Rearrangements and chromosomal evolution

David Sankoff

Comparisons of the genome sequences of related species suggests varying patterns of chromosomal rearrangements in different evolutionary lineages. In this review, I focus on the quantitative characterization of rearrangement processes and discuss specific inventories that have been compiled to date. Of particular interest are the statistical distribution of the lengths of inverted or locally transposed chromosome fragments (notably very short ones), inhomogeneities in susceptibility to evolutionary breakpoints in chromosomal regions, the relative importance of genome doubling in the history of multicellular eukaryotes, and of lateral transfer versus gene gain and loss in prokaryotes. These developments provide challenges to computational biologists to refine, revise and scale up mathematical models and algorithms for analyzing genome rearrangements.

Addresses
Department of Mathematics and Statistics, University of Ottawa, 585 King Edward Avenue, Ottawa, K1N 6N5, Canada
e-mail: sankoff@uottawa.ca

Current Opinion in Genetics & Development 2003, **13**:583–587

This review comes from a themed issue on
Genomes and evolution
Edited by Evan Eichler and Nipam Patel

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2003.10.006

Introduction

Analytical paradigms for studying genomic evolution have been in the making for several years [1,2•]. With the availability of essentially complete genome sequences, these paradigms can now be tested, revised and scaled up. Pairs of closely related genomes sharing large conserved chromosomal segments provide clear evidence of the nature of individual rearrangement events, although in relatively small numbers. Larger sets of genomes within a family, class or order can provide data about the rate of rearrangement and how it varies, but because remaining conserved segments are short and subject to local 'microrearrangements', there will be ambiguity about details of individual events.

In this review, I bypass sequence-level processes that affect *intensive* genome variables such as base composition and codon usage, which are invariant under rearrangement operations, in favour of evolutionary events that can alter *extensive* variables such as gene order, genome size or ploidy.

Purely combinatorial algorithms for comparing genomes are essentially model-free, requiring no *a priori* quantitative characterization of the processes of genome rearrangement. It is possible to introduce various parameters into the optimizing criterion to increase realism, but this leads to algorithmically much harder problems. Nevertheless, given the large eukaryotic genome sequences that are available, answers to the following questions could aid in reconstructing the details of genomic evolution. With what frequency does each type of rearrangement operation occur in a given evolutionary lineage? What regions or sites on the chromosome are susceptible to a particular type of change? How large is the DNA segment inverted, transposed, deleted, translocated, or duplicated? Partial answers are now known for several lineages and I discuss how to integrate this information into inference about evolutionary genomics.

Rearrangement repertories, regions and rates in eukaryotes

Comparisons of parts of the *Drosophila repleta* genome with *D. melanogaster* [3] suggest an overall rate of ~0.05 paracentric inversions (corresponding to 0.1 chromosomal breakages) per megabase per million years, with virtually no interchromosomal translocation. A similar regime has been reported for the genus *Anopheles* (mosquito) [4], although the more distant comparison between the fruit fly and the mosquito [5] does suggest considerable translocation. In comparing the nematode *Caenorhabditis elegans* with *C. briggsae*, Coghlan and Wolfe [6] have estimated the ratio of translocations to inversions to transpositions to be 1:1:2, and a rearrangement rate at least four times that of *Drosophila*.

The subtelomeric regions of the chromosomes of the human malaria parasite *Plasmodium falciparum* [7] display extensive inter-chromosomal exchange between subtelomeric regions also observed in several parasitic species, including *Encephalitozoon cuniculi* [8]. This likely facilitates selective antigenic variability. A pattern of telomeric translocation has also been clearly documented through the comparison of the genome sequences of *Saccharomyces cerevisiae* and three closely related yeasts: *S. paradoxus*, *S. mikatae* and *S. bayanus* [9].

Over a much longer history, yeast genomes have undergone thousands of inversions, many of them involving very few, or even single genes [10]. This was previously shown to be a characteristic of many smaller eukaryotic genomes [11], but inversions of a few hundred or a few thousand base pairs of DNA are now also known to be very common even in higher eukaryotes [12••].

Transpositions of duplicated genomic segments along human chromosomes, and especially into pericentromeric and subtelomeric regions of multiple chromosomes, has revealed itself to be a major component of recent human [13•,14] and primate evolution [15] as well as in mice [16]. Transpositions of small clusters of genes between chromosomes are suggested to play an important role in the evolution of plants [17].

Whether or not modern genomes show evidence of ancient tetraploidization events, and how many, remain acrimonious issues, with debates raging around the human, *Arabidopsis* and yeast genomes. Wong, Butler and Wolfe [18] have shown that the hemiascomycete yeasts can be neatly divided into two groups: the *Saccharomyces cerevisiae* lineage whose members show consistent evidence of an ancestral polyploidization event, and those that diverged from the *Saccharomyces* line before this event, showing no such evidence. Recent analyses confirm tetraploidization occurring in the *Arabidopsis* lineage since its split from the rest of the dicots [19]. Finally, the continuing dissent of Hughes [20] has been refuted [21,22•], the weight of evidence showing that the early vertebrate lineage saw at least one round of genome duplication [23,24].

Lateral transfer of genes from one organism to another is a mechanism for introducing new genes or paralogs of existing genes into a genome. This interferes with phylogenetic inference and, particularly when it involves transfer of 'islands', multigene DNA fragments [25], the identification of vertically conserved segments and clusters, and hence of rearrangement. Early enthusiasm (or concern) over large numbers of inter-kingdom transfers inferred to exist in the human genome [26] was largely dispelled by more conservative analyses [27], on the basis of the genome-internal duplication and loss process. Ongoing inter-kingdom transfer into eukaryotes (e.g. from prokaryotic symbionts) would now seem pertinent mostly to unicellular organisms or to transfer from organellar to nuclear genomes [28]. Furthermore, in the inventory of eubacterial and archaeal genome sequences, intergenome transfer is quantitatively dwarfed by processes of gene gain — by duplication and the emergence of new genes — and especially by gene loss [29]. The latter process is accelerated in symbionts and parasites [8].

Another eukaryotic process, which has not yet been addressed in algorithms for the inference of translocation, and chromosome fission and fusion, is new centromere formation [30].

Genome-level inventories of translocations, inversions, transpositions, duplications, and losses resulting from comparisons of genomic sequences represent a major new advance over the traditions of comparative mapping. But to bring these data to bear on our understanding of the mechanisms of chromosomal evolution, they must be interpreted in the light of the formal theories of genome rearrangement that have been developed in the frameworks of combinatorial algorithmics and probabilistic models.

Loss of signal in prokaryote comparison

With notable exceptions [31], comparative genomics studies of prokaryotes have not focused on genome-wide rates of rearrangement or sizes of conserved segments in the way eukaryotic comparisons have for the simple reason that gene adjacency is poorly conserved even among relatively closely related organisms. Gene content changes radically through both loss and horizontal acquisition, especially among pathogens, and even conserved groups of functionally related genes maintain their proximities in a loose way, if at all. The eukaryote-based notion of a conserved segment with identical gene content and gene order must be replaced with a concept of gene cluster [32], run [31], team [33], and so on, where order, contiguity and even strandedness may be relaxed to some extent. A series of tests has been proposed for investigating whether such loose groupings are statistically significant compared to chance co-occurrences in the genomes being examined [32].

The paradox of conservation of loose proximity despite rapid loss of strict adjacency has traditionally been resolved by appealing to selective functional pressures or by invoking operon-like concepts, but may simply be caused by a prevalence of the same sort of short-range rearrangements described above for eukaryotes. Probabilistic modeling [34] shows that a regime of short inversions scrambles gene clusters without dispersing them widely, whereas a regime of long inversions results in many conserved adjacencies within segments widely distributed within the genome, for comparable average amounts of total evolution.

Simulation experiments [35] show that as inversions accumulate on the prokaryotic genome, the details of the evolutionary signal are rapidly lost, so that although some overall degree of rearrangement may be estimated, the particular events and their sequence are obscured. The clear exceptions are found in very short inversions, evidence for which tends to persist for long periods of time despite being overlaid by many other rearrangements. Moreover, controlled analyses of many pairs of bacterial genomes detect and validate significant numbers of short inversions.

Analytical advances

In the field of combinatorial optimization, much attention has been devoted to improving the efficiency of the Hannenhalli–Pevzner algorithm for inferring the reversal (inversion) or translocation history between two genomes [36–38] and to generalizing this to take into account gene

insertion and loss [39,40], and paralogy [41]. Much effort has also gone into extending two-genome comparisons towards multiple genome analyses in the phylogenetic context [42], as reviewed in [43•].

Several publications have recently appeared on probabilistic models of rearrangement with a view to making them amenable to statistical inference [44–47]. In another algorithmic development, the problem of reconstructing the gene order of the chromosomes of a genome just as it underwent tetraploidization, based on the rearranged (translocated and inverted) genome of a modern-day descendant, has been solved in complete generality [48].

The Nadeau–Taylor approach to comparative mapping as reformulated in terms of gene order has seen several theoretical improvements, related to observability of conserved segments, chromosomal inhomogeneities in susceptibility to breakpoints, chromosome-dependent rearrangement rates, marker density and improved estimation [12••,49,50].

Inhomogeneities

A survey of rearrangement breakpoints culled from tumour karyotypes and infertility tests showed that whereas somatic cell chromosomal aberrations tended to cluster arm-medially over the set of autosomal arms, with certain arms also showing strong telomeric concentrations, breakpoints in the genomes of normal translocation carriers showed a uniform random distribution, at least at the level of resolution determined by chromosome bands [51]. Algorithmic approaches to link tumour genome rearrangements with evolutionary breakpoint theory are underway [52•].

At higher levels of resolution, perusal of comparisons between the human and mouse genome sequences with the help of tools such as those developed by Kent *et al.* [53••] reveals, in each genome, large numbers of relatively short regions containing elements corresponding to parts of several different chromosomes in the other species. It is unclear to what extent these sites reflect repeated reciprocal translocation, transposition of small regions, duplication processes, conversion, or errors in assembly or homology algorithms.

Formal whole-genome comparisons have relied largely on the combinatorial algorithms of Hannenhalli and Pevzner. Until recently, these could only be applied to small genomes (e.g. mitochondria, chloroplasts and eubacteria), the difficulty with eukaryotic nuclear genomes being not so much the computational cost but rather the absence of comprehensive lists of genes and their orthologs. Pevzner and Tesler recently showed how to bypass gene finding and ortholog identification by using a large number (5×10^5) of small 'anchor' sequences (10^3–10^5 bp), homologous in two genomes, which are then combined to construct large syntenic blocks (>1 Mb) [12••]. The blocks are then input as unresolved items into a variant of the original Hannenhalli–Pevzner algorithm.

The use of large unresolved blocks, and the (largely unavoidable) neglect of blocks $<10^6$ elements, however, introduces a great deal of noise into the analysis. Deleting blocks, however small, and combining others, interferes with the historical signals about rearrangement inherent in the order of the blocks. This leads to reconstructions containing erroneous inversions and translocations, and to artifactual inferences about inhomogeneities in genomic susceptibility to rearrangements.

It is nonetheless obvious from browsing structured comparative databases [53••] that there are highly jumbled small regions between some of the syntenic blocks and elsewhere in the human–mouse comparison. There is also evidence from multiple sources that some larger regions have heightened susceptibility to rearrangement across the evolutionary spectrum, whether pericentromerically [13•,16] or subtelomerically [7–9,14], where there is high repeat density [6], high recombination rate [53••] or between pairs of tRNA genes [9]. There is also substantial evidence that other regions are refractory to rearrangement in that particular gene neighborhood organization may have functionally selective advantages [53••,54]. The quantitative study of these inhomogeneities in rearrangement susceptibilities is just beginning, and refined versions of the Pevzner–Tesler approach promise to play a key role.

Conclusions

The wealth of eukaryotic genome sequence and related comparative data now being produced are enabling the newly quantitative study of genome rearrangement data at the gene-order level and at finer resolutions. One of the more striking results is the variety of patterns of rearrangement, including what types of chromosomal change, where they occur, and how frequently, in different evolutionary lineages. Much attention has been paid to the prevalence and consequences of segmental duplication. At the sequence level, an unexpected pattern of small changes has emerged, as a result of inversion, transposition, duplication, deletion or insertion of a few thousand or a few hundred base pairs, though the effect of this, if any, on gene order in the higher eukaryotes remains to be ascertained. An underdeveloped area remains the functional consequences of chromosomal rearrangement at the evolutionary time scale, with only a few pioneering systematic studies [55•] attempting to go beyond the neutralist null hypothesis.

In prokaryotes, this diversity in change processes seems even more pronounced, often even among closely related organisms, with high rates of gene loss, often environmentally explicable, and gene gain caused by horizontal

transfer adding to the gene order differences caused by inversion (again possibly with a predominance of short inversions) and other internal movements.

At the analytical level, although there is much ongoing work on modeling, statistical analysis and algorithmics, application of these methods to large nuclear genomes has been impeded by the methodological prerequisites for complete information on orthology. Thus a preoccupation with whole-genome alignment, gene annotation and ortholog identification has postponed, until recently, the upscaling of these methods to large genomes. Breakthroughs in the comparison of human, mouse, rat and other genomes at many scales simultaneously [53••] and in the adaptation of gene-order rearrangement methods to genome-sequence data [12••] should clear this logjam.

Acknowledgements

This work supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada. The author is a Fellow of the Program in Evolutionary Biology of the Canadian Institute for Advanced Research and holds the Canada Research Chair in Mathematical Genomics.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Sankoff D, Nadeau JH (Eds): *Comparative Genomics: Gene Order Dynamics, Map Alignment and the Evolution of Gene Families*. Dordrecht, NL: Kluwer; 2000.

2. • Sankoff D, El-Mabrouk N: **Genome rearrangement**. In *Current Topics in Computational Biology*. Edited by Jiang T, Smith T, Xu Y, Zhang M. Cambridge, MA: Massachusetts Institute of Technology Press; 2002:135-155.
A review of the combinatorial optimization algorithms inspired by the study of genome rearrangements.

3. Gonzalez J, Ranz JM, Ruiz A: **Chromosomal elements evolve at different rates in the *Drosophila* genome**. *Genetics* 2002, **161**:1137-1154.

4. Sharakhov IV, Serazin AC, Grushko OG, Dana A, Lobo N, Hillenmeyer ME, Westerman R, Romero-Severson J, Costantini C, Sagnon N *et al.*: **Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus***. *Science* 2002, **298**:182-185.

5. Zdobnov EM, von Mering C, Letunic I, Torrents D, Suyama M, Copley RR, Christophides GK, Thomasova D, Holt RA, Subramanian GM *et al.*: **Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster***. *Science* 2002, **298**:149-159.

6. Coghlan A, Wolfe KH: **Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila***. *Genome Res* 2002, **12**:857-867.

7. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S *et al.*: **Genome sequence of the human malaria parasite *Plasmodium falciparum***. *Nature* 2002, **419**:498-511.

8. Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P *et al.*: **Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi***. *Nature* 2001, **414**:450-453.

9. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES: **Sequencing and comparison of yeast species to identify genes and regulatory elements**. *Nature* 2003, **423**:241-254.

10. Seoighe C, Federspiel N, Jones T, Hansen N, Bivolarovic V, Surzycki R, Tamse R, Komp C, Huizar L, Davis RW *et al.*: **Prevalence of small inversions in yeast gene order evolution**. *Proc Natl Acad Sci USA* 2000, **97**:14433-14437.

11. McLysaght A, Seoighe C, Wolfe KH: **High frequency of inversions during eukaryote gene order evolution**. In *Comparative Genomics*. Edited by Sankoff D, Nadeau JH. Dordrecht NL: Kluwer Academic Press; 2000:47-58.

12. •• Pevzner P, Tesler G: **Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution**. *Proc Natl Acad Sci USA* 2003, **100**:7672-7677.
A breakthrough in adapting rearrangement algorithms to nuclear genomes, bypassing gene-finding and ortholog identifications. These findings are suggestive of new directions for modeling, algorithms, and further comparative sequencing.

13. • Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE: **Recent segmental duplications in the human genome**. *Science* 2002, **297**:1003-1007.
An assessment of the particular role of segmental duplications in recent primate history. These are associated with gene diversification, regions of genomic instability and genetic disease.

14. Mefford HC, Trask BJ: **The complex structure and dynamic evolution of human subtelomeres**. *Nat Rev Genet* 2002, **3**:91-102. [Published erratum appears in *Nat Rev Genet* 2002, **3**:229.]

15. Samonte RV, Eichler EE: **Segmental duplications and the evolution of the primate genome**. *Nat Rev Genet* 2002, **3**:65-72.

16. Thomas JW, Schueler MG, Summers TJ, Blakesley RW, McDowell JC, Thomas PJ, Idol JR, Maduro VV, Lee-Lin SQ, Touchman JW *et al.*: **Pericentromeric duplications in the laboratory mouse**. *Genome Res* 2003, **13**:55-63.

17. Salse J, Piegu B, Cooke R, Delseny M: **Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project**. *Nucleic Acids Res* 2002, **30**:2316-2328.

18. Wong S, Butler G, Wolfe KH: **Gene order evolution and paleopolyploidy in hemiascomycete yeasts**. *Proc Natl Acad Sci USA* 2002, **99**:9272-9277.

19. Bowers JE, Chapman BA, Rong J, Paterson AH: **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events**. *Nature* 2003, **422**:433-438.

20. Hughes AL, Friedman R: **2R or not 2R: testing hypotheses of genome duplication in early vertebrates**. *J Struct Funct Genomics* 2003, **3**:85-93.

21. Gu X, Wang Y, Gu J: **Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution**. *Nat Genet* 2002, **31**:205-209.

22. • Hokamp K, McLysaght A, Wolfe KH: **The 2R hypothesis and the human genome sequence**. *J Struct Funct Genomics* 2003, **3**:95-110.
An account of the current status of the genome doubling controversy pertaining to the vertebrate lineage, with critique of both phylogeny-based and map-based arguments. The authors use a whole-genome approach (at the gene-order level) to establish the likelihood of a tetraploidization event early in vertebrate history.

23. McLysaght A, Hokamp K, Wolfe KH: **Extensive genomic duplication during early chordate evolution**. *Nat Genet* 2002, **31**:200-204.

24. Durand D: **Vertebrate evolution: doubling and shuffling with a full deck**. *Trends Genet* 2003, **19**:2-5.

25. Hsiao W, Wan I, Jones SJ, Brinkman FS: **IslandPath: aiding detection of genomic islands in prokaryotes**. *Bioinformatics* 2003, **19**:418-420.

26. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**:860-921.

27. Salzberg SL, White O, Peterson J, Eisen JA: **Microbial genes in the human genome: lateral transfer or gene loss?** *Science* 2001, **292**:1903-1906.

28. Doolittle WF, Boucher Y, Nesbo CL, Douady CJ, Andersson JO, Roger AJ: **How big is the iceberg of which organellar genes in nuclear genomes are but the tip?** *Philos Trans R Soc Lond B Biol Sci* 2003, **358**:39-57.

29. Kunin V, Ouzounis CA: **The balance of driving forces during genome evolution in prokaryotes**. *Genome Res* 2003, **13**:1589-1594.

30. Amor DJ, Choo KH: **Neocentromeres: role in human disease, evolution, and centromere study**. *Am J Hum Genet* 2002, **71**:695-714.

31. Tamames J: **Evolution of gene order conservation in prokaryotes**. *Genome Biol* 2001, **2**:RESEARCH0020.

32. Durand D, Sankoff D: **Testing for gene clusters**. *J Comput Biol* 2003, **10**:453-482.

33. Bergeron A, Corteel S, Raffinot M: **The algorithmic of gene teams**. In *Algorithms in Bioinformatics, Proceedings of the Second International Workshop*. Edited by Gusfield D, Guigo R. Berlin: Springer Lecture Notes in Computer Science 2452; 2002:464-476.

34. Sankoff D: **Short inversions and conserved gene clusters**. *Bioinformatics* 2002, **18**:1305-1308.

35. Lefebvre J-F, El-Mabrouk N, Tillier E, Sankoff D: **Detection and validation of single-gene inversions**. *Bioinformatics* 2003, **19**:i190-i196.

36. Bader DA, Moret BM, Yan M: **A linear-time algorithm for computing inversion distance between signed permutations with an experimental study**. *J Comput Biol* 2001, **8**:483-491.

37. Bergeron A: **A very elementary presentation of the Hannenhalli-Pevzner theory**. *Discrete Applied Mathematics* 2003, in press.

38. Tesler G: **GRIMM: genome rearrangements web server**. *Bioinformatics* 2002, **18**:492-493.

39. El-Mabrouk N: **Sorting signed permutations by reversals and insertions/deletions of contiguous segments**. *J Discrete Algorithms* 2001, **1**:105-122.

40. Marron M, Swenson KM, Moret BME: **Genomic distances under deletions and insertions**. *Proceedings of COCOON, the Ninth International Computing and Combinatorics Conference: 2003*; Berlin: Springer Lecture Notes in Computer Science; **2697**:537-547.

41. Sankoff D: **Genome rearrangement with gene families**. *Bioinformatics* 1999, **15**:909-917.

42. Bourque G, Pevzner PA: **Genome-scale evolution: reconstructing gene orders in the ancestral species**. *Genome Res* 2002, **12**:26-36.

43. • Tang J, Moret BME: **Scaling up accurate phylogenetic reconstruction from gene-order data**. *Bioinformatics* 2003, **19**:i305-i312.

The authors review ongoing efforts to develop rearrangement-based phylogenetics and report on experiments to infer larger trees.

44. Sankoff D, Blanchette M: **Comparative genomics via phylogenetic invariants for Jukes-Cantor semigroups**. In *Stochastic Models: A Conference in Honour of Professor Donald A Dawson*. Edited by Gorostiza LG, Ivanoff BG. Providence, RI: American Mathematical Society; 2000:399-418.

45. Wang LS: **Genome rearrangement phylogeny using Weighbor**. In *Algorithms in Bioinformatics, Proceedings of the Second International Workshop*. Edited by Gusfield D, Guigo R. Berlin: Springer Lecture Notes in Computer Science 2452; 2002:112-125.

46. York TL, Durrett R, Nielsen R: **Bayesian estimation of the number of inversions in the history of two chromosomes**. *J Comput Biol* 2002, **9**:805-818.

47. Miklós I: **MCMC genome rearrangement**. *Bioinformatics* 2003, in press.

48. El-Mabrouk N, Sankoff D: **The reconstruction of doubled genomes**. *SIAM J Computing* 2003, **32**:754-792.

49. Sankoff D, Parent MN, Bryant D: **Accuracy and robustness of analyses based on numbers of genes in observed segments**. In *Comparative Genomics: Gene Order Dynamics, Map Alignment and the Evolution of Gene Families*. Edited by Sankoff D, Nadeau JH. Dordrecht, NL: Kluwer; 2000:299-306.

50. Housworth EA, Postlethwait J: **Measures of synteny conservation between species pairs**. *Genetics* 2002, **162**:441-448.

51. Sankoff D, Deneault M, Turbis P, Allen CP: **Chromosomal distributions of breakpoints in cancer, infertility and evolution**. *Theor Popul Biol* 2002, **61**:497-501.

52. • Pevzner PA, Raphael B: **Genome rearrangements in tumor cells: reconstructing tumor genome architecture**. *Bioinformatics* 2003, in press.

The introduction of rearrangement algorithmics to analysis of the development of a tumour genome.

53. •• Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D: **Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes**. *Proc Natl Acad Sci USA* 2003, **100**:11484-11489.

A comprehensive account of the similarities and differences between the human and mouse genome sequences, based on a judicious approach to prioritizing competing local alignments and to embedding shorter alignments within gaps of larger alignments. Gives the first attempt at inventorying the inversions, transpositions, deletions and other rearrangements, at scales of analysis from 10^2–10^7 bp. An illustration of the power of the UCSC genome browser.

54. Pal C, Hurst LD: **Evidence for co-evolution of gene order and recombination rate**. *Nat Genet* 2003, **33**:392-395.

55. • Lercher MJ, Urrutia AO, Hurst LD: **Clustering of housekeeping genes provides a unified model of gene order in the human genome**. *Nat Genet* 2002, **31**:180-183.

A systematic approach to questions of functional non-randomness in gene order, suggesting that genes cluster according to tissue-specific expression patterns.