# Vertebrate genome evolution: a slow shuffle or a big bang?

Nick G.C. Smith,[1]* Robert Knight,[2] and Laurence D. Hurst[1]

Summary

In vertebrates it is often found that if one considers a group of genes clustered on a certain chromosome, then the homologues of those genes often form another cluster on a different chromosome. There are four explanations, not necessarily mutually exclusive, to explain how such homologous clusters appeared. Homologous clusters are expected at a low probability even if genes are distributed at random. The duplication of a subset of the genome might create homologous clusters, as would a duplication of the entire genome. Alternatively, it may be adaptive for certain combinations of genes to cluster, although clearly the genes must have duplicated prior to rearrangement into clusters. Molecular phylogenetics provides a means to examine the origins of homologous clusters, although it is difficult to discriminate between the different explanations using current data. However, with more extensive sequencing and mapping of vertebrate genomes, especially those of the early diverging chordates, it should soon become possible to resolve the origins of homologous clusters. *BioEssays* 21:697–703, 1999.


## Introduction

It is often found that if one considers a cluster of genes in eukaryotes, then the homologs (assigned on the basis of sequence similarity) of those genes often form another cluster on a different chromosome.(1) The regions encompassing the Hox clusters are a well-known example.(2) We use the term ''homologous clusters'' to indicate such sets of clusters. At present, it is not known either how or why homologous clusters came to exist. Here we review the evidence with regard to a number of hypotheses, and tentatively suggest directions for future research.

The most popular explanations for the origins of homologous clusters invoke block duplication, either of the entire genome, or of a subset of the genome.(3) For example, it now appears that the Hox clusters have arisen as the result of three separate duplication events.(4) Such explanations suggest that homologous clusters arise in a single event (a ''big bang''), and have not been obscured by the subsequent effects of mutation and selection.

An alternative hypothesis suggests that homologous clusters may have arisen because of their adaptive value. A recent article by Hughes(5) on the phylogeny of genes in one family of clusters extends previous results(6) and claims to provide support for the adaptive theory, under which gene duplication and homologous cluster formation need not occur simultaneously. Homologous genes are generated by duplication, and then a series of genome rearrangements take place, which create adaptive clusters of linked genes (a "slow shuffle"). If selection favours similar sets of homologous genes come together, then homologous clusters are the result.

Coincidence may seem a poor explanation of homologous clusters, but until we know how prevalent homologous clusters are, then the effects of chance cannot be discounted. After all, given the huge number of genes in the vertebrate genome, and the existence of many families of homologous genes, the existence of a few homologous clusters is only to be expected.

## Alternative models applied to a specific case

We shall consider alternative explanations for the homologous clusters recently investigated by Hughes.(5) The clusters

[1]Department of Biology and Biochemistry, University of Bath, Bath, UK.
[2]Division of Zoology, School of Animal and Microbial Sciences, University of Reading, Reading, UK.
*Correspondence to: Nick Smith, School of Biological Sciences, University of Sussex, Brighton, BN1 9QG, UK. E-mail: n.g.c.smith@sussex.ac.uk

in question are found on chromosomes 6 (6q21.3), 9 (9q33–34), and 1(1p21–25 1p13) in humans, which are homologous with those on mouse chromosome 17, 2, and 1 respectively.[7,8] The regions on human chromosome 6 and mouse chromosome 17 are both near the major histocompatibility complex (MHC). There is some evidence for a fourth cluster in humans on chromosome 19.[7,8] These homologous clusters contain sequenced genes from nine families: Retinoid X Receptor (RXR) genes, Collagen (COL) genes, ATP-binding Cassette Transporter (ABC) genes, Proteasome Component β (PSMB) genes, Notch (NOTCH) genes, Pre-B-cell Leukemia Transcription Factor (PBX) genes, Tenascin (TEN) genes, C3/C4/C5 Complement Component (C3/C4/C5) genes, and Heat Shock Protein 70 (HSP70) genes (gene family abbreviations in parentheses as given by Hughes[5]). Coincidence seems unlikely to explain the existence of four homologous clusters, although we cannot rule out this possibility by considering a single case.

Such a pattern of homologous clusters can be explained by two or three independent duplication events, depending on whether the duplications were of the entire genome or a subset of the genome (see Fig. 1A). An initial duplication event affecting a single chromosomal region (D1) yielded two duplicated regions in different parts of the genome. Duplications of both of these regions then produced the homologous clusters now found on human chromosomes 1, 9, 6, and 19. The second stage of duplications could have been a single polyploidisation event (D2 and D3 simultaneous) with two duplications required overall, or two separate block duplications (D2 and D3 separate) with three duplications required overall.

Hughes reasoned that if block duplication explains the patterns, then a phylogeny of the genes should reflect this.[5] For simplicity we shall call the genes from the nine gene families *A* to *I*, with a suffix being used to indicate the human chromosome (i.e., *A6* is the *A* gene on human chromosome 6). If block duplication explains the pattern then a phylogenetic analysis should reveal the same time to common ancestor for *A6* and *A9* (i.e., time back to initial duplication event), as would be found from a phylogenetic reconstruction of the time of common ancestry of *B6* and *B9*, *C6* and *C9*, and so on. Further if the block duplication were the result of tetraploidization at the base of the vertebrates, as often conjectured,[3] there should be no non-vertebrate genes appearing in the phylogeny within the groupings defined by branches for the genes on chromosomes 6, 19, 9, and 1 (see Fig. 1B). All the invertebrate genes should appear as a mass sister grouping to the vertebrate copies.

If, alternatively, *A6* and *A9* diverged at the origin of the eukaryotes (for example) and independently came into a cluster with *B* genes, *C* genes etc., then human *A6* might well be more closely related to the *Drosophila* homolog of *A6* than to human *A9* (see Fig. 1C). At the same time, *B6* and *B9* could have had a different time to a common ancestor altogether.

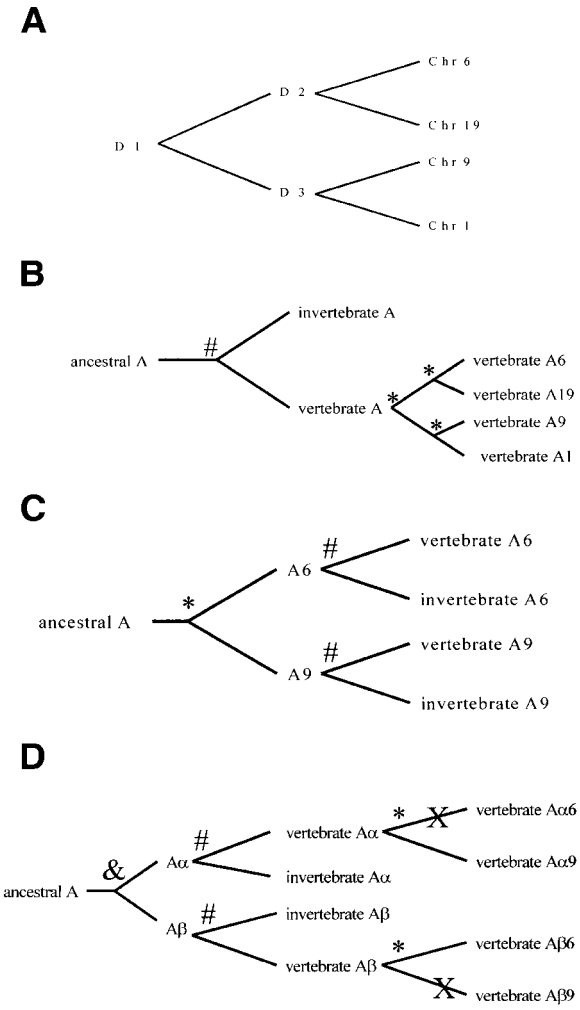


Figure 1. A scheme for the origin of duplicated regions on human chromosomes 6, 19, 1, and 9 is given in A. Three duplication events are given by D1, D2, and D3. The ancestral chromosome duplicated to give two chromosomes, both of which then duplicated to give the four present day chromosomes. B illustrates the fact that if related gene clusters have arisen in the vertebrate lineage, then the divergence times of such regions in vertebrates should be after the origin of the vertebrates. The split of the vertebrates and invertebrates is represented by #, and block duplications are given by *. C shows that if gene duplications occurred before the origin of the vertebrates, then invertebrate genes could cluster with vertebrate genes. D demonstrates how gene loss can lead to misleading estimates of divergence time. If the two remaining copies of gene A in the vertebrates are Aα9 and Aβ6, then their time of divergence will be prior to the time of the vertebrate-invertebrate split, despite block duplications involving A occurring in the vertebrate lineage. An ancient tandem duplication is given by &, and X portrays gene loss.

What patterns of molecular phylogeny did Hughes find for the gene families in the homologous clusters? For three to five of the genes, early vertebrate duplication (prior to the divergence of jawed and jawless vertebrates) is found, as expected of block duplication and/or early vertebrate polyploidization. But the remaining four showed wide ranging patterns, with divergences all before the origin of the vertebrates, some going back to the common ancestor of animals and fungi and one even going back as far as the eukaryote-eubacterial common ancestor.

This, Hughes argues, ''decisively rejects'' the hypothesis of block duplication. However the null hypothesis that he rejects supposes that no gene loss follows a block duplication. Imagine instead that gene *A* in the ancestral vertebrate genome existed in two tandem copies, rather than one. These we can call *A*α and *A*β (see Fig. 1D). The tandem duplication event might have occurred any time prior to the vertebrate expansion: at the origin of eukaryotes, at the origins of animals, whenever. We consider a tandem duplication event because the two duplicates would remain closely linked. Now imagine that after the block duplication *A*α is lost from the pre-chromosome 6, but that on the pre-chromosome 9 *A*β is lost. Under the methodology employed by Hughes such a pattern of events will lead to rejection of the hypothesis of early vertebrate block duplication. This is because when we do a phylogeny of the *A* genes, the one left on chromosome 6 (*A*β) has a most recent common ancestor with that on chromosome 9 (*A*α) sometime well before the vertebrates, at the time of the duplication of the original A gene into *A*α and *A*β.

Such pre-vertebrate tandem duplication events followed by reciprocal loss (*A*α on one and *A*β on the other) yield a model consistent with vertebrate block duplication (see Fig. 2A and B).[7,8] The block duplication model also requires that the clusters on 1 and 9 have a recent common ancestry and that the loss events occurred before this duplication. The phylogenetic evidence is consistent with this (note similar patterns of gene loss on chromosomes 1 and 9 in Fig. 2B).

A similar argument invoking gene loss has been used to account for the inconsistent phylogenetic data of related clusters of insulin group genes and aromatic amino acid hydroxylase genes.[9] Duplication dates of some of these genes prior to the origin of the vertebrates would seem, at first, to invalidate the hypothesis of block duplication at the base of the vertebrates. When gene loss and gene conversion are taken into account, however, the results are consistent with block duplication.

## Discriminating between alternative models in the specific case

From the analysis above it appears that a model of block duplication that incorporates gene loss cannot be rejected, even if a model without gene loss can be rejected. Which model is the most consistent with existing data? Hughes rightly complains that advocates of block duplication often come up with rather elaborate post-hoc hypotheses to argue their way out of tight corners. It is certainly awkward to have to propose the existence of presently unidentified duplications. But then how parsimonious is it to suppose that three to five of the clustered genes duplicated at the same time, but that there was no block duplication?

The assumption of no gene loss following gene duplication seems unreasonably prohibitive. The generation of identical duplicate genes implies that the loss of one of the two copies is unlikely to have deleterious effects. From the distributions of the number of genes in human gene families Nadeau and Sankoff have estimated that the rate of gene loss is about the same as the rate of functional retention and divergence.[10] In other words about 50% of duplicates are driven to non-functionality by mutation. There are two possible routes for functional duplicates to be retained. New functions may evolve, especially if many genes are multifunctional.[11] Alternatively, the duplicates may be retained as a buffer against developmental error.[12,13] However, these routes are unlikely to be available to all genes, which means that a certain proportion of duplicates will be lost. Further evidence of gene loss comes from Hox cluster genes: different paralogs have been lost in the different clusters, and furthermore patterns of gene loss differ between lineages (e.g., man and pufferfish[14]). In light of the scheme we presented above to suggest how gene loss can lead to the misallocation of paralogy (Fig. 1D), the fact that gene loss does occur highlights the need for caution in assigning paralogy without adequate phylogenetic evidence. By adequate phylogenetic evidence we mean that one should include groups of organisms that enable the reliability of paralogy assignments to be tested (such as early chordates, see below).

In favour of the block duplication hypothesis, the timings of the gene duplication and gene loss events proposed in Figure 2B seem reasonable. If ancient tandem duplications have been maintained for many millions of years, we expect many of the tandem duplicates to have evolved different functions. Only when further duplications occur can gene loss be expected. The reciprocal nature of gene loss, whereby one α copy and one β copy is lost, also seems reasonable given the different functions of the α and β copies.

Additionally, if the criterion in model discrimination is parsimony, it should be noted that Hughes' alternative model seems even more ornately baroque than the block duplication model, and gene loss. One must suppose that five of the genes (say *A*, *B*, *C*, *D*, and *E*) underwent duplication at the same time (in a block or separately) but that the others (*F*, *G*, *H*, and *I*) had duplicated earlier. If the first five were block duplicated then one must suggest that one copy of *F* goes to each of the four homologous clusters. If there was no block duplication one must suggest that all nine gene families

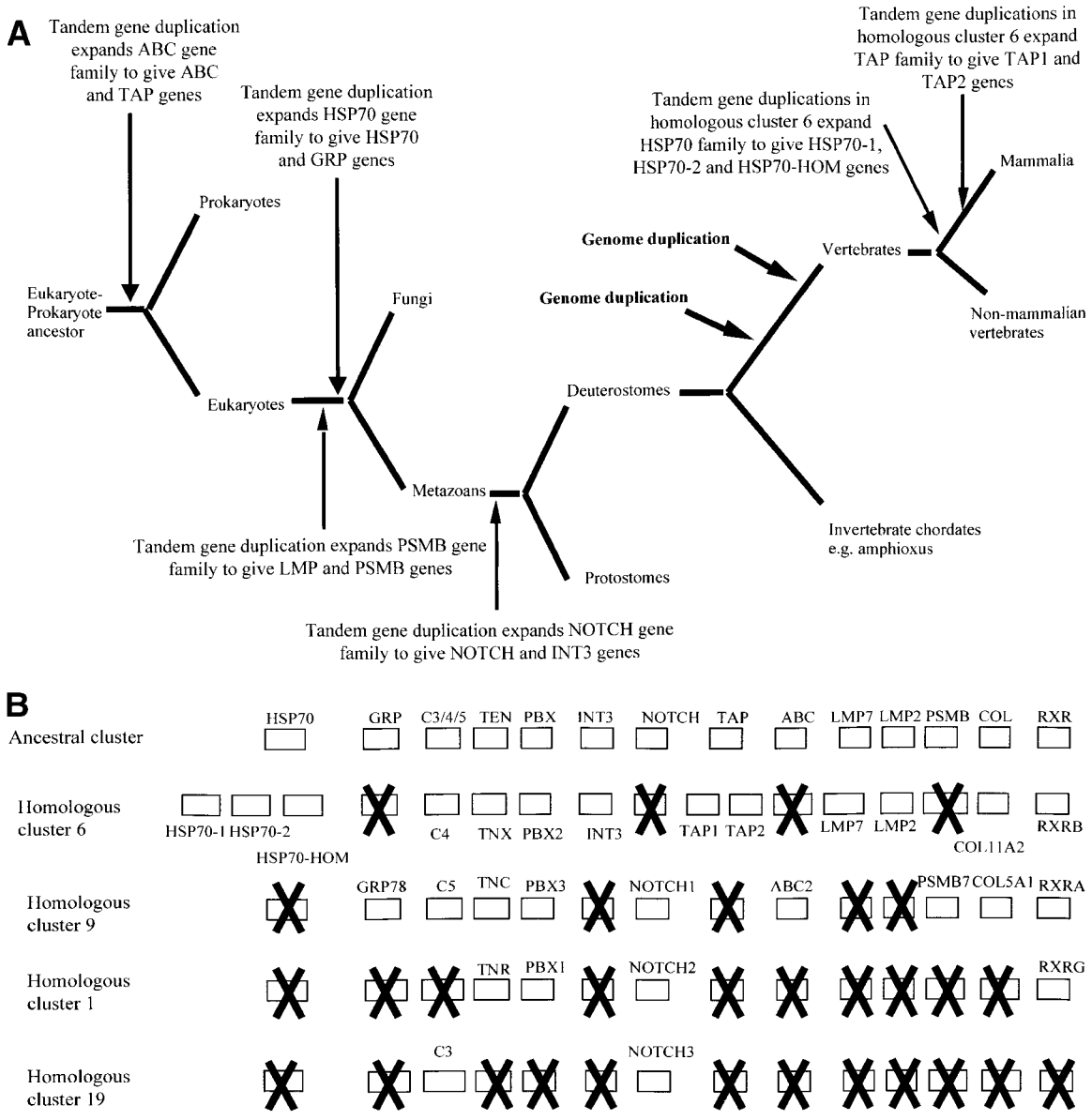


Figure 2. A model for the origin of the homologous clusters seen in humans at 1p13, q21–25, 9q33–34, 6p21.3, and 19p13.3, involving chromosomal duplications (A), tandem gene duplications (A), and gene loss (B). Early tandem duplications which occur prior to the vertebrates can be estimated from the topology of the gene trees. An initial duplication of the ancestral cluster in the vertebrate ancestor would generate the precursors for homologous clusters 6/19 and 1/9, and a second round of duplications at a later point generated the homologous clusters at 1, 9, 6, and 19 (see A). After each chromosomal duplication, extensive gene loss creates the pattern observed today (see B). Crosses indicate that either the gene has been lost or not yet found (in particular, many genes may be undiscovered within homologous cluster 19). The gene family abbreviations are given in the text.

independently came together with at least one member of each gene family in each homologous cluster.

There is one issue on which block duplication and adaptive models of homologous cluster formation are agreed: a cluster of genes must have formed at some point. The question then becomes whether the cluster came together once (block duplication hypothesis) or several times (adaptive hypothesis). Hybrid models, in which the block duplication and adaptive hypotheses are combined, are also possible. The original cluster might have come together for adaptive reasons, and then block duplications might have created the homologous clusters. Alternatively, block duplica-

tions might have created many homologous clusters, of which only those favoured by selection now remain (L. Lundin, personal communication).

Why might the original cluster have formed, or why might homologous clusters be maintained? It might be selectively advantageous for genes whose products interact to be linked, perhaps in order to enable better regulation. Such an interaction of genes within an homologous cluster appears to apply to the present case. The Proteasome Component β gene family contains the two genes *LMP2* and *LMP7*. These combine to make the LMP+ proteasome that breakdown proteins into peptides that are presented by MHC class I molecules. The peptides are transported across the endoplasmic reticulum (ER) membrane by a dimeric transporter Tap. The *Tap* gene belongs to the ATP-binding Cassette Transporter gene family. Of course it may be that this relationship between members of the homologous clusters may be just a coincidence.

However, the combination of functional interaction and physical linkage is by no means unique (for a review see Reference 15). For example, in *Caenorhabditis*[16] two different enzymes are needed for trimerizing collagen and these are encoded within one operon. The best data for selection on linkage come from meiotic drive genes[17] and supergene complexes.[18] The problem with such examples is that we are aware of no work that reliably predicts the regularity of such coincidences.

As for the evolutionary history of the homologous clusters studied by Hughes, we seem to be left with two models, one of block duplication and one of selection, both of which are really quite complicated. How can we discriminate between the two alternatives? One potentially informative piece of evidence concerns the form of the ancestral vertebrate clusters. Under the block duplication and gene loss model tandem duplications prior to the origin of the vertebrates are required for four gene families (Fig. 2A and B). Tandem duplications lead to close linkage of duplicates, which makes the misallocation of paralogy more likely (Fig. 1D). But the alternative model of ancient duplications followed by gene shuffling makes no specific prediction of tandem duplicates within the gene clusters.

Hence, if one examines the ancestral cluster in an early diverging chordate such as amphioxus or a tunicate and finds tandem duplications within the cluster, then that is good evidence in favour of the block duplication and gene loss model. Furthermore if one assumes that gene conversion has not taken place, then the time of common ancestry of such tandem duplicates and, say, human genes should be the time of the divergence of the early chordates. On the other hand, the absence of tandem duplicates in early chordates might be due to gene loss, and does not necessarily favour the adaptive hypothesis. Tandems duplicates may well have been maintained for many millions of years (from the ancient duplication event until the divergence of the early chordates), and then lost in early diverging chordates due to changing selective pressures, if the biotic environment was rapidly evolving at the time of the vertebrate radiation.

## General explanations for the evolution of homologous clusters

In the specific case described earlier, our basis for discriminating between the alternative models of homologous cluster evolution was how well the models fitted the predicted molecular phylogenies. The problem with this general method lies in the assumption that our predicted molecular phylogenies are correct. We have already shown how the correct identification of paralogs is complicated if gene duplication is followed by random gene loss. In addition, tree reconstruction is sensitive to differences in evolutionary rates. Long branch attraction, whereby the longest branches are inferred as outgroups irrespective of the true phylogeny, can be the result of variation in evolutionary rates.[19] Furthermore, the detection of fast evolving sequences is hampered by mutational saturation.[20] Differences in rates of evolution between genes within a homologous cluster could be the result of differences in mutation or selection. Given that one of the ways in which newly formed duplicates survive is through the acquisition of new roles, directional selection is likely to have affected many of the genes in the homologous clusters. Even homeobox genes, commonly thought to be highly conserved and subject to only purifying selection, have been affected by positive selection.[21]

Putting aside problems of tree reconstruction, is there any way we can decide between the different hypotheses for the general phenomenon of homologous clusters? Until we attempt to address general explanations for the evolution of homologous clusters our understanding is restricted to the level of anecdote. In principle, it might be possible to evaluate the likelihoods of evolutionary histories such as that presented in Figure 2B, if one had sufficient information on mutational biases. It should be possible to determine the relative frequencies of gene loss, different sorts of translocations, and different sorts of duplications. For example, comparative mapping data have been used to estimate distributions of rearrangement breakpoints.[22] From such mutational data, it should be possible to predict the null (coincidence) expectation for the frequency of homologous clusters. Only if the null expectation is significantly less than the observed frequency need we consider any further explanations.

However, the mutational processes involved in the evolution of homologous clusters are unlikely to occur evenly across entire genomes, and will almost certainly be subject to local effects, which will complicate matters. As an illustration of the problems inherent in determining mutational biases, consider gene loss. Developmental genes are thought to be unusual in that relatively simple mutations, either in the gene itself or in cis-regulatory elements, can provide new functions for otherwise redundant duplicated genes. Many developmen-

tal proteins have a modular structure, which allows temporal or spatial expression patterns to be altered by relatively small mutational changes, while the evolutionarily conserved core functions are retained.[23] In contrast, metabolic genes have very limited possibilities of gaining a new function prior to loss by mutational drift due to their rigid functional requirements. Thus the relative likelihood of gene loss following duplication will depend on the class of the duplicated gene, as well as other factors such as the extent to which there is the potential and the need for further adaptation of the gene in question.

Could the use of gene order data help us to understand the evolution of homologous clusters? If gene order is conserved between homologous clusters then the probability that the homologous clusters arose by chance becomes even more remote. Such an approach has been used by Pebusque et al.[24] to infer a quadruplicated region on human chromosomes 4, 5, 8, and 10. They started from a region on chromosome 8 for which they possessed detailed mapping information, then searched for paralogs, and finally deduced duplication events from mapping and phylogenetic data. Such a study suggests that duplicated regions may be common in mammalian genomes, and support reports of extensive regional duplications (see Reference 25), although the criteria commonly used to infer paralogy are perhaps too broad.[26]

## The adaptive hypothesis vs. the block duplication hypotheses

Can gene order help to decide between the adaptive and block duplication hypotheses? It appears not, since the two hypotheses cannot be readily discriminated on the basis of linkage patterns of related clusters. Local gene shuffling by inversions can upset the initially conserved gene order of block duplicated clusters (as in yeast[27]), and it is not clear whether adaptive gene clustering is likely to imply adaptive gene ordering. In the case of the homologous clusters described above, gene order appears to be weakly conserved between the clusters on human chromosomes 6 and 9.[6]

The preponderance of homologous clusters may well affect the likelihood of the alternative hypotheses. The existence in yeast of 55 duplicated regions extending across half the genome[27] strongly suggests that a block duplication event is responsible. Two problems suggest that the adaptive hypothesis is unlikely to explain so many duplicated regions. First, it may be reasonable to suppose that selection might favour physical linkage of some genes, but is such selection likely for a high proportion of all genes? Second, has enough time elapsed for sufficient mutational events to have caused a thorough reshuffling of the entire genome?

Further suggestive evidence might be obtained from consideration of the intronic content of genes within related clusters. If the multiple copies of a gene have arisen by block duplication, then one would expect similar intronic contents. However, if multiple copies have been spread around the genome by retrotransposition, as is consistent with the adaptive hypothesis, then one copy might contain introns whereas the other copies are intronless. Therefore, a preponderance of intronless genes would provide suggestive evidence in favour of the adaptive hypothesis, unless it could be demonstrated that the original, later to be duplicated, cluster contained intronless genes, in which case all copies should be intronless. This proposed test is unlikely to be conclusive because retrotransposition is not the only mode of gene movement, and because subsequent intron evolution may well have obscured any intron patterns, which might have been generated during homologous cluster evolution.

## Polyploidisation or local block duplications?

If the block duplication hypothesis for the evolution of homologous clusters is favoured over the adaptive hypothesis there remains the question of whether the block duplications were small and independent, or whether they involved the entire genome. One approach would be to consider the estimated time of duplication for a number of homologous clusters. The timing of a duplication could be estimated from the shape of the tree, by comparing the duplication event with the appearance of various groups. This method is dependent on the accuracy of tree reconstruction.

In the case of yeast, Wolfe and Shields[27] provided two reasons to believe that a polyploidisation event, rather than many independent and small block duplications, was responsible for the large number of duplicated regions. First, the orientation of both homologous clusters with respect to the centromere was the same in a significantly high proportion of cases. This suggests that the duplications were not independent. Second, no triplicated regions were found, which argues against a series of duplications spread over time.

## Conclusions

Until we have the complete genomes of a number of vertebrates, and in particular those of the early chordates, it will be difficult to evaluate general hypotheses for the evolution of homologous clusters. When we possess such information, we shall be able to address a number of questions. How common are homologous clusters within different genomes? Are there more homologous clusters than we would have expected by chance? What sort of genes does one find within the homologous clusters? Is gene order, and also orientation with respect to the centromere, conserved between homologous clusters? And finally, did these homologous clusters all arise at the same time, at some ancestral vertebrate polyploidisation event?

In the meantime (until about 2005 for the human genome[28]), we are restricted to the level of anecdote, and we should not infer from a few case studies assumptions about

the entire genome. One way in which case studies may prove profitable is in examining the hypothesis that homologous clusters may be adaptive. A fuller understanding of the regulation of genes within clusters might show how regulation is dependent on physical proximity, or perhaps even the relative position and orientations of a number of genes. The regulation of transgenic rearranged homologous clusters would then be different from the wild type, and a transgenic phenotype would provide concrete evidence that selection not only cares about the sequence of a gene but also its context with respect to other genes.

## Acknowledgments

The authors thank Peter Holland, Seb Shimeld, L. Lundin, Adam Wilkins, and two anonymous referees.

## References

1. Lundin LG. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics 1993;16: 1–9.
2. Ruddle FH, Bentley KL, Murtha MT, Risch N. Gene loss and gain in the evolution of the vertebrates. Development 1994;155–161.
3. Sharman AC, Holland PWH. Conservation, duplication, and divergence of developmental genes during chordate evolution. Netherlands J Zool 1996;46:47–67.
4. Bailey WJ, Kim J, Wagner GP, Ruddle FH. Phylogenetic reconstruction of vertebrate Hox cluster duplications. Mol Biol Evol 1997;14:843–853.
5. Hughes AL. Phylogenetic tests of the hypothesis of block duplication of homologous genes on human chromosomes 6, 9, and 1. Mol Biol Evol 1998;15:854–870.
6. Endo T, Imanishi T, Gojobori T, Inoko H. Evolutionary significance of intra-genome duplications on human chromosomes. Gene 1997;205: 19–27.
7. Kasahara M, Hayashi M, Tanaka K, Inoko H, Sugaya K, Ikemura T, Ishibashi T. Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. Proc Natl Acad Sci USA 1996;93:9096–9101.
8. Katsanis N, Fitzgibbon J, Fisher EMC. Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. Genomics 1996;35:101–108.
9. Patton SJ, Luke GN, Holland PWH. Complex history of a chromosomal paralogy region: Insights from amphioxus aromatic amino acid hydroxylase genes and insulin-related genes. Mol Biol Evol 1998;15:1373–1380.
10. Nadeau JH, Sankoff D. Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics 1997;147:1259–1266.
11. Wagner A. The fate of duplicated genes: loss or new function? Bioessays 1998;20:785–788.
12. Wilkins AS. Canalization: a molecular genetic perspective. Bioessays 1997;19:257–262.
13. Cooke J, Nowak MA, Boerlijst M, Maynard-Smith J. Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet 1997;13:360–364.
14. Holland PWH. Vertebrate evolution: something fishy about Hox genes. Current Biology 1997;7:R570–R572.
15. Hurst LD. The evolution of genomic anatomy. TREE 1998; 14:108–112.
16. Blumenthal T. Gene clusters and polycistronic transcription in eukaryotes. Bioessays 1998;20:480–487.
17. Lyttle TW. Segregation distorters. Ann Rev Genet 1991;25:511–557.
18. Ferris PJ, Goodenough UW. The mating-type locus of *Chlamydomonas reinhardtii* contains highly rearranged DNA sequences. Cell 1994;76:1135–1145.
19. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 1978;27:401–410.
20. Philippe H, Laurent J. How good are deep phylogenetic trees? Curr Opin Genet Dev 1998;8:616–623.
21. Ting CT, Tsaur SC, Wu ML, Wu CI. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 1998;282:1501–1504.
22. Nadeau JH, Sankoff D. Counting on comparative maps. Trends Genet 1998;14:495–501.
23. Duboule D, Wilkins AS. The evolution of 'bricolage'. Trends Genet 1998;14:54–59.
24. Pebusque M-J, Coulier F, Birnbaum D, Pontarotti P. Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol Biol Evol 1998;15:1145–1159.
25. http://www.cib.nig.ac.jp/dda/timanish/poster/poster.html.
26. Skrabanek L, Wolfe KH. Eukaryote genome duplication—where's the evidence? Curr Opin Genet Dev 1998;8:694–700.
27. Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 1997;387:708–713.
28. Rowen L, Mahairas G, Hood L. Sequencing the human genome. Science 1997;278:605–607.