Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Genetics & Development

Genome duplication and the origin of the vertebrate skeleton

GuangJun Zhang[1,*] and Martin J Cohn[1,2]

During vertebrate embryonic development, tissue patterning and differentiation are regulated by members of multigene families. Evolutionary expansion of these families is thought to have played a role in the evolution of anatomical complexity, including the origins of new cell and tissue types. A defining feature of vertebrates is an endoskeleton, the primary components of which are cartilage and bone. The molecular control of skeletal patterning has been the subject of intensive investigation for over two decades. More recently, comparative studies of organisms at key phylogenetic positions have highlighted the importance of gene duplication in the evolutionary diversification of connective tissues. Understanding the natural histories of gene families involved in skeletogenesis is therefore central to the issue of vertebrate skeletal evolution.

Addresses
[1] Department of Zoology, University of Florida, Cancer/Genetics Research Complex, P.O. Box 103610, Gainesville, FL 32610, United States
[2] Department of Anatomy & Cell Biology, University of Florida, Cancer/Genetics Research Complex, P.O. Box 103610, Gainesville, FL 32610, United States

Corresponding author: Cohn, Martin J (cohn@zoology.ufl.edu)

*Present address: Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States.

Current Opinion in Genetics & Development 2008, **18**:387–393

This review comes from a themed issue on
Pattern formation and developmental mechanisms
Edited by Ottoline Leyser and Olivier Pourquié

Available online 18th September 2008

0959-437X/$ – see front matter
Published by Elsevier Ltd.

DOI 10.1016/j.gde.2008.07.009

Introduction

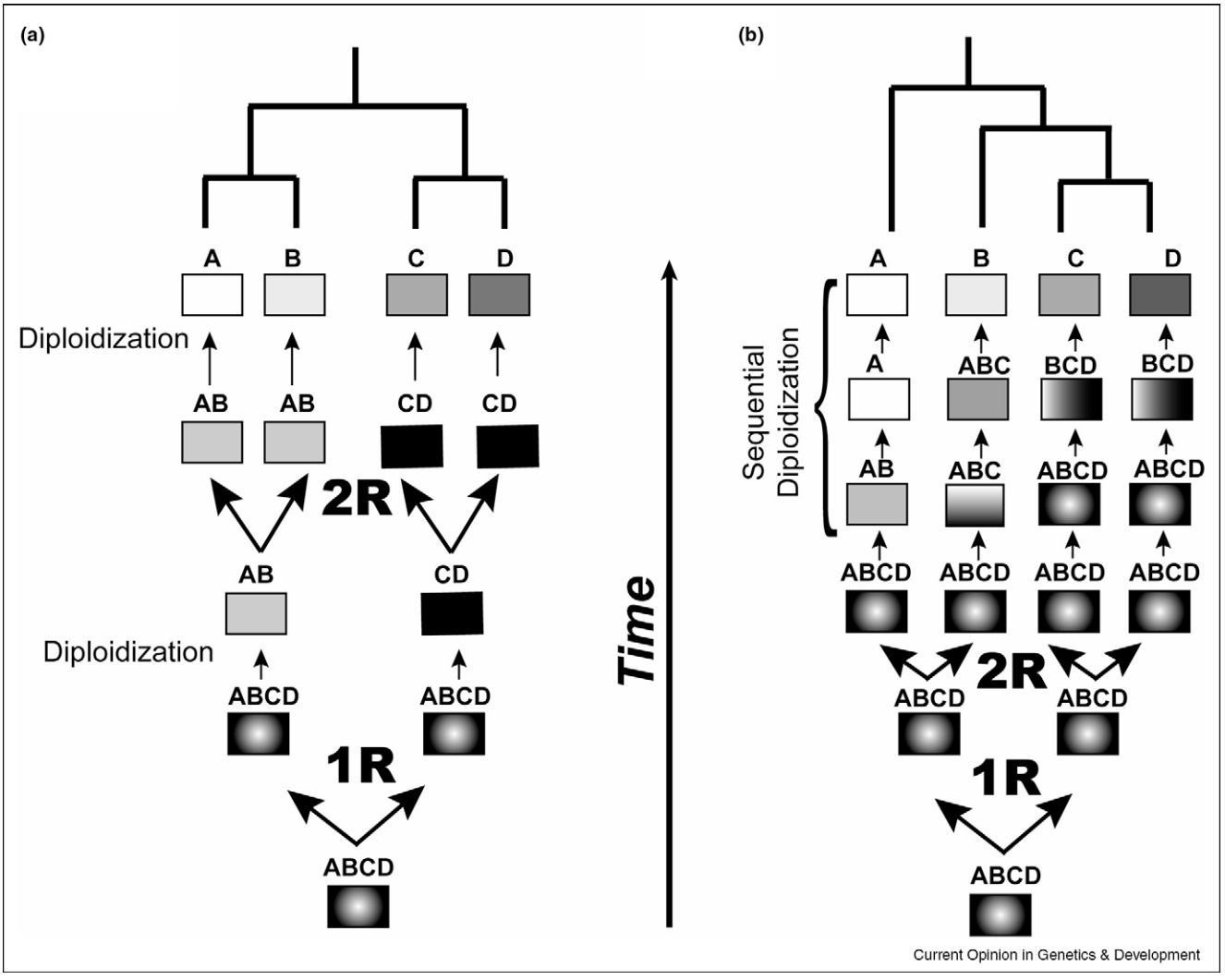
The molecular evolution of the genetic toolkit that regulates pattern formation is reflected in the broader history of the vertebrate genome. Since its publication in 1970, Susumu Ohno's paradigm-shifting book, *Evolution by Gene Duplication*, has had a profound influence on our thinking about the relationship between evolution of the genome and evolution of organismal design [1]. Ohno proposed that during the evolution of vertebrates, the entire genome underwent two rounds of polyploidization, which has come to be known as the 2R (for 'two rounds') hypothesis. Although there is compelling evidence from a multitude of phylogenetic and comparative genomic analyses that large-scale gene and/or whole genome duplications (WGDs) occurred, the 2R hypothesis has been a source of extensive debate. The recent sequencing of the lancelet genome provides strong support for two rounds of WGD occurring between the origin of chordates and the origin of jawed vertebrates [2••], though the precise timing of these events, the mechanisms by which polyploidy arose, and the effects of gene and genome duplications on the evolution of vertebrate development are open questions. In this review, we discuss the mechanisms of genome duplication, the state of the 2R hypothesis, and the implications of recent work on basal chordates and cyclostomes for our understanding of vertebrate skeletal evolution.

Whole genome duplication: cell biology and evolutionary signatures

Polyploidy can arise by several different mechanisms, some of which cause the duplication of the endogenous genome of a species, whereas others involve the fusion of related genomes. Interspecific hybridization can result in allopolyploidy, in which two diploid germ cells from different species fuse to generate two pairs of complementary chromosomes (the similarity of chromosomes duplicated by such fusions reflects a level of homology known as 'homeology'). Autopolyploidy, by contrast, refers to a doubling of the genome within a species, which can arise as a consequence of cellular anomalies such as chromosomal nondisjunction, endoreplication (incomplete division after DNA replication), or disrupted/delayed cell cycle progression [3]. Although the most extensive and informative work on polyploidization has been done in plants, examples are found throughout the animal kingdom, including most of the extant classes of vertebrates.

The mode of polyploidization is relevant to the question of how many duplications occurred during the evolutionary history of a given taxonomic group, especially when topology of gene trees is used to make inferences about patterns of duplication. For example, two rounds of WGD would be expected to result first in Gene ABCD giving rise to Genes A/B and C/D, after which A/B would produce A and B, and C/D would produce C and D. Thus, at first glance, two tetraploidy events (1:2 and 2:4) would be expected to result in a phylogenetic tree that groups the genes in a symmetrical pattern ((A,B),(C,D)) (Figure 1a). However, different mechanisms of polyploidization can yield different tree topologies [5–7]. Although two rounds of allotetraploidy would be expected to yield symmetrical trees, two rounds of autotetraploidy in close succession may result in asymmetrical trees, such as (A,(B,(C,D))), a

Figure 1



Symmetrical and asymmetrical gene tree topologies can arise from two rounds of genome duplication. (**a**) Two rounds of genome duplication involving allotetraploidization followed by diploidization can result in symmetrical genes trees. Symmetrical topologies can also follow autotetraploidization, if diploidization occurs before the second duplication event. (**b**) Autotetraploidy can produce asymmetrical tree topologies if the second duplication closely follows the first, before diploidization. After the second duplication, diplodization (and crossover) can then occur randomly and at different rates, resulting in sequential divergence of the homeologs. For further details, see [5–7].

topology often observed within vertebrate multigene families (Figure 1b). The underlying logic is relatively simple. If repeated rounds of tetraploidization are followed by complete diploidization (degeneration of the tetraploid genome such that it returns to the diploid state), then symmetrical trees would be predicted [5–7]. However if the second round of tetraploidization occurs more rapidly, before diploidization has been completed, then the four homeologous chromosomes should be able to pair freely in this 'octoploid' state. Gradual sequence divergence of the four 'homeologs' can lead to asymmetrical gene trees, and crossover can displace genes that were previously linked [5–7]. The importance of these models is that they illustrate how different biological mechanisms of genome duplication can leave a range of signatures in the genomes of descendant species, and decoding these signatures requires that one examine multiple lines of evidence, including gene number, organization and number of paralogy regions, tree topology, and tree congruence.

Fates of duplicated genes

The short-term consequences of polyploidy are crucial to a discussion of WGD as an evolutionary mechanism, as it is the survival and fitness of the organism harboring the newly duplicated genome that determines whether the polyploid state persists or becomes a dead end. For sexually reproducing organisms, an immediate challenge is embryo viability and, for those that develop normally, their disease state and

mating compatibility. In mammals, changes in ploidy most often lead to embryonic lethality or disease, though an interesting exception is the South American rat *Tympanoctomys barrerae*, which has two sex chromosomes and 100 autosomes [4]. Studies of experimentally induced polyploidy in rodents have shown that tolerance to polyploidy can be influenced by genetic background [8]. For organisms that are viable (and do not suffer from disease) following a WGD event, population genetics, mating compatibility, and fitness pose further challenges to the propagation of the tetraploid genome [9••].

Several potential outcomes may follow gene duplication events. Most often, duplicated genes are lost before they become fixed in the population, and those that do become fixed are frequently eliminated by degenerative mutation accumulation [10,11]. Widespread gene loss is one factor that contributes to asymmetrical tree topology and to the large number of vertebrate gene families that contain fewer than four paralogs [10–13]. The fraction of duplicated genes that are retained has been the focus of considerable attention from investigators interested in the evolution of developmental mechanisms, largely because it is these duplicates that have the potential to acquire novel functions, a process known as neofunctionalization. The classical view of neofunctionalization is that so long as one copy of the gene retains its original function, the other copy would be released from selective pressure and, therefore, would have new freedom to acquire novel functions [1]. A third possible outcome is that duplicated genes may undergo subfunctionalization, in which the preduplication state is partitioned between the two duplicate copies such that they exhibit complementary expression patterns or functions [10]. During subfunctionalization, duplicated genes accumulate degenerative mutations that result in each of the duplicates retaining only a subset of the ancestral expression pattern or function. This model, known as the duplication–degeneration–complementation (DDC) model, is particularly attractive because it is based on each gene acquiring degenerative mutations, which occur more frequently than beneficial mutations [10,11,14••]. Therefore, one would expect subfunctionalization to be the second most common outcome of a duplication event, occurring less frequently than nonfunctionalization but more than neofunctionalization. There is growing evidence for subfunctionalization playing an important role in the evolution of developmental modularity [15•,16,17]. For example, the compartmentalized expression of *Pax1* and *Pax9* in the developing vertebrae may have arisen by an initial acquisition of a single novel expression domain in the somites, and a later partitioning into two distinct regions of the sclerotome [17].

Genome duplications in vertebrates: the 2R hypothesis

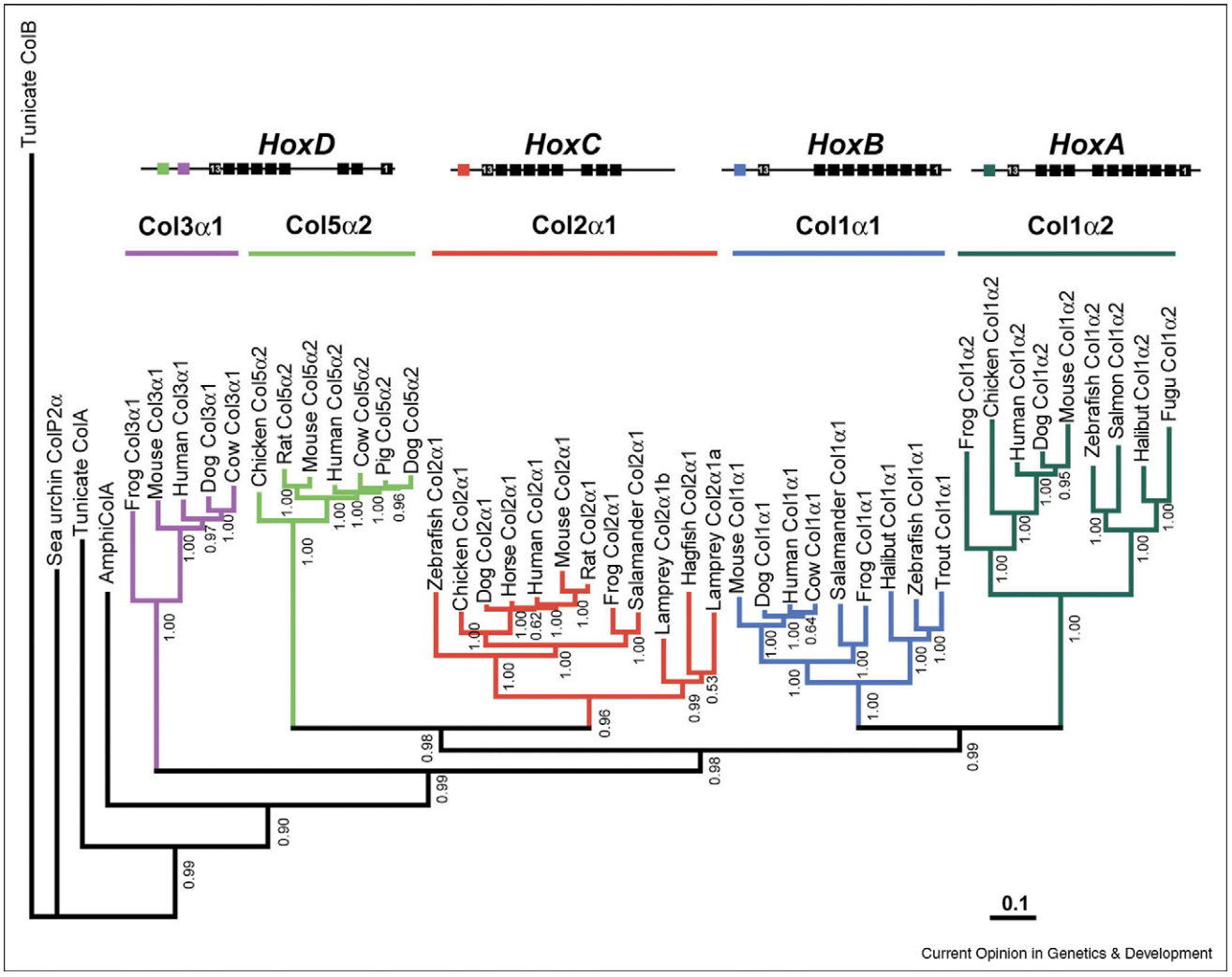
The precise number of genome duplications that occurred during vertebrate evolution has been difficult to infer from singular data sets, be they phylogenetic tree topologies, gene family sizes, numbers of duplicated genes, or paralogy maps [18,19]. In part, this is because few gene families have maintained their original post-duplication numbers, genomic structure, or evolutionary rates, and changes in any one of these parameters can generate a misleading evolutionary signature. The few exceptions, such as the clustered Hox and Major Histocompatibility Complex (MHC) genes, proved to be the most useful in early reconstructions of the ancestral chordate genome, though even these have complexities in their evolutionary histories that have made for challenging interpretations [20–22]. Nonetheless, the number and organization of Hox gene clusters in chordates provide some of the most compelling support for Ohno's 2R hypothesis, which proposed that two rounds of duplication occurred early in vertebrate evolution. A single Hox cluster is found in lancelets and echinoderms, whereas four Hox clusters exist in the chondrichthyans and sarcopterygians examined to date [2••,23,24•]. Ray-finned fishes (actinopterygians) further duplicated their Hox clusters during an additional round of genome duplication [13,25•,26]. Similarly strong evidence for two rounds of WGD (or at least large-bloc duplications) comes from analyses of genes in the MHC. The human MHC lies on the p arm of chromosome 6, and three paralogous regions have been found on human chromosomes 1, 9, and 19 [27]. Like the Hox complex, a single MHC-like complex exists both in lancelets and ascidians [28,29].

Evidence for the 2R hypothesis is not restricted to the large gene clusters described above. Chromosomal regions containing linked groups of paralogous genes (termed 'paralogons') are widespread in vertebrate genomes. Early estimates based on draft genome sequences placed the number of paralogons containing at least 3 genes at over 1000 [30,31]. In an elegant comparative study of the complete genome sequences of *Ciona*, *Fugu*, *Mus*, and *Homo*, Dehal and Boore identified in humans nearly 3000 paralogous gene pairs that duplicated before the divergence of fish and tetrapods [19]. Even more striking is their finding of 454 'tetra-paralogons' (paralogy regions that occur at exactly four places in the genome) distributed across all but one of the human autosomes and on the X chromosome [19]. Widespread tetra-paralogy is precisely the pattern that one would expect to find following two rounds of WGD, and although it should be noted that these are tetra-paralogous *regions* and not necessarily tetra-paralogous *genes* (a much rarer phenomenon because of the frequency of gene loss following duplication), such signatures in the genome provide strong support for the 2R hypothesis. The lancelet genome sequence recently confirmed that jawed vertebrates have quadruple conserved synteny relative to lancelets, providing the most conclusive data for two rounds of WGD occurring after the divergence of cephalochordates from the lineage leading to jawed vertebrates [2••].

Genome duplications in vertebrates: alternatives to 2R

Although the lancelet genome data are likely to resolve many of the questions surrounding the 2R debate, it should be noted that some investigators have suggested that more than two rounds of WGD predated origin of jawed vertebrates [32], whereas others have argued that paralogons (such as the Hox clusters) were assembled not only by whole genome duplication but also by small-scale gene duplications and chromosomal rearrangements [33]. Comparative studies of the Hox clusters have been particularly informative, and other gene families have provided additional insights into the number and timing of genome duplications in vertebrate evolution. With respect to skeletal evolution, an interesting case is the evolution of the collagen gene family. Collagens originated ~800 million years ago and diversified into two major groups, the fibril-forming and nonfibril-forming collagens. The fibrillar collagen family consists of A, B, and C clades [34,35]. Clade A contains collagen types I, II, III, and Vα2, which are linked to the Hox clusters and have been proposed to share their duplication history [32]. *Col1α2* is linked to the *HoxA* cluster, *Col1α1* is linked to *HoxB*, *Col2α1* is linked to *HoxC*, and both *Col5α2* and *Col3α1* are linked to the *HoxD* cluster (Figure 2). On the basis of a phylogenetic study of the Hox-linked collagens, Bailey *et al.* suggested that the strongest support existed for a (D(A(B,C))) relationship among the Hox clusters, and they proposed that this was achieved by three rounds of cluster duplication ($1 \rightarrow 2 \rightarrow 3 \rightarrow 4$). A more recent analysis of clade A collagens in jawed and jawless vertebrates showed a topology of (Col3α1(Col5α2,Col2α1)(Col1α1,Col1α2)) (Figure 2 and Ref. [36•]). If these genes coduplicated

Figure 2



Phylogenetic relationships among chordate clade A fibrillar collagen protein sequences and their linkage to the HoxA–D clusters. The four mouse Hox clusters are shown above each collagen clade, and the colored boxes represent each of the linked collagen genes. Tree is after [36•] and the Hox-Collagen cluster diagrams are after [32].

with the Hox clusters, this would support a Hox topology of (D,C)(B,A).

Two additional findings bear on the question of the number and timing of the duplication events. Firstly, both lampreys and hagfish have true orthologs of *Col2α1* [36•,37••], which encodes the most abundant protein in vertebrate cartilage. Given that *Col2α1* is linked to the *HoxC* cluster in jawed vertebrates, and that *Col2α1* was present in the common ancestor of cyclostomes and gnathostomes, an obvious question is whether the *HoxC* cluster also predated the divergence of these lineages. Lampreys are estimated to have three or four Hox clusters, and hagfishes may have as many as seven Hox clusters [38–41]. Three independent analyses of the lamprey Hox clusters have confirmed that additional duplications occurred in the lamprey lineage after its divergence from the gnathostome lineage [38–40], and two copies of Col2α1 were also identified in lampreys [37••]. Force *et al.* suggest that at least one of these duplication events predated the lamprey–gnathostome divergence [38]. In support of their hypothesis is their finding that several lamprey Hox sequences group with gnathostome C and D clusters, which is consistent with the common ancestor of lampreys and gnathostomes possessing an A/B cluster and a C/D cluster. Whether the four gnathostome HoxA–D clusters were present in the common ancestor of jawed and jawless vertebrates has been difficult to resolve (the extreme divergence of lamprey sequences clouds their orthology to gnathostome genes), though answers may come with the completion of the lamprey genome sequence.

The second issue stems from the work on the ParaHox genes of jawless fishes, which has led to the suggestion that the first round of duplication occurred in the gnathostome lineage after the divergence of cyclostomes [42•]. Phylogenetically, the hagfish ParaHox genes lie outside of the gnathostome Cdx and Gsx clades, which raises the possibility that the hagfish lineage diverged before the genome duplications that gave rise to the multiple Hox and ParaHox clusters of gnathostomes [42•]. In their analysis of lamprey Hox DNA sequences, Fried *et al.* supported the idea that lamprey and gnathostome Hox genes arose independently from a single cluster in their common ancestor [41], though this contrasts with studies using protein datasets [38,39]. If hagfishes do indeed have seven Hox clusters and one ParaHox cluster, then this would suggest either that large-scale loss of the ParaHox genes followed the genome duplications, or the less probable possibility that hagfish Hox clusters arose by a mechanism other than WGD. Whatever the status of the Hox clusters before the divergence of jawed and jawless vertebrates, there is little doubt lampreys, and possibly hagfish, underwent successive rounds of independent duplication. Support for this comes not only from the Hox clusters, but also from the lack of strict orthology among many other gene families. Escriva *et al.* analyzed 33 gene families in lampreys and hagfishes and argued that the different topologies reflect a complex history of genome duplications, smaller-scale gene duplications and a high rate of secondary gene loss. They concluded that 'two major periods of gene duplications' occurred before the separation of the lamprey and gnathostome lineages, with additional duplications occurring after their divergence [12].

Genome duplication and the origin of the vertebrate skeleton

The vertebrate skeleton has been proposed to be one of the innovations that resulted from the amplification of gene number by WGD [43]. Hox genes and fibrillar collagens play essential roles in patterning and differentiation of vertebrate connective tissues, including cartilage and bone. Duplication of *Hox* clusters and the linked clade A fibrillar collagen genes may have played a key role in this process, in that expansion of these two families could have provided mechanisms both for the diversification of connective tissue cell types and for differential regulation of their patterning and growth. This raises two central questions; what is the evolutionary origin of the vertebrate chondrocyte, and what is the relationship between increased gene number and its evolutionary divergence from the ancestral cell type? It has been suggested that the notochord may represent a primitive form of cartilage, based on their many shared structural, cellular, and molecular properties, and that vertebrate chondrocytes may have evolved from notochordal cells [44,45]. In gnathostomes, the notochord and/or notochordal sheath expresses most of the vertebrate clade A fibrillar collagen genes [36•]. Lancelets and ascidians have been reported to possess a single clade A fibrillar collagen gene that is expressed in the notochord, supporting the idea that an ancestral *ColA* gene was expressed in the notochord of stem-group chordates [36•,37••,46•]. Thus, the duplication and divergence of the clade A collagen genes in stem-group vertebrates may have facilitated the evolutionary origin of chondrocytes from notochordal cells [36•]. Interestingly, notochordal cells of lancelets have been found to express several other orthologs of vertebrate chondrogenic genes, including SoxE, SoxD, Twist, and Ets [47], suggesting that many of the genetic components required for cartilage differentiation were expressed in the chordate notochord before the origin of vertebrates. Understanding how the genetic cassettes for skeletogenesis were assembled during evolution will require further work on the nature of the regulatory interactions among these genes in an unduplicated genome, like that of the lancelet. Whether the earliest cartilaginous tissues were derived from mesoderm, neural crest, or were acellular (possibly secreted from endoderm) remains unclear [36•,37••,47,48••]. Reconstruction of skeletal development in the vertebrate common ancestor will be most accurate if it combines inferences from the positions of skeletal elements in fossil

vertebrates with data on the developmental origin of homologous structures in extant taxa.

Conclusions

As the number of complete genome sequences from vertebrates and basal chordates has increased in recent years, so too has the evidence in support of Ohno's proposal that WGD played a major role in the evolution of vertebrates. Comparative developmental studies show that neofunctionalization and subfunctionalization are indeed mechanisms of evolutionary change. Of the vertebrate organ systems, the skeleton offers the unique opportunity to integrate comparative development and genomics with paleontological data, and therefore has the potential to address the *who*, *what*, *when*, *where*, and *how* questions about skeletal evolution. Major questions still remain to be resolved, such as the precise number of genome duplications, the timing of these events relative to the origins of the major chordate clades, and the impact of genome duplication on the evolution of developmental modularity and complexity.

Acknowledgements

We thank Charlie Baer and Ashley Seifert for providing critical comments on the manuscript.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Ohno S: **Evolution by Gene Duplication**. New York: Springer; 1970.

2. •• Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK *et al.*: **The amphioxus genome and the evolution of the chordate karyotype**. *Nature* 2008, **453**:1064-1071.
This paper reports the draft genome sequence of a lancelet (or amphioxus), and confirms its status as the most basal living chordate. Comparison with vertebrate genomes supports two rounds of whole genome duplication between the origin of chordates and jawed vertebrates.

3. Storchova Z, Pellman D: **From polyploidy to aneuploidy, genome instability and cancer**. *Nat Rev Mol Cell Biol* 2004, **5**:45-54.

4. Gallardo MH, Bickham JW, Honeycutt RL, Ojeda RA, Kohler N: **Discovery of tetraploidy in a mammal**. *Nature* 1999, **401**:341.

5. Furlong RF, Holland PW: **Were vertebrates octoploid?** *Philos Trans R Soc Lond B: Biol Sci* 2002, **357**:531-544.

6. Gibson TJ, Spring J: **Evidence in favour of ancient octaploidy in the vertebrate genome**. *Biochem Soc Trans* 2000, **28**:259-264.

7. Wolfe KH: **Yesterday's polyploids and the mystery of diploidization**. *Nat Rev Genet* 2001, **2**:333-341.

8. Eakin GS, Behringer RR: **Tetraploid development in the mouse**. *Dev Dyn* 2003, **228**:751-766.

9. •• Otto SP: **The evolutionary consequences of polyploidy**. *Cell* 2007, **131**:452-462.
The author provides an excellent review of the mechanisms of polyploidy and consequences of genome duplication through a discussion of short-term and long-term effects on the biology of polyploid organisms.

10. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations**. *Genetics* 1999, **151**:1531-1545.

11. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes**. *Science* 2000, **290**:1151-1155.

12. Escriva H, Manzon L, Youson J, Laudet V: **Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution**. *Mol Biol Evol* 2002, **19**:1440-1450.

13. Panopoulou G, Poustka AJ: **Timing and mechanism of ancient vertebrate genome duplications — the adventure of a hypothesis**. *Trends Genet* 2005, **21**:559-567.

14. •• Lynch M: **The Origins of Genome Architecture**. Sunderland, MA: Sinauer Assocs., Inc.; 2007.
Lynch's book is a contemporary and insightful treatment of evolution by genome duplication. This is the definitive bridge between Ohno's classic book [1] and the latest progress in genome evolution.

15. • Kleinjan DA, Bancewicz RM, Gautier P, Dahm R, Schonthaler HB, Damante G, Seawright A, Hever AM, Yeyati PL, van Heyningen V *et al.*: **Subfunctionalization of duplicated zebrafish pax6 genes by *cis*-regulatory divergence**. *PLoS Genet* 2008, **4**:e29.
By mapping the zebrafish eye mutant *sunrise* and conducting transgenesis experiments in mice and zebrafish, the authors show that *cis*-regulatory evolution drove the divergence of pax6a and pax6b.

16. Yan YL, Willoughby J, Liu D, Crump JG, Wilson C, Miller CT, Singer A, Kimmel C, Westerfield M, Postlethwait JH: **A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development**. *Development* 2005, **132**:1069-1083.

17. Mise T, Iijima M, Inohaya K, Kudo A, Wada H: **Function of Pax1 and Pax9 in the sclerotome of medaka fish**. *Genesis* 2008, **46**:185-192.

18. Horton AC, Mahadevan NR, Ruvinsky I, Gibson-Brown JJ: **Phylogenetic analyses alone are insufficient to determine whether genome duplication(s) occurred during early vertebrate evolution**. *J Exp Zoolog B: Mol Dev Evol* 2003, **299**:41-53.

19. Dehal P, Boore JL: **Two rounds of whole genome duplication in the ancestral vertebrate**. *PLoS Biol* 2005, **3**:e314.

20. Hughes AL: **Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history**. *J Mol Evol* 1999, **48**:565-576.

21. Hughes AL, da Silva J, Friedman R: **Ancient genome duplications did not structure the human Hox-bearing chromosomes**. *Genome Res* 2001, **11**:771-780.

22. Martin AP: **Increasing genomic complexity by gene duplication and the origin of vertebrates**. *Am Nat* 1999, **154**:111-128.

23. Amemiya CT, Prohaska SJ, Hill-Force A, Cook A, Wasserscheid J, Ferrier DE, Pascual-Anaya J, Garcia-Fernandez J, Dewar K, Stadler PF: **The amphioxus Hox cluster: characterization, comparative genomics, and evolution**. *J Exp Zoolog B: Mol Dev Evol* 2008, **310**:465-477.

24. • Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC *et al.*: **Survey sequencing and comparative analysis of the elephant shark (*Callorhinchus milii*) genome**. *PLoS Biol* 2007, **5**:e101.
Using a shotgun-based survey sequencing approach, the authors report that the genome of a shark shares more synteny and sequence similarity with humans than does the zebrafish. Although sharks are more distantly related to humans, unlike teleosts they have not undergone the additional round genome duplication.

25. • Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP: **The ''fish-specific'' Hox cluster duplication is coincident with the origin of teleosts**. *Mol Biol Evol* 2006, **23**:121-136.
A paper surveyed the teleost Hox clusters in bony fishes, and provides us the evidences that all the genome neoteleosts underwent extra round of WGD.

26. Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y: **Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates**. *Proc Natl Acad Sci U S A* 2004, **101**:1638-1643.

27. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, Lush MJ, Povey S, Talbot CC Jr, Wright MW *et al.*: **Gene map of the extended human MHC**. *Nat Rev Genet* 2004, **5**:889-899.

28. Castro LF, Furlong RF, Holland PW: **An antecedent of the MHC-linked genomic region in amphioxus**. *Immunogenetics* 2004, **55**:782-784.

29. Azumi K, De Santis R, De Tomaso A, Rigoutsos I, Yoshizaki F, Pinto MR, Marino R, Shida K, Ikeda M, Arai M *et al.*: **Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: ''waiting for Godot''**. *Immunogenetics* 2003, **55**:570-581.

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

31. 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.

32. Bailey WJ, Kim J, Wagner GP, Ruddle FH: **Phylogenetic reconstruction of vertebrate Hox cluster duplications**. *Mol Biol Evol* 1997, **14**:843-853.

33. Abbasi AA, Grzeschik KH: **An insight into the phylogenetic history of HOX linked gene families in vertebrates**. *BMC Evol Biol* 2007, **7**:239.

34. Runnegar B: **Collagen gene construction and evolution**. *J Mol Evol* 1985, **22**:141-149.

35. Morvan-Dubois G, Le Guellec D, Garrone R, Zylberberg L, Bonnaud L: **Phylogenetic analysis of vertebrate fibrillar collagen locates the position of zebrafish alpha3(I) and suggests an evolutionary link between collagen alpha chains and hox clusters**. *J Mol Evol* 2003, **57**:501-514.

36. • Zhang G, Cohn MJ: **Hagfish and lancelet fibrillar collagens reveal that type II collagen-based cartilage evolved in stem vertebrates**. *Proc Natl Acad Sci U S A* 2006, **103**:16829-16833.
Taking a comparative developmental approach, this paper and Ref. [37••] report that jawless fishes have Col2α1 genes and Collagen II-based cartilage, pushing the origin of the collagenous skeleton back to the common ancestor of all living vertebrates. Additionally, lancelets have a single ancestral fibrillar collagen gene expressed in the notochord. This study shows that expansion of the fibrillar collagen gene family coincided with origin of the vertebrate skeleton.

37. •• Zhang G, Miyamoto MM, Cohn MJ: **Lamprey type II collagen and Sox9 reveal an ancient origin of the vertebrate collagenous skeleton**. *Proc Natl Acad Sci U S A* 2006, **103**:3180-3185.
See annotation to Ref. [36•].

38. Force A, Amores A, Postlethwait JH: **Hox cluster organization in the jawless vertebrate *Petromyzon marinus***. *J Exp Zool* 2002, **294**:30-46.

39. Irvine SQ, Carr JL, Bailey WJ, Kawasaki K, Shimizu N, Amemiya CT, Ruddle FH: **Genomic analysis of Hox clusters in the sea lamprey *Petromyzon marinus***. *J Exp Zool* 2002, **294**:47-62.

40. Stadler PF, Fried C, Prohaska SJ, Bailey WJ, Misof BY, Ruddle FH, Wagner GP: **Evidence for independent Hox gene duplications in the hagfish lineage: a PCR-based gene inventory of *Eptatretus stoutii***. *Mol Phylogenet Evol* 2004, **32**:686-694.

41. Fried C, Prohaska SJ, Stadler PF: **Independent Hox-cluster duplications in lampreys**. *J Exp Zoolog B: Mol Dev Evol* 2003, **299**:18-25.

42. • Furlong RF, Younger R, Kasahara M, Reinhardt R, Thorndyke M, Holland PW: **A degenerate ParaHox gene cluster in a degenerate vertebrate**. *Mol Biol Evol* 2007, **24**:2681-2686.
To test the 2R hypothesis, the authors characterize the ParaHox cluster in hagfishes, and suggest that the first genome duplication in vertebrates postdated the divergence of hagfishes from the lineage leading to jawed vertebrates.

43. Shimeld SM, Holland PW: **Vertebrate innovations**. *Proc Natl Acad Sci U S A* 2000, **97**:4449-4452.

44. Cole AG, Hall BK: **Cartilage is a metazoan tissue; integrating data from nonvertebrate sources**. *Acta Zool (Stockholm)* 2004, **85**:69-80.

45. Stemple DL: **Structure and function of the notochord: an essential organ for chordate development**. *Development* 2005, **132**:2503-2512.

46. • Wada H, Okuyama M, Satoh N, Zhang S: **Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny**. *Evol Dev* 2006, **8**:370-377.
This paper showed the two chordates, asidians and lancelets, fibrillar collagen gene expression patterns and discussed the origin of veretbrate skeletons.

47. Meulemans D, Bronner-Fraser M: **Insights from amphioxus into the evolution of vertebrate cartilage**. *PLoS ONE* 2007, **2**:e787.

48. •• Rychel AL, Swalla BJ: **Development and evolution of chordate cartilage**. *J Exp Zoolog B: Mol Dev Evol* 2007, **308**:325-335.
The authors show that SoxE and fibrillar collagen genes are expressed in pharyngeal endoderm of lancelets and hemichordates, adjacent to the acellular cartilage. They raise the possibility that the secretion of acellular matrix by pharyngeal endoderm was the primitive mode of chondrogenesis for deuterostomes, and the gene network was later co-opted by neural crest cells.