughes, *Mol. Biol. Evol.* **20**, 154 (2003).
42. J. E. Bowers, B. A. Chapman, J. Rong, A. H. Paterson, *Nature* **422**, 433 (2003).
43. K. Wolfe, D. Shields, *Nature* **387**, 708 (1997).
44. A. McLysaght, K. Hokamp, K. H. Wolfe, *Nature Genet.* **31**, 200 (2002).
45. X. Gu, Y. Wang, J. Gu, *Nature Genet.* **31**, 205 (2002).
46. N. El-Mabrouk, D. Sankoff, *SIAM (Soc. Industr. Appl. Math.) J. Comput.* **32**, 745 (2003).
47. D. Sankoff, *Bioinformatics* **15**, 909 (1999).
48. L. Zhang, B. Ma, L. Wang, Y. Xu, *Bioinformatics* **19**, in press.
49. W. Ramakrishna *et al.*, *Genetics* **162**, 1389 (2002).
50. H. Ranson *et al.*, *Science* **298**, 179 (2002).
51. P. Dehal *et al.*, *Science* **293**, 104 (2001).
52. P. Stankiewicz, J. R. Lupski, *Trends Genet.* **18**, 74 (2002).
53. M. Lynch, J. S. Conery, *Science* **290**, 1151 (2000).
54. G. Liu *et al.*, *Genome Res.* **13**, 358 (2003).
55. M. A. Batzer, P. L. Deininger, *Nature Rev. Genet.* **3**, 370 (2002).
56. R. J. Mural *et al.*, *Science* **296**, 1661 (2002).
57. P. SanMiguel, B. S. Gaut, A. Tikhonov, Y. Nakajima, J. L. Bennetzen, *Nature Genet.* **20**, 43 (1998).
58. B. S. Gaut, M. Le Thierry d'Ennequin, A. S. Peek, M. C. Sawkins, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7008 (2000).
59. K. M. Devos, J. K. Brown, J. L. Bennetzen, *Genome Res.* **12**, 1075 (2002).
60. B. Charlesworth, C. H. Langley, W. Stephan, *Genetics* **112**, 947 (1986).
61. See www.genome.ucsc.edu.
62. L. W. Hillier *et al.*, *Nature* **424**, 157 (2003).
63. See www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome.
64. We thank our colleagues for access to unpublished data in the preparation of this review and R. Clark for technical assistance in the preparation of display items. E.E.E. is supported by NIH grants (HG02385, GM58815, and HD43569). D.S. holds the Canada Research Chair in Mathematical Genomics and is a Fellow of the Canadian Institute for Advanced Research. We apologize to all colleagues whose work we could not cite because of space constraints.