VIEWPOINT

Prokaryotic Chromosomes and Disease

Jörg Hacker,* Ute Hentschel, Ulrich Dobrindt

Recent insights into bacterial genome organization and function have improved our understanding of the nature of pathogenic bacteria and their ability to cause disease. It is becoming increasingly clear that the bacterial chromosome constantly undergoes structural changes due to gene acquisition and loss, recombination, and mutational events that have an impact on the pathogenic potential of the bacterium. Even though the bacterial genome includes additional genetic elements, the chromosome represents the most important entity in this context. Here, we will show that various processes of genomic instability have an influence on the many manifestations of infectious disease.

Roughly 5000 bacterial species have been described, representing a mere 0.5 to 1% of the total number of prokaryotes. Only an extremely small portion of these microbes, about 200 species, are known to cause disease in humans. Yet, for some of the most feared diseases, the infection dose required may be exceedingly small: it takes on average only 10 microbes of *Yersinia pestis* to cause bubonic plague and only 100 microbes of certain *Shigella* species to initiate severe dysentery. Considering the impact that pathogenic microorganisms had on human history and considering that infectious disease is still the principal threat to human health today, it is important to ask how pathogenic bacteria cause disease.

Virtually every niche of the human body that can be colonized by bacteria is prone to infection. Fortunately, most bacteria residing in or on the human body are harmless commensals, for example, those that occur in the intestine. Current theory holds that the majority of disease-causing bacteria from the intestine may have been derived from commensals that have acquired genes from foreign sources turning them into pathogens. Another important mechanism by which harmless bacteria may turn into pathogens is change of host or host niche, upon which their virulence potential is frequently revealed to its full extent. Certain bacterial diseases caused by *Y. pestis*, *Salmonella enterica, Borrelia burgdorferi*, or multiresistant enterococci are dramatic examples underscoring the relevance of the host side of infection.

With the advent of DNA sequencing, it has become possible to correlate infectious disease with prokaryotic genome structure. The sequence data of more than 50 fully annotated genomes of pathogenic and nonpathogenic bacteria have allowed the identification of unifying patterns as well as differences among genomes of pathogenic and closely related, nonpathogenic bacteria. It has also revealed mechanisms that promote genome plasticity, such as horizontal gene transfer, genome reduction, genome rearrangements, and the generation of point mutations. Moreover, the discovery of superintegrons has altered our understanding of infectious disease. Here, the relationships between genome evolution and disease that have emerged recently are discussed.

Institut für Molekulare Infektionsbiologie, Universität Würzburg, Röntgenring 11, 97070 Würzburg, Germany.

*To whom correspondence should be addressed. E-mail: j.hacker@mail.uni-wuerzburg.de

Evolution of Pathogens

Horizontal gene transfer (HGT) is the process by which genetic information is passed from one bacterial genome to another (*1*, *2*) (Fig. 1). HGT is especially important in the evolution of pathogenic lifestyles as infection-related factors can be transmitted in a single-step integration event. The three most important characteristics are as follows:

Antibiotic resistance. Resistance determinants of Gram-negative bacteria are often associated with mobile or transferable genetic elements such as plasmids, integrons, super-integrons, and complex transposons. Furthermore, pathogenic variants of Gram-positive cocci (e.g., *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Enterococcus faecalis*) causing severe cases of sepsis and catheter-associated infections in hospitals carry genomic islands encoding methicillin resistance and/or large transposons responsible for, e.g., vancomycin resistance (*3*–*5*). Integrons are natural cloning and expression systems that incorporate open reading frames and convert them into functional genes. This allows the accumulation of large arrays of gene cassettes that can eventually be transferred as a whole between different replicons (*6*). They are also the primary system for antibiotic resistance and virulence gene capture in Gram-negative enterobacteria. Superintegrons represent another type of integron that occurs in many genera of the γ-Proteobacteria and are far superior in their ability to "stockpile" gene cassettes of different functions, including virulence traits.

Pathogenicity. Virulence genes are frequently located on mobile or formerly mobile genetic elements including pathogenicity islands (PAIs) that are present in Gram-negative and Gram-positive bacteria (*7*–*9*). PAIs represent large chromosomal regions of horizontally acquired DNA that are believed to have evolved from former lysogenic bacteriophages and plasmids. The subsequent bacterial acquisition of virulence-associated factors encoded on different mobile genetic elements indicates a functional interdependency between such factors. Accordingly, the virulence factor SseI encoded by the Gifsy 2 phage in *S. enterica* sv. Typhimurium is secreted by a type III secretion system that is itself encoded on the pathogenicity island SPI-2. Moreover, it is becoming increasingly clear that independent transfer events can have synergistic effects. For example, this phage also encodes another virulence factor, GtgE, (*10*) and a superoxide dismutase, SodC, which acts as a fitness factor. The combination of phage- and PAI-encoded factors, both offensive and defensive, supports infections due to *S. enterica* sv. Typhimurium.

The integration of newly acquired genetic elements into general regulatory circuits as well as the coordination of their expression is a prerequisite for optimal function. The genes *mgtC* and *sopD2*, involved in the invasive phenotype of *S. enterica*, are regulated by the two-component regulatory systems *pho*P/Q- and *ssrA*/B, respectively (*11*, *12*). In the first case, the horizontally acquired gene comes under the control of preexisting regulators. In the latter case, the regulator itself was introduced on a pathogenicity island and has come to control the regulation of transcription of phage-encoded genes. The mechanisms by which newly acquired elements are harnessed by preexisting networks and by which compatibility of different genetic systems is ensured are as yet unkown.

Fitness traits. Many horizontally acquired determinants are involved in metabolic adaptation and increasing survival of the bacterium. These traits are found in commensal and pathogenic bacteria alike. For example, the so-called "high pathogenicity island" initially described in the highly virulent *Yersiniae* has subsequently been found in nonpathogenic enterobacteria (*13*, *14*). Comparative analysis of the complete genome sequences of *Escherichia coli* and *S. enterica* variants has revealed that none of the phenotype traits that distinguish the two species are attributable to individual point mutations. Instead, species-specific traits derive from functions encoded

either by horizontally acquired genes (e.g., lactose, citrate, and propanediol utilization, indole production) or from the loss of ancestral genes (e.g., alkaline phosphatase).

There is evidence that HGT between bacteria can occur during infection, for example with *Campylobacter jejunii,* which causes diarrhea (*15*), or even during passage of *Y. pestis* through an insect vector (*16*). Although these and other observations suggest that environmental stress can stimulate HGT, the signals that trigger this event in vivo are, for the most part, unknown. Transfer RNA (tRNA) genes frequently serve as integration sites for mobile elements, but the mechanism for this selective integration is unknown. Because tRNA gene sequences are highly conserved, they may increase the host range of a mobile element. Additionally, tRNA genes are generally transcriptionally active ensuring immediate expression of acquired genes. After integration of mobile genetic elements into tRNA genes by site-specific recombination, the tRNA genes remain functional. They also exhibit symmetric nucleotide sequences in the stem loops facilitating the binding of integrases. The association with particular so-called "minor" tRNAs may also have modulatory effects on the translational efficiency of target genes (*17*).

Genome Reduction in Pathogenic and Symbiotic Bacteria

Because bacterial genomes are not growing ever larger in size, the acquisition of foreign genetic elements must be counterbalanced by the loss of native genes. Deletional bias is a major force shaping bacterial genomes. In some cases, the loss of gene function may provide a selective advantage, as exemplified by the beneficial loss of metabolic genes (termed "black holes") (*18*). Many unexpressed pseudogenes of the pathogen *Y. pestis* are functional in other *Yersinia* species (*19*), implying that gene loss contributes to the adaptation of *Y. pestis* to its insect vector, which is a prerequisite for transmission of this pathogen from rodents via pest fleas to humans.

Analysis of genome sequence information of various obligately intracellular bacteria (pathogenic or symbiotic), such as *Chlamydia* spp., *Rickettsia* spp., *Buchnera aphidicola*, and *Blochmannia* spp., shows that these bacterial genomes have lost large amounts of DNA. This phenomenon also emphasizes the similar mechanisms between pathogens and symbionts (*20*, *21*). Genes that confer metabolic traits necessary for niche adaptation are maintained, whereas those that do not provide a selective benefit are lost. Eventually, the optimization of these processes shapes the genome architecture of a microorganism (Fig. 1). In bacteria that have been associated with hosts for evolutionarily long periods of time, the genome structure frequently reflects the lifestyle of the bacterium (*22*). Accordingly, intracellular symbionts contain genes encoding beneficial functions that may supplement nutrition of their hosts, whereas intracellular parasites eventually cause host damage.

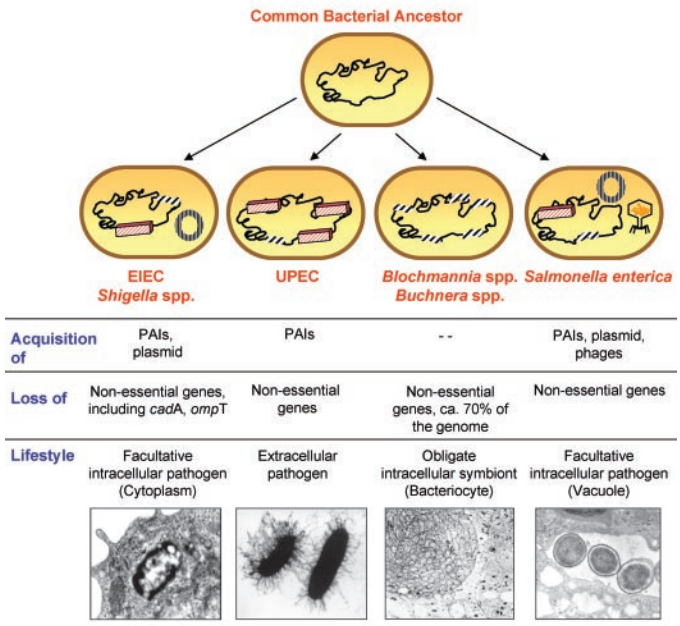

Fig. 1. Evolution of different variants of pathogenic and symbiotic γ-proteobacterial variants by acquisition and loss of genetic information from a common bacterial ancestor (e.g., *S. enterica*, *Shigella* spp., uropathogenic *E. coli* and the endosymbionts of aphids, *B. aphidicola*, and ants, *Blochmannia* spp. Abbreviations are as follows: *cad*A, lysine decarboxylase-encoding gene; *omp*T, outer membrane protein T-encoding gene; PAI, pathogenicity island; EIEC, enteroinvasive *E. coli*; UPEC, uropathogenic *E. coli*.

It is tempting to speculate that the loss of genetic information is programmed in some way to ensure long-term persistence in the host. This hypothesis is supported by the observation that the virulence potential of uropathogenic *E. coli* isolated from acute infections differs markedly from those recovered from chronic infections. This phenotypic modulation under in vivo conditions leads to an irreversible loss of genes, gene blocks, or even entire PAIs during infection (*7*). Apparently, less virulent variants are better adapted for a long-term colonization than their highly pathogenic counterparts. The signals and enzymes involved in the directed loss of genetic information or "phase variation," i.e., the switch between an "on" and "off" status of gene expression, during the course of the infection remain to be resolved.

DNA Rearrangements

Bacterial genomes constantly undergo rearrangements. DNA repeats and gene paralogs can mediate intragenomic recombination events that can simultaneously alter the expression of disease-associated genes. Genome rearrangements often play a role in surface structure variation to circumvent confrontation with the host immune system (Fig. 2). Phase variation has been described for type 1 fimbriae (Fim) expression in pathogenic *E. coli*. Type I fimbriae production is increased during urinary tract infection promoting colonization by uropathogenic *E. coli* strains. Phase variation results from a stimulation of FimB recombinase expression in vivo (*23*); however, the stimuli for the preferential in vivo "on" status of the *fim* switch are not known. Transposition and precise excision of accessory genetic elements [e.g., insertion sequences (IS)] can also cause phase variation, e.g., for biofilm formation of the nosocomial pathogen *S. epidermidis* (*24*).

The genome of *B. burgdorferi*, the causative agent of lyme disease, undergoes dynamic rearrangements within the chromosome and among the 12 linear and 9 circular plasmids. A substantial fraction of the genome is made up of paralogous genes. About 5% of the chromosomal genes and an estimated 15% of plasmid genes as well as many pseudogenes encode for lipoproteins. Lipoproteins are important surface structures and targets for the host immune response. *Borrelia* apparently uses recombination to vary its surface structures with both homologous and nonhomologous mechanisms being involved in switching or recombining of these paralogs (*25*).

The most striking feature of pathogenic *Neisseria* species (*N. gonorrhoeae* and *N. meningitidis*, which cause gonorrhea and meningitis, respectively) is the amount of repetitive DNA in the chromosome. Repeat-mediated rearrangements facilitate cell surface genes moving around on the chromosome, allowing "silent genes" to be positioned next to "on" switches where they become active. Other repeat sequences may facilitate rearrangements of DNA within cell surface genes. Internal shuffling of these genes changes the encoded proteins, and each generation of bacteria presents a different appearance to the immune system. Phase variation by slipped-strand mispair-

ing (SSM) causes changes in the number of repeats thus changing the coding frame of the gene. Examples include capsule-, hemoglobin receptor (HmbR)-, or opaque-protein expression in *Neisseria* (*26*), which seem to be variably induced in different stages of disease. Additionally, antibiotic resistance can be caused by DNA rearrangements, such as remodeling of penicillin-binding protein encoding genes that result in penicillin resistance of *Streptococcus pneumoniae*, a severe pathogen of the respiratory tract (*27*).

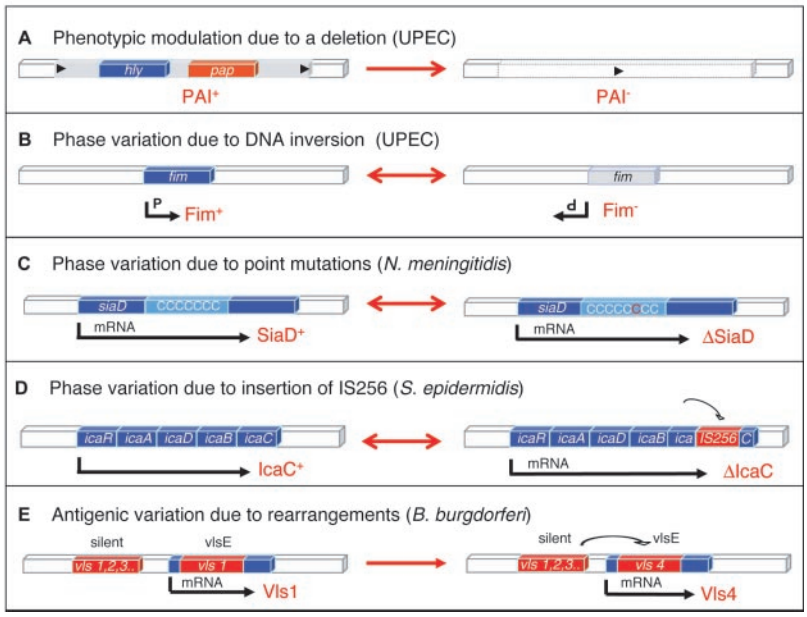

Fig. 2. Mechanisms contributing to chromosomal variability of pathogens. Chromosomal variations can result from (**A**) phenotypic modulation, e.g., the deletion of a pathogenicity island encoding α-hemolysin (*hly*) and P-fimbriae (*pap*) in uropathogenic *E. coli* (UPEC); (**B**) phase variation, e.g., the inversion of DNA elements such as the *fim* promoter switch directing type 1 fimbriae (Fim) expression in UPEC; (**C**) "slipped strand mispairing," e.g., phase variation by point mutation within the *siaD* gene required for capsule expression in *N. meningitidis*; (**D**) phase variation of biofilm formation by insertion or excision of an insertion sequence element (IS*256*) into the extracellular polysaccharide-encoding *ica* gene cluster of *S. epidermidis*; and (**E**) antigenic variation by DNA rearrangements in a variable surface-exposed lipoprotein gene cassette *vlsE* of *B. burgdorferi*. Abbreviations are as follows: C, cytosin; P, promoter; PAI, pathogenicity island.

Adaptation and Variation of Mutation Rates

The majority of bacteria seem to pass repeatedly through periods of increased mutation rates during their evolutionary history. However, a link between high mutation rates and virulence potential cannot be generalized at this point (*28*). In *E. coli* and *S. enterica* sv. Typhimurium, *mutS* mutators that are deficient in DNA mismatch repair accelerate the mutation rates and relax the barriers that normally restrict homologous recombination. Interestingly, the *mutS* gene belongs to a recombinational hot spot within the *E. coli* and *Salmonella* chromosome (*mutS-rpoS* region), suggesting that *mutS* itself may also be subject to horizontal transfer. Rescue of defective *mutS* alleles with wild-type sequences by HGT may be a mechanism for stabilizing adaptive changes promoted by *mutS* mutators and has been reported to occur in nature (*29*, *30*).

More than one-third of cystic fibrosis (CF) patients harboring *Pseudomonas aeruginosa* are infected by mutator strains (*31*), whereas no mutator strains were found among *P. aeruginosa* isolated from lungs of non-CF patients in this study. It is noteworthy that a correlation has been observed between high mutation rates and multiple antibiotic resistance. CF patients are also infected with *S. aureus* mutator strains (*32*). Differences in mutation rates affect SSM and, therefore, phase-variable expression of hemoglobin receptor genes in *N. meningitidis* (*33*). Thus, mutator bacteria may gain an advantage in certain pathologies.

Point mutations resulting in single-nucleotide polymorphisms (SNPs) can lead to genetic alterations that provide a selective advantage during the course of a single infection, epidemic spread, or the long-term evolution of virulence. Allelic variations of fimbrial adhesins in *E. coli* and *S. enterica* sv. Typhimurium can determine host specificity and tissue tropism and can serve as a molecular bridge from commensal to pathogenic lifestyles. For example, naturally occurring point substitutions in FimH alleles, coding for the type 1 fimbrial adhesin of uropathogenic *E. coli*, result in higher affinity for monomannose (and type IV collagen) receptors than most intestinal commensal isolates. This correlates with an increased tropism for uroepithelium and bladder colonization. In *S. enterica* sv. Typhimurium, SNPs in the type 1 fimbrial adhesin gene produce important differences in HEp-2 cell binding, biofilm formation, and host-colonization (*34*, *35*). These findings underscore the great impact of mutations as generators of diversity.

Future Challenges

New insights regarding the mechanisms of infectious disease have been gained in the wake of large-scale genome sequencing. Most importantly, comparative and functional genomics have helped unravel the magnitude of horizontal gene transfer and its impact on prokaryotic genome evolution. The continued understanding of these processes provides us with a vision of how genome dynamics may contribute to infectious disease. It has also become apparent that evolutionary events are accelerated during infection. In figurative terms, disease can be regarded as an "evolutionary pressure cooker" rather than Darwin's "warm little pond." The research accomplishments of the past few years have provided, for the first time, insights into the evolutionary origins of infectious disease. Future questions that must be addressed are: What is the in vivo relevance of horizontal gene transfer during the course of an infection? Is HGT a programmed event, how is it regulated, and what might the signals be? By what mechanisms does the genome maintain stability and at the same time flexibility in the face of environmental challenge and how does it protect function? What is the function of the large number of unknown genes that are located on horizontally acquired elements? In summary, we are able, for the first time, to illuminate the dynamical processes of genome evolution and to correlate these findings with infectious disease.

References and Notes

1. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature* **405**, 299 (2000).
2. U. Dobrindt *et al.*, *J. Bacteriol.* **185**, 1831 (2003).
3. I. T. Paulsen *et al.*, *Science* **299**, 2071 (2003).
4. M. Kuroda *et al.*, *Lancet* **357**, 1225 (2001).
5. A. Umeda, F. Garnier, P. Courvalin, M. Galimand, *J. Antimicrob. Chemother.* **50**, 253 (2002).
6. D. A. Rowe-Magnus, A. M. Guerout, L. Biskri, P. Bouige, D. Mazel, *Genome Res.* **13**, 428 (2003).
7. J. Hacker, J. B. Kaper, *Annu. Rev. Microbiol.* **54**, 641 (2000).
8. E. F. Boyd, H. Brussow, *Trends Microbiol.* **10**, 521 (2002).
9. N. Shankar, A. S. Baghdayan, M. S. Gilmore, *Nature* **417**, 746 (2002).
10. T. D. Ho *et al.*, *J. Bacteriol.* **184**, 5234 (2002).
11. E. A. Groisman, *J. Bacteriol.* **183**, 1835 (2001).
12. J. H. Brumell *et al.*, *Traffic* **4**, 36 (2003).
13. J. Hacker, E. Carniel, *EMBO Rep.* **2**, 376 (2001).
14. T. A. Ölschläger *et al.*, *J. Bacteriol.* **185**, 1107 (2003).
15. P. Boer *et al.*, *Mol. Microbiol.* **44**, 351 (2002).
16. B. J. Hinnebusch, M. L. Rosso, T. G. Schwan, E. Carniel, *Mol. Microbiol.* **46**, 349 (2002).
17. U. Dobrindt, L. Emödy, I. Gentschev, W. Goebel, J. Hacker, *Mol. Genet. Genomics* **267**, 370 (2002).
18. A. T. Maurelli, R. E. Fernandez, C. A. Bloch, C. K. Rode, A. Fasano, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3943 (1998).
19. J. Parkhill *et al.*, *Nature* **413**, 523 (2001).
20. U. Hentschel, M. Steinert, J. Hacker, *Trends Microbiol.* **8**, 226 (2000).
21. R. Gross, J. Hacker, W. Goebel, *Mol. Microbiol.* **47**, 1749 (2003).
22. N. A. Moran, *Cell* **108**, 583 (2002).
23. N. W. T. Gunther, V. Lockatell, D. E. Johnson, H. L. Mobley, *Infect. Immun.* **69**, 2838 (2001).
24. I. Lössner, K. Dietrich, D. Dittrich, J. Hacker, W. Ziebuhr, *J. Bacteriol.* **184**, 4709 (2002).
25. S. Casjens *et al.*, *Mol. Microbiol.* **35**, 490 (2000).
26. N. J. Saunders *et al.*, *Mol. Microbiol.* **37**, 207 (2000).
27. R. Hakenbeck *et al.*, *Infect. Immun.* **69**, 2477 (2001).
28. B. Picard *et al.*, *Infect. Immun.* **69**, 9 (2001).

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