OXFORD

Briefings in Functional Genomics, 2015, 1–7

doi: 10.1093/bfgp/elv032
Review paper

Toward understanding the evolution of vertebrate gene regulatory networks: comparative genomics and epigenomic approaches

Juan R. Martinez-Morales

Corresponding author. Juan R. Martinez-Morales, Centro Andaluz de Biología del Desarrollo (CSIC/UPO/JA), 41013 Sevilla, Spain. Tel.: +34 954977827; E-mail: jrmarmor@upo.es

Abstract

Vertebrates, as most animal phyla, originated >500 million years ago during the Cambrian explosion, and progressively radiated into the extant classes. Inferring the evolutionary history of the group requires understanding the architecture of the developmental programs that constrain the vertebrate anatomy. Here, I review recent comparative genomic and epigenomic studies, based on ChIP-seq and chromatin accessibility, which focus on the identification of functionally equivalent *cis*-regulatory modules among species. This pioneer work, primarily centered in the mammalian lineage, has set the groundwork for further studies in representative vertebrate and chordate species. Mapping of active regulatory regions across lineages will shed new light on the evolutionary forces stabilizing ancestral developmental programs, as well as allowing their variation to sustain morphological adaptations on the inherited vertebrate body plan.

Key words: vertebrate evolution; phylotypic period; conserved non-coding elements (CNEs); cis-regulatory modules (CRMs); comparative epigenomics

Introduction

The vertebrate body plan emerged from the basic chordate blueprint, with an axial notochord, a dorsal neural tube flanked by segmented trunk muscles, pharyngeal gill slits and a ventral heart. The three main groups within the phylum *Chordates*, cephalochordates, tunicates and vertebrates (here, referred as vertebrates for simplicity), share this archetypal design [1]. Assembled over this chordate architecture, a number of vertebrate-specific traits (i.e. synapomorphies) evolved during the Cambrian radiation. The most relevant of all these innovations is the elaboration of a new head over the ancestral trunk [2–4] (Figure 1A). It has been postulated that the appearance of novel cell populations such as the neural crest and ectodermal placodes (initially under relaxed evolutionary pressure) could have supported the evolution of tissue innovations [5, 6]. Emerging predatory jaws and elaborated sensory organs, together with the acquisition of both an endoskeleton and an increasingly complex brain, may have permitted the transition from a filter-feeding to an active predation lifestyle [2, 7]. The 2R hypothesis proposes the occurrence of two rounds of whole genome duplication (WGD) as an event closely related to the origin and early evolution of vertebrates [8–10]. It has been repeatedly suggested that these duplications, followed by sub-functionalization of duplicated gene copies, act as a permissive (and even instructive) factor for the emergence of the evolutionary innovations observed in vertebrates [8, 11, 12]. In line with this notion, the extra WGD observed in actinopterygians (3R) has also been postulated as a driving mechanism for the fast adaptive radiation observed in teleosts [13–15]. Opposing this view, some authors have argued that the anatomical analysis of fossil vertebrates does not support a link between WGDs and the emergence of evolutionary novelties [16].

Riding a wave of evolutionary plasticity, vertebrates diverged into a fascinating range of morphological adaptations. This becomes evident either when comparing the fossil skeletons of extinct animals or simply by contemplating the variety of modern species (>60 000 vertebrate species have been described). This morphological diversity, apparent in adult animals,

Juan R. Martinez-Morales is group leader at the Centro Andaluz de Biología del Desarrollo (CSIC/UPO, Seville). His laboratory focuses on understanding the mechanisms involved in the morphogenesis of the vertebrate eye and the evolutionary emergence of vertebrate innovations.

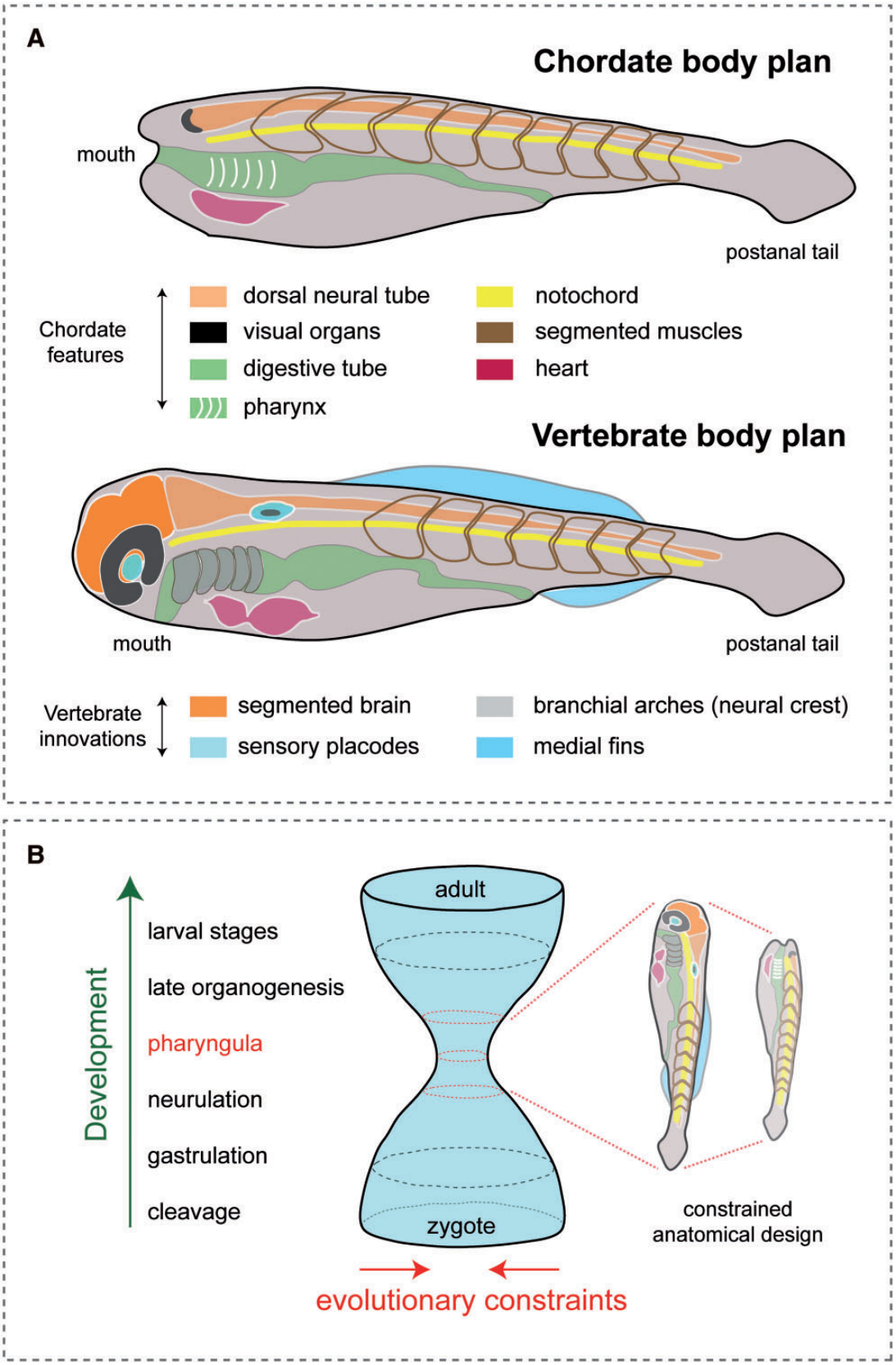

Figure 1. The vertebrate body plan and the hourglass model. (**A**) Schematic representation of the characteristic anatomical features of the chordate and vertebrate body plans. Common chordate/vertebrate structures as well as vertebrate innovations are represented. (**B**) The hourglass model, showing evolutionary constraints at the phylotypic window is depicted.

narrows considerably when embryonic forms are compared. On the basis of this observation, the 19th century embryologist Ernst Von Baer formulated the 'Laws of Development', which can be summarized in a fundamental idea: during embryogenesis, the organisms proceed acquiring first the most general traits of the clade they belong and then the specific features of the individual species. This concept remains valid, being central to our current understanding of the relationship between ontogeny and phylogeny [17, 18]. Possibly the main refinement to Von Baer's laws of embryology is the modern view that the point of maximum homology does not occur early during segmentation and gastrulation stages, but later at mid-development at the so-called phylotypic period, which in vertebrates coincides with the pharyngula stages [19, 20] (Figure 1B).

During the last decades of the 20th century, developmental genetics emerged (i.e. from the convergence of classical embryology, genetics and molecular biology) as a field focused on understanding the genetic bases of organ formation. Most of the key regulatory genes were identified in vertebrate and invertebrate models and, surprisingly at the time, many turned to be functionally conserved across metazoans [21]. It soon became obvious that these developmental genes do not act in isolation, but rather as central nodes of ancestral gene regulatory networks (GRNs) responsible for the establishment and maintenance of tissue identity through a hierarchical series of developmental decisions. According to this view, conserved body plans would depend on the conservation of core GRN or 'kernels', and conversely, the emergence of new animal tissues and organs would entail the re-elaboration of ancestral GRNs [22, 23].

A genetic interpretation of phenotypic convergence during mid-embryogenesis is the hourglass model, which postulates that the preservation of the body plan depends on constrained molecular mechanisms active at the phylotypic stage [24]. Although a number of molecular analyses could not confirm the hourglass model [25, 26], more recent advances in transcriptomics have allowed the identification of an evolutionary restrained molecular signature at the phylotypic period. Studies based on either the comparison of transcriptome profiles during the ontogeny of related species [27–29], or the examination of the transcriptome age index (i.e. phylotranscriptomics) through embryogenesis in a single species [30, 31], have succeeded in the identification of an hourglass pattern. The nature of the evolutionary forces maintaining phylotypic restraints over time has provided fertile ground for hypothesis [21, 24, 32]. However, the exact mechanistic description of these forces, which seem to be active at the phylotypic period in extant species [31], still remains as a challenging task. Moreover, understanding vertebrates' evolutionary origin and further radiation requires gaining insight into many other fundamental questions such as: how novel developmental programs were evolutionary assembled to sustain vertebrate innovations? Do these programs depend on the re-elaboration of ancestral GRNs? Among the GRNs conserved in vertebrates, which are more resilient to evolutionary change and why? Is the conservation of these networks the ultimate cause of the observed phylotypic constraint? Can different regulatory circuits converge into similar anatomical designs? Whereas gene centric analyses provided by classical developmental genetic approaches were essential to understand the basic programs involved in tissue specification (e.g. [33, 34]), genome-wide approaches will be fundamental to answer many of the questions formulated above. In the following brief, we review recent functional genomic and comparative epigenomic studies that shed light on the identification of *cis*-regulatory modules (CRMs) across vertebrates.

Comparative analysis of conserved non-coding elements during vertebrate development and evolution

CRMs are commonly described as genomic regions (generally non-coding) containing clusters of transcription factor binding sites (TFBSs), which control a subcomponent of the overall expression pattern of a given gene [35]. The ability of a CRM (although CRMs include enhancers, silencers, promoters, locus-control regions and insulators, in this review, we have focused mainly on those modules acting as tissue-specific enhancers) to behave as an enhancer (i.e. driving expression in a spatio-temporal restricted manner) can be experimentally assessed in transgenesis experiments. The individual analysis of genomic regions in enhancer assays is however an arduous task, and several approaches have been undertaken for genome-wide CRM prediction (reviewed by [36, 37]). The sequencing of an increasing number of vertebrate species [38–45] opened the possibility of multi-species genomic alignments, and thus the indirect prediction of CRM, either through the identification of TFBSs clusters or simply by DNA alignment between related species (known as phylogenetic footprinting).

Comparative genomics has been extensively used to reveal conserved non-coding elements (CNEs) in several species including vertebrates, for which a collection of approximately 3000 highly conserved non-coding elements (hCNEs) has been identified [46, 47]. Similarly, an independent collection of conserved regulatory elements is associated to developmental genes in the invertebrate group [48]. Vertebrate and invertebrate sets of CNEs are almost non-overlapping, with only a few anecdotic examples of trans-phyletic conservation of regulatory modules [49–51]. Interestingly, most of the vertebrate CNEs are in the vicinity of genes with important morphogenetic roles during development [47, 52]. In fact, several functional studies indicate that a large proportion of the vertebrate's CNEs contain enhancer or silencer elements that regulate the expression of developmental genes in specific domains of the embryo [52–55]. These initial observations suggested that CNEs could be used as a proxy to identify CRMs; however, several lines of evidence have shown the limitations of this assumption. First, not all CNEs seem to behave as tissue-specific drivers in transgenesis assays [54]. Moreover, the operational logic of a CNE may be species specific, depending both on a particular combination of TFBSs, and on the expression and activity of the transcriptional regulators that bind this element. Thus, it is not infrequent that conserved sequences behave differently when injected in different species [56]. Finally, a number of reports have described enhancers that despite being associated to the same gene and driving expression in a similar manner do not show apparent sequence conservation [57–59]. This can be partially explained because of the limitations of the computational methods used to detect sequence conservation in 'covert' elements [60]. Through pairwise comparisons among distantly related species, Taher *et al.* [60] were able to identify 1500 pairs of human and zebrafish elements showing no conservation by direct alignment, but that could be assigned to common CNEs when the frog genome was intercalated.

Functionally conserved covert elements, with similar regulatory logic embedded into a divergent sequence background, might represent an important fraction of the CRMs in

vertebrates. This could be particularly the case in teleosts' genomes, in which CNEs seem to evolve faster than in other groups [61, 62]. The pervasive existence of covert regulatory elements provides an explanation to the following paradox: despite the obvious conservation of a basic anatomical design in vertebrates, and to a lesser extent also in chordates, only a few dozen mammalian CNEs can be traced back in jawless fishes and even a smaller number in cephalochordates [49, 51, 63, 64]. This apparent lack of homology is even more puzzling, taking into account that half of the most conserved CNEs behave as active enhancers precisely during mid-embryogenesis [54, 65], the period of maximum morphological constraint. The assumption of completely different GRNs that independently converge on the same body plan in vertebrates and lower chordates (i.e. tunicates and cephalochordates) seems unlikely. The alternative hypothesis, the active maintenance of core GRNs to sustain the chordate/vertebrate architecture, entails the existence of functionally conserved CRMs without apparent sequence conservation among chordate subphyla. This lack of overall sequence conservation in otherwise functionally equivalent CRMs argues for a hidden homology in regulatory elements derived from an ancestral sequence (i.e. cover elements containing similar TFBS clusters). Alternatively, GRNs maintenance through evolution can also be explained by the *de novo* re-elaboration of modules, not related at the phylogenetic level, but sharing the same TFBSs regulatory logic. Gaining insight into the evolutionary history of vertebrate GRNs requires then direct methodological approaches (i.e. not based in sequence conservation) to identify active CRMs. The main conclusions of the initial studies using comparative epigenomics as a research tool will be discussed in the following section.

Comparative epigenomics in mammals and beyond

The development during the past decade of massive parallel sequencing platforms represented a gigantic technological leap that has transformed epigenomic studies, among other research fields [66]. Based on these new technologies, a number of direct methods have been established to identify CRMs at a genome scale, including mainly the ChIP-seq and open chromatin analyses. These approaches constitute core analytical tools in ENCODE and Roadmap Epigenomics, both reference projects for the systematic mapping of CRMs in the human genome [67, 68].

Among vertebrates, comparative global mapping of TF-binding events by ChIP-seq has been explored particularly in mammals. These studies have focused on individual tissues and associated tissue-specific factors such as OCT4 and NANOG in embryonic stem cells [69], CEBPA and HNF4A in liver [70] or PPARG in adipose tissue [71]. The overall conclusion of these works is that a large fraction of the TF-binding events is species specific, thus implying a rapid regulatory turnover of compensatory modifications to maintain similar gene transcriptional output [72]. Epigenomics analyses in *Drosophila* species reached independently similar conclusions on the specificity of the binding events [73]. In fact, TFBSs turnover is also obvious in closely related mouse species [74], and even among individuals for whom TF-binding loses are frequently associated to genomic polymorphisms [75]. This fast evolutionary dynamics of TF-binding precludes, to a large extent, comparative epigenomics studies including far-related vertebrates. As an illustrative example of this limitation, only 6–8% of CEBPA binding is conserved between human and opossum, and the conservation is further reduced to as little as 2% when human and chicken are examined [70].

Histone modifications in general, and H3K27ac and H3K4me3 in particular, provide a molecular signature of the activity state of CRMs [76–78], and constitute an apparently universal regulatory code in eumetazoans [79]. These epigenetic marks have been used to reveal active enhancers across mammalian species, focusing on specific cell types and tissues such as embryonic stem cells [80], limbs [81] and hepatic tissue [82]. The data obtained through the analysis of histone modifications' binding provide an additional layer of information to that of genomic conservation and TF binding, and contribute to a comprehensive interpretation of the genome regulatory logic [80]. Although more conserved among species than TFBSs [70], a substantial fraction of the CRMs revealed by epigenetic marks are lineage specific. Thus, comparative studies in mammals have described a rapid evolution of enhancers, often by exaptation of ancestral DNA, and a higher conservation of active promoters [80–83]. This type of information can now be used as an important tool, both to investigate how morphological and physiological adaptations evolved within a particular lineage [81, 82], and to understand the nature of the evolutionary constraints that maintain the vertebrate body plan. Regarding this last aspect, the dynamics of epigenetic marks during embryogenesis has been explored either for the whole zebrafish embryo [84] or for specific mouse tissues [83]. Interestingly, both studies show that enhancers active during mid-embryogenesis show a higher degree of sequence conservation. The comparative analysis of epigenetic marks at the phylotypic window in two far-related teleosts, zebrafish and medaka, has allowed the identification of a set of conserved and active vertebrate enhancers (≈700), which regulate the expression of TF with fundamental roles in tissue specification [85]. It is likely that these TF represent 'hub' genes within the GRNs operating during the phylotypic period. To which extent, other phylotypic enhancers showing no apparent sequence conservation may contribute to the regulation of the core GRNs is currently unclear.

In addition to ChIP-seq approaches, open chromatin detection methods such as DNAseI hypersensitivity sites (DHSs) [86–88], FAIRE [89] and more recently ATAC-seq [90, 91] have been used to investigate *cis*-regulatory landscapes in different tissue types. Recent DHSs studies between human and mouse tissues have allowed a comparative analysis of TF occupancy at nucleotide resolution [92, 93]. A first conclusion from these works is that, despite the existence of a common TF recognition lexicon in mammals, only 22% of the individual TF footprints are conserved. This low conservation argues again for an extensive evolutionary turnover at the TFBSs level and is in line with the conclusions from previous ChIP-seq analyses [69, 70, 82]. Interestingly, this conservation percentage increases to 44% when TF to TF connections are explored, and reaches 95% when the architectures of specific GRNs are examined [92].

Conclusions and future directions

Comparative epigenomics is emerging as a valuable tool in evolutionary biology. Pioneer studies examining mammalian species have shown the potentiality of these novel analytical methods to investigate the emergence of lineage-specific traits, as well as to identify ancient regulatory blocks responsible for the common features of the entire clade. Comprehensive maps of enhancers, promoters and open chromatin should now be obtained for representative organisms of the vertebrate and chordate evolutionary tree. The direct analysis of their CRMs

dynamics during development should shed new light on the architecture of ancestral GRNs conserved both in vertebrates and chordates. Moreover, this information will be instrumental to infer how developmental programs evolved to sustain morphological adaptations [94], and how novel cell types and tissues emerged to eventually transform the vertebrate body plan.

The development of new bioinformatic tools to uncover hidden sequence homology, in conjunction with ChIP-seq and chromatin accessibility approaches, should provide enough analytical depth to identify relevant CRMs at each transition during vertebrates evolutionary history. To the classical transgenesis approaches to test elements' functionality *in vivo*, now we can add recent advances in genome editing technologies provided by the CRISPR–Cas9 system [95, 96]. These new methods open the possibility of removing and engineering CRMs in a variety of model organisms, and hence will constitute a powerful tool to functionally validate predictions derived from epigenomic studies.

Glossary terms

Assay for transposase-accessible chromatin followed by deep sequencing (ATAC-seq). This genome-wide technology allows investigating chromatin accessibility. The method is based on the Tn5-mediated *in vitro* transposition of sequencing adapters into open chromatin regions.

Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq). The method is used for genome-wide isolation and sequencing of DNA fragments associated to a given protein after covalent cross-linking.

***Cis*-regulatory modules (CRMs).** They are defined as DNA elements (usually non-coding) able to influence the expression of one or more genes. CRMs contain specific arrangements of transcription factor binding sites (TFBSs) determining their role as silencers, enhancers or insulators. CRMs can be either evolutionary conserved or not.

Conserved non-coding elements (CNEs). They are defined as non-coding DNA fragments identified by sequence homology between species. CNEs may act as CRMs or not.

Covert regulatory elements. CRMs without apparent sequence conservation between species, according to conventional alignment algorithms, but which show functional equivalence. Cover elements often containing similar TFBSs clusters.

DNA exaptation. Evolutionary adaptation of DNA sequences to novel regulatory functions, which are different from their ancestral role. In this review, refers specifically to the evolution of CRMs.

DNAseI hypersensitivity sites (DHSs). Regions of open chromatin accessible for DNaseI digestion. Genome-wide sequencing of these sites (DNase-seq) provides a comprehensive mapping of chromatin accessibility.

Formaldehyde-assisted isolation of regulatory elements (FAIRE). This genome-wide technology allows investigating chromatin accessibility. It is based on the differential efficiency of formaldehyde cross-linking between nucleosome-bound DNA and open chromatin regions.

Gene regulatory network (GRN). Interrelated collection of transcription factors and CRMs assembled into an operational circuit that determines the expression of all its components (nodes).

Phylogenetic footprinting. Comparative genomics approach used for the identification of CNEs through multi-species DNA alignment.

Phylotypic period. The developmental window in which all the embryonic forms within a phylum converge into common anatomical features.

Transcription factor binding sites (TFBSs). Short and specific DNA sequences that act as binding/recognition motifs for transcription factors. They constitute the minimal functional units within *cis*-regulatory modules.

Key points

- Understanding the origin and evolution of the vertebrate body plan requires the analysis of CRMs in different species and developmental stages.
- Maintenance of the vertebrate/chordate anatomical features is not accompanied by the conservation of CRMs at the sequence level.
- Recent advances in comparative epigenomics provide a direct method to identify active CRMs in vertebrate species.
- These novel analytical tools can now be applied to investigate enhacers' evolution in the vertebrate clade.

Funding

This work was supported by grants BFU2011-22916, BFU2014-53765-P and P11-CVI-7256 (to J.R.M.M.).

References

1. Wada H, Satoh N. Patterning the protochordate neural tube. *Curr Opin Neurobiol* 2001;**11**:16–21.
2. Gans C, Northcutt RG. Neural crest and the origin of vertebrates: a new head. *Science* 1983;**220**:268–72.
3. Shimeld SM, Holland PW. Vertebrate innovations. *Proc Natl Acad Sci USA* 2000;**97**:4449–52.
4. Holland ND, Chen J. Origin and early evolution of the vertebrates: new insights from advances in molecular biology, anatomy, and palaeontology. *Bioessays* 2001;**23**:142–51.
5. Arendt D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet* 2008;**9**:868–82.
6. Martinez-Morales JR, Henrich T, Ramialison M, *et al*. New genes in the evolution of the neural crest differentiation program. *Genome Biol* 2007;**8**:R36.
7. Manzanares M, Nieto MA. A celebration of the new head and an evaluation of the new mouth. *Neuron* 2003;**37**:895–8.
8. Ohno S. *Evolution by Gene Duplication*. New York: Springer-Verlag, 1970.
9. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 2005;**3**:e314.
10. Kasahara M. The 2R hypothesis: an update. *Curr Opin Immunol* 2007;**19**:547–52.

11. Holland PW, Garcia-Fernandez J, Williams NA, *et al*. Gene duplications and the origins of vertebrate development. *Dev Suppl* 1994:125–33.
12. Blomme T, Vandepoele K, De Bodt, S *et al*. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol* 2006;**7**:R43.
13. Amores A, Force A, Yan YL, *et al*. Zebrafish hox clusters and vertebrate genome evolution. *Science* 1998;**282**:1711–14.
14. Taylor JS, Van de Peer Y, Braasch I, *et al*. Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci* 2001;**356**:1661–79.
15. Christoffels A, Koh EG, Chia JM, *et al*. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol* 2004;**21**(6):1146–51.
16. Donoghue PC, Purnell MA. Genome duplication, extinction and vertebrate evolution. *Trends Ecol Evol* 2005;**20**:312–19.
17. Abzhanov A. von Baer's law for the ages: lost and found principles of developmental evolution. *Trends Genet* 2013;**29**:712–22.
18. Gould SJ. *Ontogeny and Phylogeny*. Cambridge, MA: Belknap Press of Harvard University Press, 1977.
19. Slack JM, Holland PW, Graham CF. The zootype and the phylotypic stage. *Nature* 1993;**361**:490–2.
20. Richardson MK. Heterochrony and the phylotypic period. *Dev Biol* 1995;**172**:412–21.
21. Raff RA. *The Shape of Life: Genes, Development, and the Evolution of Animal Form*. Chicago, IL: University of Chicago Press, 1996.
22. Davidson EH, Erwin DH. Gene regulatory networks and the evolution of animal body plans. *Science* 2006;**311**:796–800.
23. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 2008;**134**:25–36.
24. Duboule D. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev Suppl* 1994:135–42.
25. Roux J, Robinson-Rechavi M. Developmental constraints on vertebrate genome evolution. *PLoS Genet* 2008;**4**:e1000311.
26. Comte A, Roux J, Robinson-Rechavi M. Molecular signaling in zebrafish development and the vertebrate phylotypic period. *Evol Dev* 2010;**12**:144–56.
27. Gerstein MB, Rozowsky J, Yan KK, *et al*. Comparative analysis of the transcriptome across distant species. *Nature* 2014;**512**:445–8.
28. Kalinka AT, Varga KM, Gerrard DT, *et al*. Gene expression divergence recapitulates the developmental hourglass model. *Nature* 2010;**468**:811–14.
29. Irie N, Kuratani S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun* 2011;**2**:248.
30. Domazet-Loso T, Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 2010;**468**:815–18.
31. Drost HG, Gabel A, Grosse I, *et al*. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol Biol Evol* 2015;**32**(5):1221–31.
32. Kalinka AT, Tomancak P. The evolution of early animal embryos: conservation or divergence? *Trends Ecol Evol* 2012;**27**:385–93.
33. Satou Y, Satoh N. Gene regulatory networks for the development and evolution of the chordate heart. *Genes Dev* 2006;**20**:2634–8.
34. Nikitina N, Sauka-Spengler T, Bronner-Fraser M. Chapter 1: Gene regulatory networks in neural crest development and evolution. *Curr Top Dev Biol* 2009;**86**:1–14.
35. Davidson EH. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. Amsterdam: Academic Press, 2006.
36. Nelson AC, Wardle FC. Conserved non-coding elements and cis regulation: actions speak louder than words. *Development* 2013;**140**:1385–95.
37. Jeziorska DM, Jordan KW, Vance KW. A systems biology approach to understanding cis-regulatory module function. *Semin Cell Dev Biol* 2009;**20**:856–62.
38. McPherson JD, Marra M, Hillier L, *et al*. A physical map of the human genome. *Nature* 2001;**409**:934–41.
39. Venter JC, Adams MD, Myers EW, *et al*. The sequence of the human genome. *Science* 2001;**291**:1304–51.
40. Waterston RH, Lindblad-Toh K, Birney E, *et al*. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;**420**:520–62.
41. Brenner S, Elgar G, Sandford R, *et al*. Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. *Nature* 1993;**366**:265–8.
42. International_Chicken_Genome_Sequencing_Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 2004;**432**:695–716.
43. Kasahara M, Naruse K, Sasaki S, *et al*. The medaka draft genome and insights into vertebrate genome evolution. *Nature* 2007;**447**:714–19.
44. Hellsten U, Harland RM, Gilchrist MJ, *et al*. The genome of the Western clawed frog Xenopus tropicalis. *Science* 2010;**328**:633–6.
45. Howe K, Clark MD, Torroja CF, *et al*. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 2013;**496**:498–503.
46. Bejerano G, Pheasant M, Makunin I, *et al*. Ultraconserved elements in the human genome. *Science* 2004;**304**:1321–5.
47. Sandelin A, Wasserman WW, Lenhard B. ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res* 2004;**32**:W249–52.
48. Vavouri T, Walter K, Gilks WR, *et al*. Parallel evolution of conserved non-coding elements that target a common set of developmental regulatory genes from worms to humans. *Genome Biol* 2007;**8**:R15.
49. McEwen GK, Goode DK, Parker HJ, *et al*. Early evolution of conserved regulatory sequences associated with development in vertebrates. *PLoS Genet* 2009;**5**:e1000762.
50. Royo JL, Maeso I, Irimia M, *et al*. Transphyletic conservation of developmental regulatory state in animal evolution. *Proc Natl Acad Sci USA* 2011;**108**:14186–91.
51. Clarke SL, VanderMeer JE, Wenger AM, *et al*. Human developmental enhancers conserved between deuterostomes and protostomes. *PLoS Genet* 2012;**8**:e1002852.
52. Woolfe A, Goodson M, Goode DK, *et al*. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 2005;**3**:e7.
53. de la Calle-Mustienes E, Feijoo CG, Manzanares M, *et al*. A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. *Genome Res* 2005;**15**:1061–72.
54. Pennacchio LA, Ahituv N, Moses AM, *et al*. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 2006;**444**:499–502.

55. Nobrega MA, Ovcharenko I, Afzal V, *et al*. Scanning human gene deserts for long-range enhancers. *Science* 2003;**302**:413.
56. Ritter DI, Li Q, Kostka D, *et al*. The importance of being cis: evolution of orthologous fish and mammalian enhancer activity. *Mol Biol Evol* 2010;**27**:2322–32.
57. Fisher S, Grice EA, Vinton RM, *et al*. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* 2006;**312**:276–9.
58. McGaughey DM, Stine ZE, Huynh JL, *et al*. Asymmetrical distribution of non-conserved regulatory sequences at PHOX2B is reflected at the ENCODE loci and illuminates a possible genome-wide trend. *BMC Genomics* 2009;**10**:8.
59. Blow MJ, McCulley DJ, Li Z, *et al*. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet* 2010;**42**:806–10.
60. Taher L, McGaughey DM, Maragh S, *et al*. Genome-wide identification of conserved regulatory function in diverged sequences. *Genome Res* 2011;**21**:1139–49.
61. Lee AP, Kerk SY, Tan YY, *et al*. Ancient vertebrate conserved noncoding elements have been evolving rapidly in teleost fishes. *Mol Biol Evol* 2011;**28**:1205–15.
62. Venkatesh B, Kirkness EF, Loh YH, *et al*. Ancient noncoding elements conserved in the human genome. *Science* 2006;**314**:1892.
63. Holland LZ, Albalat R, Azumi K, *et al*. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res* 2008;**18**:1100–11.
64. Putnam NH, Butts T, Ferrier DE, *et al*. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 2008; **453**:1064–71.
65. Visel A, Prabhakar S, Akiyama JA, *et al*. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet* 2008;**40**:158–60.
66. Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat Rev Genet* 2010;**11**:476–86.
67. The_ENCODE_Project_Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**:57–74.
68. Roadmap_Epigenomics_Consortium, Kundaje A, Meuleman W, *et al*. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;**518**:317–30.
69. Kunarso G, Chia NY, Jeyakani J, *et al*. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* 2010;**42**:631–4.
70. Schmidt D, Wilson MD, Ballester B, *et al*. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 2010;**328**:1036–40.
71. Mikkelsen TS, Xu Z, Zhang X, *et al*. Comparative epigenomic analysis of murine and human adipogenesis. *Cell* 2010;**143**:156–69.
72. Villar D, Flicek P, Odom DT. Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nat Rev Genet* 2014;**15**:221–33.
73. He Q, Bardet AF, Patton B, *et al*. High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. *Nat Genet* 2011;**43**:414–20.
74. Stefflova K, Thybert D, Wilson MD, *et al*. Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* 2013;**154**:530–40.
75. Kasowski M, Grubert F, Heffelfinger C, *et al*. Variation in transcription factor binding among humans. *Science* 2010;**328**:232–5.
76. Ong CT, Corces VG. Enhancers: emerging roles in cell fate specification. *EMBO Rep* 2012;**13**:423–30.
77. Creyghton MP, Cheng AW, Welstead GG, *et al*. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* 2010;**107**(50):21931–6.
78. Rada-Iglesias A, Bajpai R, Swigut T, *et al*. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 2011;**470**:279–83.
79. Schwaiger M, Schonauer A, Rendeiro AF, *et al*. Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res* 2014;**24**:639–50.
80. Xiao S, Xie D, Cao X, *et al*. Comparative epigenomic annotation of regulatory DNA. *Cell* 2012;**149**:1381–92.
81. Cotney J, Leng J, Yin J, *et al*. The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* 2013;**154**:185–96.
82. Villar D, Berthelot C, Aldridge S, *et al*. Enhancer evolution across 20 mammalian species. *Cell* 2015;**160**:554–66.
83. Nord AS, Blow MJ, Attanasio C, *et al*. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* 2013;**155**:1521–31.
84. Bogdanovic O, Fernandez-Minan A, Tena JJ, *et al*. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res* 2012;**22**:2043–53.
85. Tena JJ, Gonzalez-Aguilera C, Fernandez-Minan A, *et al*. Comparative epigenomics in distantly related teleost species identifies conserved cis-regulatory nodes active during the vertebrate phylotypic period. *Genome Res* 2014;**24**(7):1075–85.
86. Neph S, Vierstra J, Stergachis AB, *et al*. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 2012;**489**:83–90.
87. Thurman RE, Rynes E, Humbert R, *et al*. The accessible chromatin landscape of the human genome. *Nature* 2012; **489**:75–82.
88. Boyle AP, Davis S, Shulha HP, *et al*. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 2008;**132**:311–22.
89. Giresi PG, Kim J, McDaniell RM, *et al*. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* 2007;**17**:877–85.
90. Buenrostro JD, Giresi PG, Zaba LC, *et al*. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 2013;**10**:1213–18.
91. Zentner GE, Henikoff S. High-resolution digital profiling of the epigenome. *Nat Rev Genet* 2014;**15**:814–27.
92. Stergachis AB, Neph S, Sandstrom R, *et al*. Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* 2014;**515**:365–70.
93. Vierstra J, Rynes E, Sandstrom R, *et al*. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* 2014;**346**:1007–12.
94. Lopez-Rios J, Duchesne A, Speziale D, *et al*. Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. *Nature* 2014;**511**:46–51.
95. Hwang WY, Fu Y, Reyon D, *et al*. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 2013;**31**:227–9.
96. Mali P, Yang L, Esvelt KM, *et al*. RNA-guided human genome engineering via Cas9. *Science* 2013;**339**:823–6.