Evolutionary developmental biology and genomics

Cristian Cañestro, Hayato Yokoi and John H. Postlethwait

Abstract | Reciprocal questions often frame studies of the evolution of developmental mechanisms. How can species share similar developmental genetic toolkits but still generate diverse life forms? Conversely, how can similar forms develop from different toolkits? Genomics bridges the gap between evolutionary and developmental biology, and can help answer these evo–devo questions in several ways. First, it informs us about historical relationships, thus orienting the direction of evolutionary diversification. Second, genomics lists all toolkit components, thereby revealing contraction and expansion of the genome and suggesting mechanisms for evolution of both developmental functions and genome architecture. Finally, comparative genomics helps us to identify conserved non-coding elements and their relationship to genome architecture and development.

Developmental genetic toolkit
A set of genes that is required for development and is shared widely among species.

Phylogenomics
Phylogenetic inference on a genome-wide scale.

Ecdysozoan
A group of protostomes that unites the phyla Arthropoda (including flies) and Nematoda (including roundworms), among others.

Institute of Neuroscience, 1425 East 13th Avenue, University of Oregon, Eugene, Oregon 97403, USA. Correspondence to J.H.P. e-mail: jpostle@uoregon.edu
doi:10.1038/nrg2226

"What characterizes the living world is both its diversity and its underlying unity." (Jacob, 1977)[1]

How can a conserved, broadly shared developmental genetic toolkit generate today's amazing diversity of life forms[2–4]? The conservation of the toolkit became evident from dramatic discoveries such as the finding that Hox genes control anterior–posterior patterning in both a fly and a mouse[5], and that the human paired box gene 6 (*PAX6*), which is necessary for eye development, can cause cells in the primordial wing of a fly embryo to become eye cells[6]. Early attempts to explain the paradox that was implied by Jacob 30 years ago suggested that phenotypic diversity derives from differences in where and when genes are expressed, rather than in the products that the genes encode[1,7]. Owing to rapid genome-sequencing technologies, biologists have access, for the first time, to the full list of toolkit components for a wide variety of species. Comparison of whole-genome sequences can reveal changes to the toolkit in an evolutionary perspective and suggest hypotheses for the origin of phenotypic diversity.

This Review explores just four topics from the various ways that genomics impacts evolutionary developmental biology, with a bias towards examples from animals with which the authors are most familiar. Because knowing lineage relationships is essential to map the orientation of trait gain and loss, we first describe the impact of genomic data on our understanding of organismal phylogenies. The second section shows how genome contraction events can help identify trait-specific genes, affect genome architecture and lead to alternative modes of development. Genome contraction reveals an inverse paradox: cases in which organisms develop fundamentally similar morphologies (phenotypic unity) despite important differences in genetic toolkits (genetic diversity). Third, we explore how genome-expansion events can augment the complexity of the developmental toolkit. The final section examines the impact of comparative genomics on understanding the influence of genome architecture on gene regulation as a force for phenotypic diversity.

Phylogenomics and developmental biology

Accurate knowledge of phylogenies among organisms is important to understand the direction of change when one lineage possesses a trait that is missing from its sister lineage. Was the trait absent from the last common ancestor and gained in one lineage? Or was it present in the last common ancestor but lost in one lineage? Phylogenomics can improve the accuracy of phylogenetic analysis by using thousands of concatenated, unambiguously aligned amino-acid positions from hundreds of genes supplied by full genome sequences from many organisms[8,9]. Furthermore, the presence or absence of rare genomic changes — such as gene fusions, transposable element insertions or intron positions — provides additional valuable markers to assess phylogenetic relatedness. Nevertheless, phylogenomic analysis has limitations and can sometimes lead to contradictory results, such as the lingering controversy regarding the validity of the taxon ecdysozoa[10]. Problems can arise due to

rapid, ancient cladogenesis, the abundance of homoplastic characters, rapidly evolving positions within proteins and rapidly evolving lineages[9]. Nevertheless, recent advances in methods to detect systematic errors, improvements in data quality, wider taxonomic sampling and the identification of new markers of biological history help to improve our knowledge of the tree of life[11].

Some recent examples show the power of phylogenomics for evolutionary developmental biology (FIG. 1). Classically, cnidarians, a basally diverging group of animals, have a radial, bag-like body plan with a body cavity that opens to the exterior through an orifice that acts both as a mouth and an anus[12]. However, phylogenomic analysis showed that a muscular parasitic worm (*Buddenbrockia plumatellae*) is a cnidarian[13]. This finding increases the known diversity of cnidarian body plans and poses new questions for understanding the genetic control of cnidarian development. Similarly, phylogenomic analysis has finally solved the enigmatic evolutionary position of *Xenoturbellida*, a ciliated marine worm that was initially thought to be related to acoelomorph flatworms; *Xenoturbellida* now is placed in a new phylum within the deuterostomes (FIG. 1) as the sister group of the ambulacraria[14]. This example broadens the known morphologies of the sister group to the chordates, our own phylum.

A third example of the power of phylogenomics dethrones cephalochordates (for example, amphioxus) as the long-assumed surviving sister lineage of vertebrates[12]. This position is now occupied by urochordates

Cladogenesis
The process in which lineages of organisms diverge into separate clades — groups of organisms, all of which are descended from a single common ancestor.

Homoplastic character
Characters that are similar owing to convergent evolution rather than derivation from a single character in the last common ancestor.

Deuterostomes
'Deutero' (second), 'stome' (mouth). Bilaterian animals in which the first opening of the embryo forms the anus, whereas a second opening forms the mouth, in contrast to protostomes — bilaterians in which the first embryonic opening forms the mouth. Deuterostomes include chordates, echinoderms and hemichordates.

Ambulacraria
A taxon containing the phyla Echinodermata (including sea stars and sea urchins) and Hemichordata (including acorn worms).

Chordates
Our own phylum, which includes three subphyla: Vertebrata (including fish, amphibia, reptiles, birds and mammals), Cephalochordates (like amphioxus) and Urochordates (like ascidians and larvaceans).

Urochordates
The subphylum of chordates that is the sister group to the vertebrates, including ascidians (or sea squirts), a class forming a tadpole larva with a chordate body plan that is destroyed by a radical metamorphosis to form a sessile adult, and larvaceans, a class of mostly planktonic animals that maintains a chordate body plan throughout its life cycle. Also called tunicates.

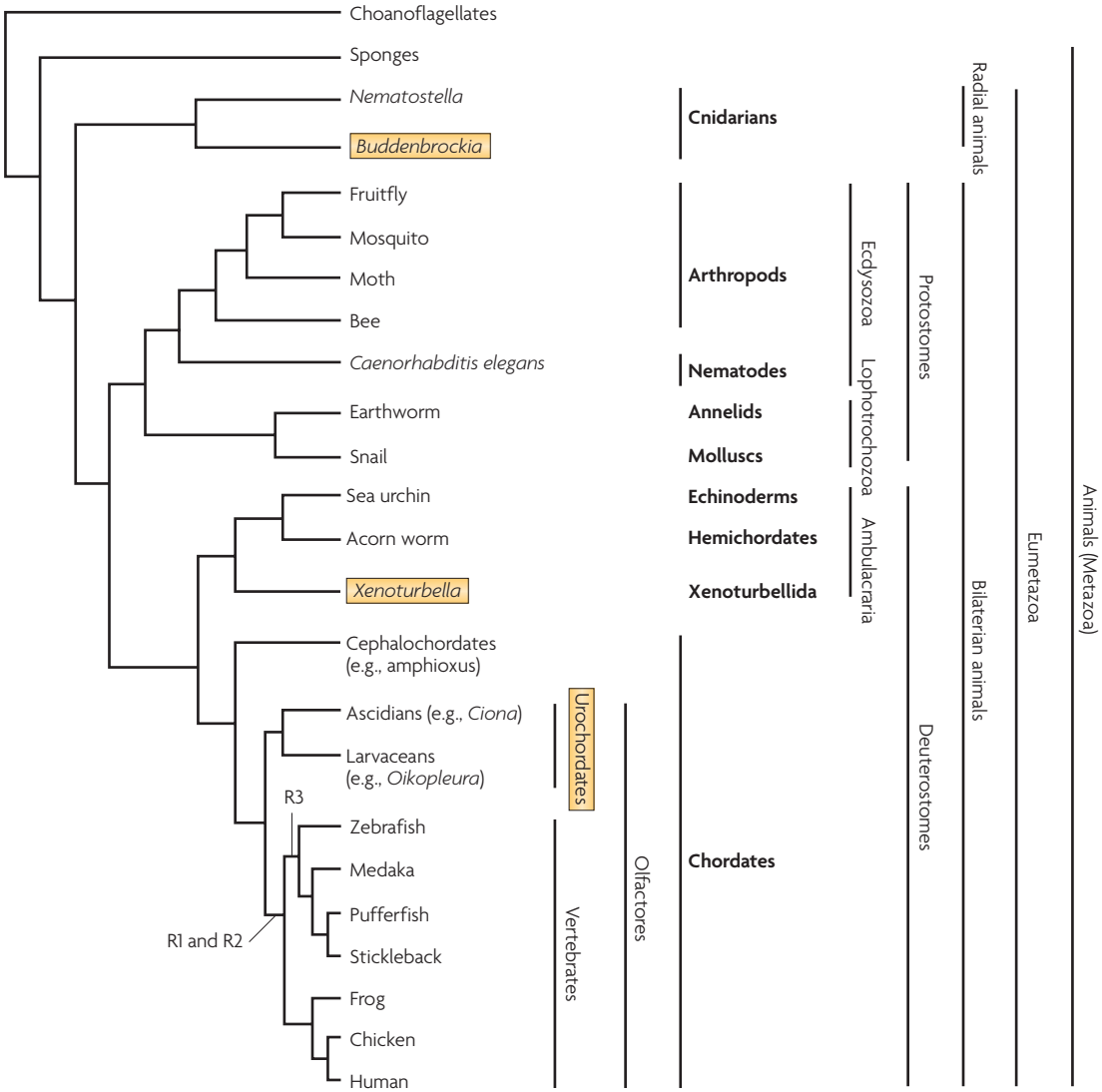

Figure 1 | **Phylogenomics improves our understanding of the historical relationships of organismal diversity.** Phylogenetic information is essential to determine whether developmental features that are present in one group, but missing from the sister group, have resulted from the gain of new genetic mechanisms in one lineage or the secondary loss of ancestral mechanisms in the other. Recent revisions are highlighted in yellow: *Buddenbrockia*, which extends the morpho-space of Cnidarians from solely bag-like animals to worm-like forms; *Xenoturbella*, which represents a new phylum of Deuterostomes; and the revision of urochordates, rather than cephalochordates, as the sister of vertebrates, which provides new insights into the origin of character states in chordates. R1–R3, rounds of whole-genome duplication.

Olfactores
A chordate taxon including the two subphyla Vertebrata and Urochordata.

DNA methylation
A DNA modification in which a methyl group is added to cytosine. Methylation inhibits gene expression and is maintained through DNA replication and cell division.

Epigenetic
Factors that affect gene action without changing nucleotide sequence. Epigenetic modifications act by changing the structure of chromatin, and are facilitated by DNA methylation and histone modification.

(for example, ascidians and larvaceans), making a new group, the olfactores[14,15]. Rare genomic changes that are shared by urochordates and vertebrates, including the domain organization of the cadherin gene family and a unique amino-acid insertion in the coding region of fibrillar collagen genes, support this conclusion[16,17]. In addition, some morphological features support the constitution of the new group, such as neural crest-like cells and epidermal placodes[18–20]. Other features in stem olfactores, such as a complex tripartite brain, might have been secondarily simplified in urochordates rather than having evolved in vertebrate phylogeny[21–23]. Future work will distinguish between features that were absent in stem olfactores and evolved in vertebrates from features that were possessed by stem olfactores and lost secondarily in urochordates (for example, REFS 24,25).

Genome contraction and development

Accurate phylogenies help us to identify traits that were present in a clade's common ancestor but were secondarily lost, which can lead to the loss of genes that were used exclusively for that trait. Conversely, genes that were assumed to be important for a given trait can be lost without the loss of the trait, the inverse paradox. This section reviews three cases of genome contraction that involve cilia and flagella, DNA methylation and retinoic acid (RA) signalling to show how analysis of genome contraction can identify genes that are important for a given trait, and how investigation of genome contraction can suggest hypotheses for the evolution of genome architecture and for the innovation of alternative modes of development.

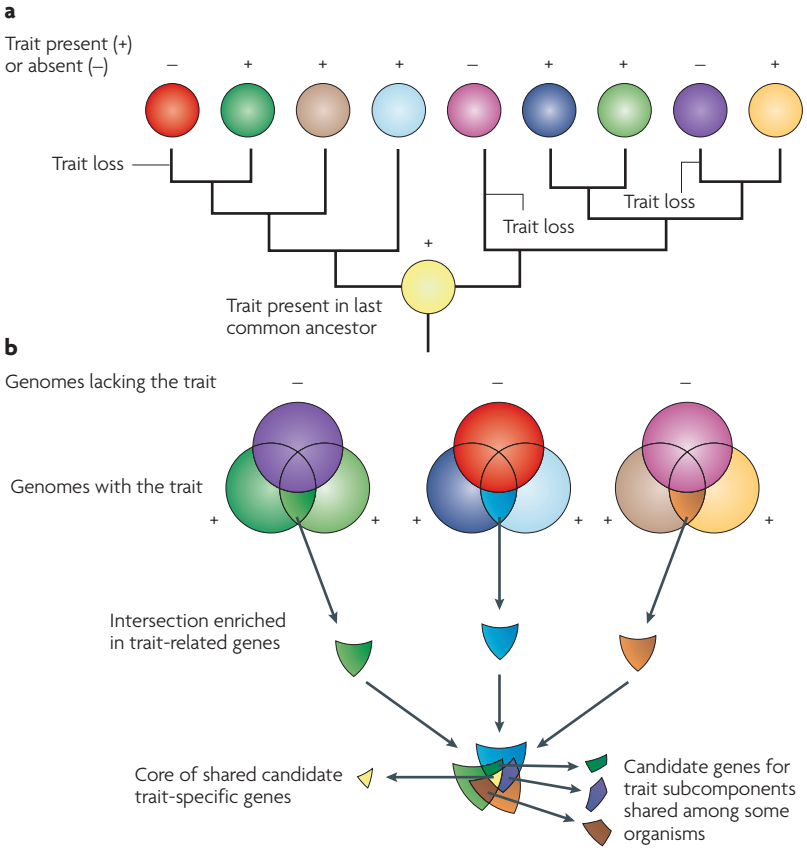

Figure 2 | **Comparative genomics is a tool to identify trait-specific genes.** Comparison of diverse genomes in a clade that have variably lost a trait (shown in part **a**) identifies candidate trait-related genes that are present in organisms that have the trait but are absent in organisms that lack the trait (shown in part **b**). Combinatorial comparisons at different levels of stringency that depend on evolutionary distance provides different sets of gene candidates that can reveal the core of trait-specific genes shared among all organisms and candidate genes for trait subcomponents shared by only some of the organisms.

Trait loss illuminates trait-specific genes. Comparative genomics provides a powerful tool to discover trait-specific genes on the basis of the assumption that most genes that are expressed exclusively in a given trait are lost if the trait is lost[26,27]. The strategy compares genomes in a clade, the members of which vary with respect to the presence or absence of an ancestral trait (FIG. 2a). The intersection of genes in genomes with the trait, after subtracting genes in genomes without the trait, is enriched in candidate trait-related genes (FIG. 2b). Genome comparisons at stringencies that are appropriate for evolutionary distance can suggest candidates for the basic core of common trait-specific genes. Comparison of different genome subgroups can identify subsets of candidate trait-related genes that are involved in variably present subcomponents of the trait (FIG. 2b). The power of this strategy was demonstrated in comparative genomics of cilia and flagella, microtubule-based organellar whips, which are important for development of left–right asymmetries, heart formation, vertebrate photoreceptors, and invertebrate mechano- and chemoreceptors[28]. Comparative genomics of organisms with cilia (such as flies, roundworms, green algae, protists and humans), and organisms that lack cilia (such as plants, yeasts and slime moulds)[27], and comparison of organisms with flagella (such as green algae, flies, roundworms, sea squirts, mice and humans) and organisms that lack flagella (such as plants)[26], identified several hundred candidate genes related to cilia or flagella. Finding more than 80% of ancestral genes that are known to be involved in cilia function verified the method. The analysis identified a novel family of proteins (OSEG) that are essential for the development of cilia in *Drosophila melanogaster*[27]. Studies *in silico*, *in vitro* and *in vivo* in *Caenorhabditis elegans* validated flagella-related genes, and identified a novel human gene (*BBS5*) as defective in Bardet–Biedl syndrome[26]. Further applications of this genomic strategy will facilitate the identification of candidate genes that are important for the development and evolution of a variety of traits.

Contraction of the DNA-methylation toolkit. Gene silencing by DNA methylation has a fundamental role in gene regulation during vertebrate development[29]. DNA methylation is an epigenetic mechanism based on cell inheritance without mutation[30]. Vertebrate genomes are heavily methylated, but the genomes of many non-vertebrates are much less methylated[31–33]. How does the evolution of this epigenetic system correlate with the evolution of developmental mechanisms, the preservation of genome architecture and the generation of phenotypic diversity?

Evolutionary changes that affect chromatin-based epigenetic systems are potentially important for phenotypic diversity. Unmethylated regions of the genome usually contain highly expressed genes that are precisely regulated by transcription factors, but highly methylated regions often contain less active, more broadly expressed genes[34,35]. This distribution suggests that DNA methylation helps to suppress spurious initiation of transcription. DNA methyltransferases (DNMTs) are key players in DNA methylation[36]. DNMT3 methylates DNA *de novo* during development, and DNMT1 guides subsequent epigenetic inheritance. DNMT2 shows low activity on DNA but higher activity on specific tRNA molecules, although its full role remains enigmatic[37]. Methylated DNA recruits methyl-CpG-binding domain proteins (MBDs) and their associated histone deacetylases (HDACs), resulting in tighter chromatin packaging and locally reduced access of transcription factors to target genes. Epigenetic mechanisms that alter chromatin condensation, and thereby help to activate or silence genes in a chromosome neighbourhood, thus regulate genes on the basis of genome architecture, a level of regulation that is superimposed on the level of gene-specific *cis*-regulatory elements[38].

Fruitflies and nematodes have little or no methylated DNA, raising doubts about its general significance for development in non-vertebrates. However, genome analyses reveal that, although the nematode *C. elegans* lacks *dnmt* genes, related nematodes preserve a *dnmt2*-related gene, suggesting a recent loss of the methylation machinery in nematode evolution[32]. Among insects, fruitflies and mosquitoes have only *DNMT2*, and a silk moth has both *Dnmt1* and *Dnmt2* (REFS 39,40), but the honeybee possesses a full set of *Dnmt* genes that are functionally comparable to their vertebrate counterparts. These results show that the full set of *Dnmt* genes was present in the last common ancestor of bilaterians, but that the *Dnmt* toolkit experienced multiple independent contractions in protostome lineages[41]. Among deuterostomes, our analyses of genome databases suggest that sea urchins, cephalochordates and ascidian urochordates have the full complement of *Dnmt* genes, whereas the larvacean urochordate *Oikopleura dioica* has only *Dnmt2*, revealing a contraction of the larvacean toolkit despite the morphological similarities of the ascidian and larvacean larvae. Our analysis of the recently available genome sequence of the cnidarian *Nematostella vectensis* identified all three *Dnmt* genes. These results show that the full set of *Dnmt* genes, which was already present in the last common ancestor of radial and bilaterian animals, has been truncated in larvaceans and multiple times in protostomes, but not in some cnidarians or in the vertebrate lineage.

Contraction of the Dnmt toolkit: an inverse paradox. What allowed the DNA-methylation toolkit to contract in some lineages but not in others? How can the fundamentally similar body plans of a bee and a fly develop either with or without regulation provided by DNA methylation? This problem illustrates the inverse paradox (genetic diversity despite body-plan unity), and any of several hypotheses might explain it. First, developmental mechanisms might have become independent of DNA methylation in a stem ecdysozoan, leading to relaxation of the selective constraints to maintain DNA methylation; lineages that have maintained DNA-methylation machinery, such as that of the honeybee, may use it for non-developmental functions such as imprinting or complex social behaviours[41]. An alternative hypothesis is that the epigenetic system of chromatin change is important for development in lineages that lack *Dnmt* genes, but alternative mechanisms provide this function. The fidelity of epigenetic gene silencing is probably increased by interactions between the DNA-methylation and histone-modification systems[42]. In flies, factors other than DNMTs may cooperate with histone modification to facilitate changes in chromatin structure[43].

Evolution of genome architecture and epigenetics. The evolution of genome architecture might help to explain how the DNA-methylation toolkit can contract in some lineages but not others. Chromosome rearrangements can disrupt coherent regions of epigenetic gene regulation because genes that are translocated from a highly methylated region could become deregulated after transfer to an undermethylated region. This suggests that the evolution of genome architecture can depend on the epigenetic system. This hypothesis predicts that syntenies should tend to be conserved between lineages that preserve ancestral epigenetic systems. This prediction agrees with results from recent genomic analyses showing that vertebrates share with cnidarians, but not with well investigated protostomes, both extensive conserved syntenies and a full *Dnmt* toolkit[44]. In the case of *O. dioica*, the loss of two of the three *Dnmt* genes might be linked to the contraction of their genome, the smallest among chordates, which was accompanied by extensive genomic rearrangements[45,46]. The lack of conserved patterns of nuclear compartmentalization, and the lack of correlation between active transcription and domains that are rich in histone-specific modifications[47], suggest that the epigenetic system might be altered in larvaceans, perhaps as an adaptation to their determinative mode of development and rapid life cycles. Future functional analysis will be needed to understand the importance of variation in the epigenetic control toolkit, the evolution of genome architecture and their impact on mechanisms of development.

Genome contraction, RA and Hox clusters. The relationship of genome architecture to gene regulation is evident in the Hox clusters, groups of tandemly duplicated genes that encode homeodomain-containing transcription factors that are important for organizing the bilaterian anterior–posterior body axis (reviewed in REF. 48). In vertebrate genomes, the order of Hox genes roughly matches both the order of expression along the body axis (spatial collinearity) and the order of expression during development (temporal collinearity)[49]. Spatial collinearity depends mainly on *cis*-regulatory elements, but temporal collinearity depends on the architecture of the Hox cluster[45,49,50].

Bilaterians
A taxon of animals with a bilaterally symmetrical body plan, in contrast to the basally diverging radiata, which have a radial body plan. Includes cnidarians such as sea anemones and jellyfish.

Synteny
(Same thread). A set of genes on the same chromosome (clearly, two genes in a fish and the orthologues of those two genes in a human are not on the same chromosome and so can't be syntenic).

Conserved synteny
A situation in which a set of syntenic genes in one species has orthologues that are syntenic in another species.

Determinative development
A developmental mode in which cell fate becomes fixed very early in embryonic development.

Hox clusters
A group of tandemly duplicated genes encoding homeodomain-containing transcription factors that control the development of animal body axes.

Collinearity
In Hox clusters, genes located 3′ in the cluster are expressed earlier (temporal collinearity) and more anteriorly (spatial collinearity) than genes that lie more 5′ in the cluster.

In vertebrates and cephalochordates, RA helps to regulate temporal collinearity[51,52]. RA gradually increases the portion of a Hox cluster that is poised outside territories of condensed chromatin, allowing genes along the cluster to gradually access transcription machinery over time[53,54]. RA positions decondensed chromatin with respect to Hox genes by chromatin remodelling induced by DNA methylation, histone methylation, acetylation and deacetylation[55,56]. RA binds to a retinoic acid receptor (RAR), which heterodimerizes with a retinoid-X receptor (RXR) at RA-response elements in or near target genes[57]. RA-activated RAR recruits protein complexes that contain histone acetyltransferases that induce gradual changes in chromatin structure. The classical genetic machinery for RA action also includes enzymes that synthesize RA (such as ALDH1A) and degrade RA (such as CYP26), which together regulate the distribution of RA during development[58].

Because the role of RA in anterior–posterior axial patterning seemed to be limited to chordates, and because the main components of RA signalling (ALDH1A, CYP26 and RAR) had been described only in chordates, it was supposed that the 'invention' of RA genetic machinery was a key innovation for development of the chordate body plan, probably mediated by Hox genes in axial patterning (reviewed in REFS 59,60). However, recent genome analyses revealed the unexpected presence of RA genetic machinery in non-chordate deuterostomes[61–63], suggesting that RA signalling is not a chordate invention, or that the chordate innovation was the redeployment of an ancient signalling system for new developmental roles, including the regulation of Hox-cluster expression. Functional analysis of RA action in the development of hemichordates, which share many developmental similarities with chordates[64], will help in evaluating these hypotheses.

Another unexpected result comes from the genome of *O. dioica*. Larvaceans are the only urochordates that maintain a chordate body plan as adults, and yet its deep genome database lacks the classical genes for RA synthesis, degradation and reception[24] (FIG. 3). Because cephalochordates, which diverged basally among chordates (FIG. 1), have the RA toolkit, it must have been secondarily lost in larvacean evolution but preserved in the ascidian lineage. A study of RA action showed that it does not cause homeotic posteriorization in larvacean embryos, in contrast to vertebrates and cephalochordates[24]. These results show that a chordate can develop the phylotypic body plan without genes for the classical morphogenetic role of RA (the inverse paradox), and suggest that larvaceans use alternative mechanisms for the development of chordate features.

Differences in RA toolkits between larvaceans and ascidians are not reflected in drastic differences in embryonic development — the inverse paradox. Evidence suggests that axial patterning independent of RA-signalling is actually a shared, derived feature of urochordates[24,65]. Excess RA in both larvaceans and ascidians seems to alter organ morphogenesis rather than causing Hox-related homeotic transformations. This finding suggests that stem urochordates evolved an alternative developmental mechanism that allowed anterior–posterior axial patterning to become independent of RA (FIG. 3). The RA machinery in ascidians may perform functions such as asexual reproduction and regeneration rather than embryonic axial patterning[66,67].

RA contraction and Hox-cluster disintegration: a model. Although Hox-cluster genes occupy contiguous regions in cephalochordate and vertebrate genomes, and perhaps did so in the last common ancestor of all bilaterians (but see REF. 49), in many genomes Hox-cluster genes are separated into two or more subclusters; for example, in the ascidian *Ciona intestinalis* nine Hox genes appear at five different genomic locations[68], and in *O. dioica* all Hox genes are individually dispersed in the genome[45] (FIG. 3). What features correlate with intact Hox clusters, and what are the consequences of Hox-cluster disintegration? Clearly, Hox-cluster disintegration will thwart the vertebrate mechanism of RA-induced gradual expansion of chromatin relaxation.

Is altered RA signalling in the axial patterning of urochordate embryos causally related to the break up of Hox clusters? The following model could explain most of the data (FIG. 3). Strong selection for rapid embryonic development and life cycle (egg to egg in less than 10 days for *O. dioica*) might simultaneously select for both determinative development, which decreases dependence on extracellular signals such as RA to establish embryonic coordinates, and genome diminution, which is often associated with chromosome rearrangements that can disperse former gene neighbours across the genome (such as *O. dioica* Hox-cluster genes) and disrupt gene regulatory mechanisms that rely on long-range enhancers or chromosome territories, as do vertebrate Hox clusters. This model is consistent with the absence of temporal collinearity of Hox expression in urochordates[45,68] (FIG. 3). Under this model, because RA signalling and DNA methylation become less important, genes that are necessary for their action, such as *Rar* genes and *Dnmt* genes, are free to degrade. Under this model, the disintegration of the Hox cluster and modification of RA signalling in stem urochordates might have led to interesting alternative mechanisms of anterior–posterior axial patterning in urochordate embryos (FIG. 3).

Genome expansion and precision tools

Whereas genome contraction can be associated with the evolution of alternative genetic mechanisms, genome expansion can contribute to the evolution of old tools into new, increasingly specialized devices. In Ohno's classical model[69], one member of a pair of duplicated genes retains the original function whereas its paralogue either disappears by accumulation of detrimental mutations (called non-functionalization[70]) or acquires rare beneficial mutations that confer new, positively selected functions (neofunctionalization). The duplication, degeneration, complementation hypothesis (or DDC model)[70] suggests a third alternative for duplicate preservation: subfunctionalization, the complementary partitioning of ancestral structural and regulatory subfunctions between two duplicate genes so that the sum of their functions equals that of the

Phylotypic body plan
The body organization shared by all members of a phylum.

Paralogues
Genes within the same species that arose by gene duplication within the lineage. For example, *Hoxa1* and *Hoxb1* in mice, or *hoxb1a* and *hoxb1b* in zebrafish.

Non-functionalization
The process whereby a pair of duplicated genes reverts to a single copy as one suffers mutations that produce a non-functional protein.

Neofunctionalization
The process whereby a pair of duplicated genes becomes permanently preserved as one copy acquires mutations conferring a new function that becomes fixed in a population by positive Darwinian selection.

DDC model
(Duplication–degeneration–complementation). A model to explain the evolution of duplicated genes including the complementary loss of subfunctions by degenerative mutations.

Subfunctionalization
The process whereby a pair of duplicated genes becomes permanently preserved because the two gene copies have reciprocally lost essential subfunctions by complementary degenerative mutations.

Subfunction
A specific subset of a gene's regulatory or structural function that, when mutated, establishes a distinct complementation group.

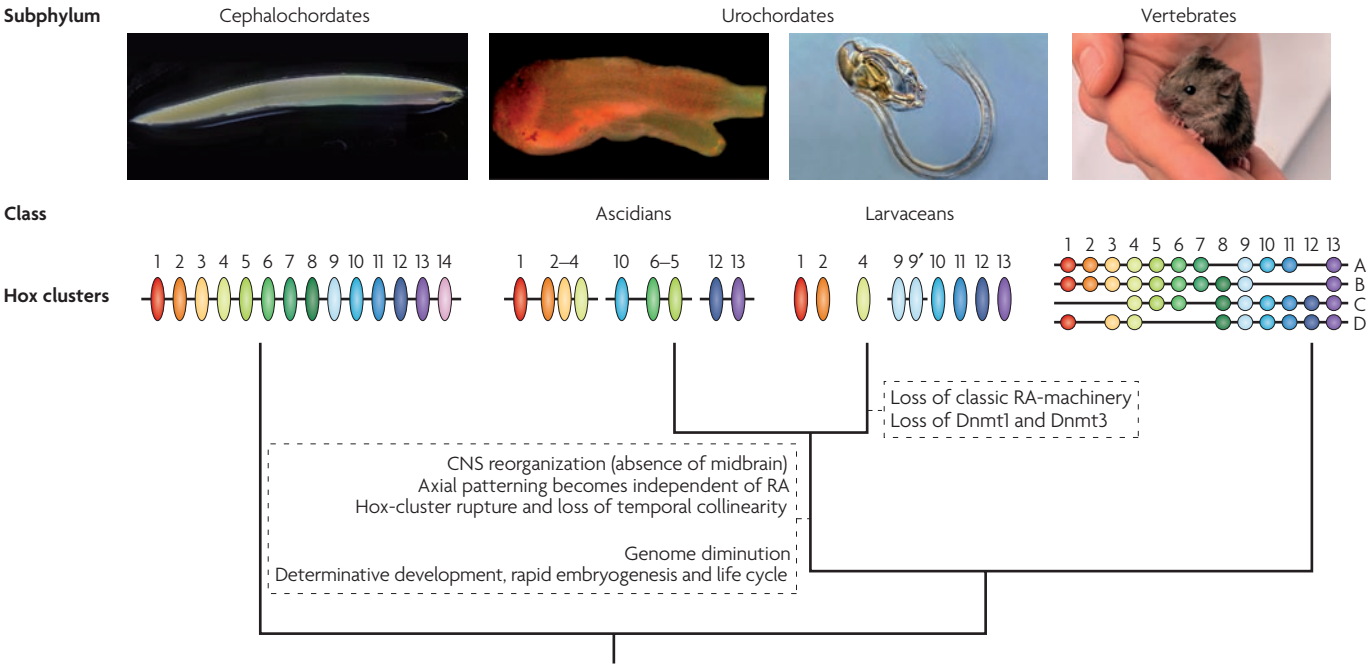

Figure 3 | **Genome contraction and morphology.** Stem urochordates adopted a determinative mode of development, reduced the size of their genomes, lost temporal collinearity of Hox-gene expression, broke up their Hox-gene cluster and lost the need to use retinoic acid (RA) for anteroposterior axial patterning associated with the reorganization of their CNS. Larvaceans lack the classic genetic machinery to synthesize, degrade and detect RA, and they also lack a complete genetic system for DNA methylation (carried out by DNA methyltransferases (Dnmts), but nevertheless build a complete chordate body plan that is retained throughout life. Mouse image courtesy of Getty Images.

parental single-copy gene (see also REF. 71). In the DDC model, it is important to distinguish between subfunctionalization, the initial event that preserves two duplicate genes, and subfunction partitioning, events that occur after the initial preservation of duplicate gene copies[70]. The DDC model predicts that evolutionary constraints on duplicated genes can differ after subfunctionalization owing to relaxed pleiotropy. Because a gene with fewer subfunctions would have fewer diverse tasks, it might more readily accommodate mutations that confer novel functions, leading to the evolution of new tools that are more specifically tailored to specific jobs and thereby contributing to the generation of phenotypic diversity.

Genome expansion and lineage divergence. Lineage-specific non-functionalization and subfunction partitioning can, in principle, provide genetic population-isolating mechanisms[72,73]. This is because F_1 hybrids from the mating of two populations that are fixed for reciprocal non-functionalized or subfunctionalized alternative gene duplicates will produce some F_2 individuals (about 1 in 16 individuals, according to Mendel) that are doubly homozygous for alleles that lack a specific paralogue or subfunction; such individuals will die if the original gene subfunction is essential. If, as after genome duplication, genes on several chromosomes independently experience DDC, then most of the F_1 offspring of two populations will be nearly sterile. This suggests that genome expansion can be an important force for evolutionary diversification[69,73,74].

It is likely that two rounds (R1 and R2) of whole-genome duplication occurred during early vertebrate evolution , and another round occurred at the base of the teleost lineage (R3)[73,75–78] (FIG. 1). Comparative analysis of teleost and human genomes revealed chromosome rearrangements that occurred over a short evolutionary time leading to rapid genome reorganization[77,79]. These events, given the appropriate ecological opportunity, might have facilitated the acquisition of vertebrate innovations and the teleost radiation.

Specialization of FGFs, tools for developmental signalling. As an example of how genome duplication provides opportunities for the evolution of specialized developmental tools, consider the functional evolution of fibroblast growth factor (FGF) gene paralogues that appeared during the vertebrate and teleost radiations. FGFs comprise a family of secreted signalling molecules that control development and homeostasis[80]. Genome analysis shows that mammals have at least 22 *Fgf* genes in seven subfamilies — *FgfA* (1/2), *FgfB* (3/7/10/22), *FgfC* (4/5/6), *FgfD* (8/17/18), *FgfE* (9/16/20), *FgfF* (11/12/13/14) and *FgfG* (19/21/23) — that seem to have expanded in R1 and R2 from seven ancestral proto-*Fgf* genes[81,82] (FIG. 4). The genomic location of *Fgf* genes helps us to infer their evolutionary origin. For example, *Fgf4* of the C group and *Fgf19* of the G group are tightly linked in a 100-kb segment, as are *Fgf6* of the C group and *Fgf23* of the G group, suggesting that *FgfC* and *FgfG* subfamilies arose as tandem duplicates before R1 and R2

Subfunction partitioning
The distribution of gene subfunctions to one or another gene duplicate subsequent to the preservation of both paralogues by subfunctionalization.

Pleiotropy
A genetic phenomenon in which a single gene affects many traits.

Population-isolating mechanisms
Traits that prevent populations of organisms from interbreeding to produce viable, fertile offspring.

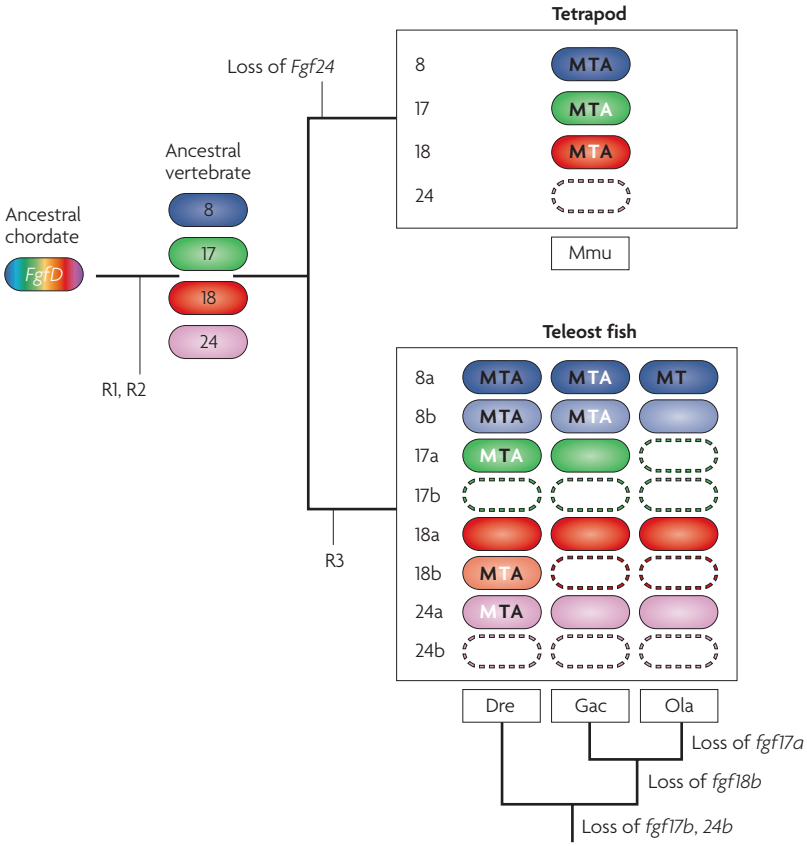

Figure 4 | **Expansion and subfunctionalization of the *FgfD* subfamily.** The figure shows the fibroblast growth factor D (*FgfD*) subfamily expansion from ancestral chordates to vertebrates by whole-genome duplication (WGD). The single *FgfD* gene in ancestral chordates gave rise to four orthologues (*Fgf8, 17, 18* and *24*) after two rounds of WGD (R1 and R2). The tetrapod lineage subsequently lost *Fgf24* but preserved *Fgf8, 17* and *18*. After separating from the tetrapod lineage, the teleost lineage experienced another round of WGD (R3), initially giving eight orthologues (*fgf8a, 8b, 17a, 17b, 18a, 18b, 24a* and *24b*). Duplicated orthologues were lost by non-functionalization or preserved by subfunctionalization in a lineage-specific manner. The presence and absence of orthologues was analysed in the Ensembl and the UTGB medaka genome databases. Phylogenetic relationships of three teleost fish are shown at the bottom, along with expression patterns associated with each gene in each species that demonstrate lineage-specific subfunction partitioning. MTA (M, midbrain–hindbrain boundary; T, tailbud; A, appendage) represents gene expression patterns. A black letter indicates expression, whereas a white letter indicates the absence of expression. Mmu, *Mus musculus* (mouse); Dre, *Danio rerio* (zebrafish); Gac, *Gasterosteus aculeatus* (stickleback); Ola, *Oryzias latipes* (medaka).

Eumetazoans
All animals (metazoa) except sponges.

Orthologues
Genes in different species that derive from the same gene in the last common ancestor of those species, for example, *Hoxb1* in mice, *HOXB1* in humans and *hoxb1a* in zebrafish.

(REF. 81). Thus, ancestral chordates are likely to have had six *Fgf* genes (*FgfA, B, C/G, D, E* and *F*). Consistent with this hypothesis for the origin of *Fgf* genes, the ascidian *C. intestinalis*, whose lineage diverged before R1 and R2 (FIG. 1), possesses at least five of the six proto-*Fgf*s, plus one unassigned *Fgf*[83]. However, few *Fgf* genes are found in genomic databases of protostomes, for example, only two in the nematode and three in the fruitfly[81,82], raising the question of when *Fgf* subfamilies evolved.

Analysis of the genome of the sea anemone *N. vectensis* (Cnidaria, FIG. 1) helped to answer this question. *Nematostella vectensis* has 13 *Fgf* genes, many of which might have arisen by lineage-specific gene duplication[84] but, according to our analysis, at least four of the six proto-*Fgf* subfamilies are present in *N. vectensis* (C.C., H.Y. & J.H.P., unpublished results: *NvFGF1D*, *FgfA*; *Nv211797*, *FgfB*; *NvFgf8A*, *FgfD*; *Nv212165*, *FgfE*). Together with the analysis of Wnt genes (cnidarians have 11 of 12 known Wnt subfamilies, whereas only 6 are present in ecdysozoans[85]), the analysis of the cnidarian genome reveals an unexpected complexity of the developmental toolkit in basally diverging metazoans, and suggests that a substantial part of the basic chordate developmental toolkit existed already in the last common ancestor of all eumetazoans. Additional genome sequences for a broader sample of organisms will provide a better picture of toolkit history, and will illuminate its consequences in developmental diversification over major animal transitions.

Teleost FgfD subfamily expansion. Gene duplication and loss can alter toolkit composition, but the assignment of toolkit orthologues for species diverging on either side of a developmental transition is important, because biologists can learn how gene functions change and developmental tools specialize only by comparing functions of orthologues in the context of accurate phylogenies. In the past, limited data sometimes rendered misleading gene homologies and incorrect gene nomenclature; for example, zebrafish *fgf8b* was initially mischaracterized as *fgf17* (REFS 86,87). Whole-genome sequences from various species help us to overcome these problems by providing access to all members of each gene family, their positions in the genome and comparative syntenic information across phylogenies.

With no gene loss, R1, R2 and R3 should have produced four orthologues in tetrapods and eight in teleosts. From an ancestral *FgfD* gene, which is represented today by *fgf8/17/18* in the *C. intestinalis* genome[83], vertebrates, taken together, have copies of four predicted paralogues (*Fgf8, 17, 18* and *24*) (FIG. 4), but only three of the four are present in tetrapods (*Fgf8, 17* and *18*), demonstrating lineage-specific loss of *Fgf24* after the divergence of teleosts and tetrapods. Six of the eight predicted paralogues are present in teleosts[86–90] (FIG. 4). Comparative analysis of teleost genomes also reveals lineage-specific loss of *Fgf* paralogues: sticklebacks and pufferfish seem to have lost one *fgf18* gene after diverging from the zebrafish lineage, and the medaka lineage seems to have lost an additional *fgf17* and *fgf18* gene. The hypothesis that lineage-specific subfunction partitioning can erect population-isolation mechanisms predicts this observed type of lineage-specific paralogue loss.

Comparative studies of the *FgfD* group help us to understand the relative roles of subfunction partitioning and non-functionalization, and the origin of novel functions in generating lineage-specific developmental differences. The expression pattern of the single *Fgf8* gene in tetrapods is similar to the summation of the expression domains of *fgf8a* and *fgf8b* in teleosts[86–88]. For example, in mice, *Fgf8* is expressed in somites and in the neural crest[91], whereas, in zebrafish and sticklebacks, *fgf8a* but not *fgf8b* is strongly expressed in somites and, reciprocally, *fgf8b* but not *fgf8a* is strongly expressed in the neural crest[87]. The DDC model predicts that this

type of complementary degeneration of subfunctions provides the opportunity for *fgf8a* to specialize for the somite function whereas its orthologue *fgf8b* could specialize for the neural crest function. Thus, the zebrafish and stickleback orthologues of *fgf8a* and *fgf8b* could form tools that are specialized for different functions in the toolkits of these two lineages.

Analysis of teleost *fgf8* expression patterns suggests that subfunction partitioning might have continued after lineages diverged. For instance, in tetrapods, *Fgf8* is essential for the formation of the midbrain–hindbrain boundary (MHB)[80]. In zebrafish and sticklebacks, both *fgf8a* and *fgf8b* are coexpressed in the MHB[86,87], and analysis of mutant zebrafish shows that *fgf8a* is essential for MHB formation. In medaka, *fgf8a* is also expressed in the MHB but, unexpectedly, it is not necessary for MHB development[92]. Furthermore, in medaka, inhibition of *fgf8a* blocks the formation of the trunk and tail[92] but, in zebrafish, only inhibition of both *fgf8a* and *fgf24* together blocks trunk and tail development[90]. Similarly, paired appendages (limbs and fins) require *Fgf8* in chickens and mice[93,94], but require *fgf24* rather than *fgf8* in fish[90]. These results show that, after the expansion of the *FgfD* subfamily, the duplicates initially retained functional redundancies (appendage function for both *Fgf8* and *Fgf24*) that were eventually resolved differently in different lineages by non-functionalization, subfunction partitioning and, presumably, the evolution of new functions. Some of these lineage-specific differences might have been in place to contribute to lineage diversification, but the hypothesis that they were causative remains to be tested.

Genome architecture and development

The previous two sections discussed studies that illustrate toolkit contraction and expansion and how changes in toolkit unity can contribute to the generation of phenotypic diversity. After genome expansion, orthologues can evolve different expression patterns in different species, and paralogues can have different expression patterns within a species. This section shows how comparative genomics can improve our understanding of the mechanisms by which these differences in expression patterns arise, and how mechanisms of conserved gene expression can relate to evolutionary stability of genome architecture.

Conserved non-coding elements. Comparative genomic analysis led to the surprising discovery that most evolutionarily conserved sequences in mammalian genomes are non-coding elements rather than protein-coding genes[95]. Conserved non-coding elements (CNEs) in amniotes tend to lie in gene-poor areas near developmental genes that encode transcription factors and morphogenetic proteins. Many CNEs include *cis*-acting regulatory elements that affect the activity of nearby genes[96]. These considerations suggest the hypothesis that variation in CNEs may contribute to lineage-specific developmental capabilities, which would be predicted by the idea that evolutionary change in gene regulation is a major force in the generation of phenotypic diversity. Recent studies on colour patterns in flies, for instance, reinforce this idea, and show how independent changes in *cis*-regulatory elements have led to gains and losses of convergent pigment patterns among flies[97].

The DDC model predicts that gene regulatory subfunctions should partition between two paralogues and, if CNEs adequately represent at least a portion of the regulatory subfunctions (note that in some cases function can be conserved even though structure is not[98]), then the hypothesis predicts that ancestral CNEs should distribute between paralogues after genome duplication. Results from comparative studies on plant and animal genomes support this prediction. From a genomic duplication event about 11 million years ago, maize inherited paralogues *liguleless2* (*lg2*) and *liguleless-related sequence-1* (*lrs1*) and, after comparing sequences with their single-copy orthologue in rice, Langham *et al.* found that of 30 original CNEs one was lost from *lg2* and two different CNEs were missing from *lrs1* (REF. 99). The *lg2* gene evolved a new role in the development of the ligule after the duplication event[99], leading to the hypothesis that the partitioning of subfunctions subsequent to the duplication might have facilitated the origin of this new gene function. Among animals, CNE evolution after R3 has been investigated. For example, two zebrafish co-orthologues of human engrailed homeobox 2 (*EN2*) have reciprocally partitioned some ancestral expression domains and CNEs, but share others redundantly[70,73]. A recent systematic analysis of seven pairs of pufferfish gene duplicates arising in R3 revealed a reciprocal loss of CNEs that are shared with the single-copy human gene, as predicted by the subfunction partitioning hypothesis[100].

Paralogues not only partition ancestral CNEs, but also possess specific, partitioned functions. For instance, after being injected into fish or frog embryos, reporter constructs driven by CNEs for *Iroquois* genes express in specific tissues that represent the endogenous expression pattern, and cognate CNEs from different taxa have conserved expression domains[101,102]. CNEs derived from R1 and R2 are evident in the human genome; for example, some of the elements that are present in the fish co-orthologues of *PAX2* are also shared with *PAX5* and *PAX8*, which are paralogues from R1 and R2 (REF. 100). More than 100 small families of CNEs that are duplicated in the human genome can drive the expression of reporter constructs in phylogenetically similar domains in zebrafish embryos[102]. These CNEs must have been present more than 550 million years ago in the ancestral pre-vertebrate genes and have apparently preserved their functions during several rounds of genome expansion. A particularly interesting case involving functional tests of CNEs by morpholino gene knockdown shows subfunction partitioning after both the R2 and R3 events: some functions associated with *HOXA1* in humans are associated with *hoxb1* co-orthologues in zebrafish, consistent with the idea that, after the duplication of the ancestral *Hoxa1/b1* gene, ancient CNEs that were retained by both *Hoxa1* and *Hoxb1* were sorted differently in the tetrapod and teleost lineages; finally, after R3, further subfunction partitioning occurred between the teleost *hoxb1a* and *hoxb1b* genes[103].

Conserved non-coding element
A DNA sequence that is maintained over evolutionary time but whose information does not ultimately appear in the sequence of a protein.

Ligule
A thin sheet on a grass leaf between the sheath and the stem.

Co-orthologues
A pair of gene duplicates, both of which are orthologues of a single gene in a different species.

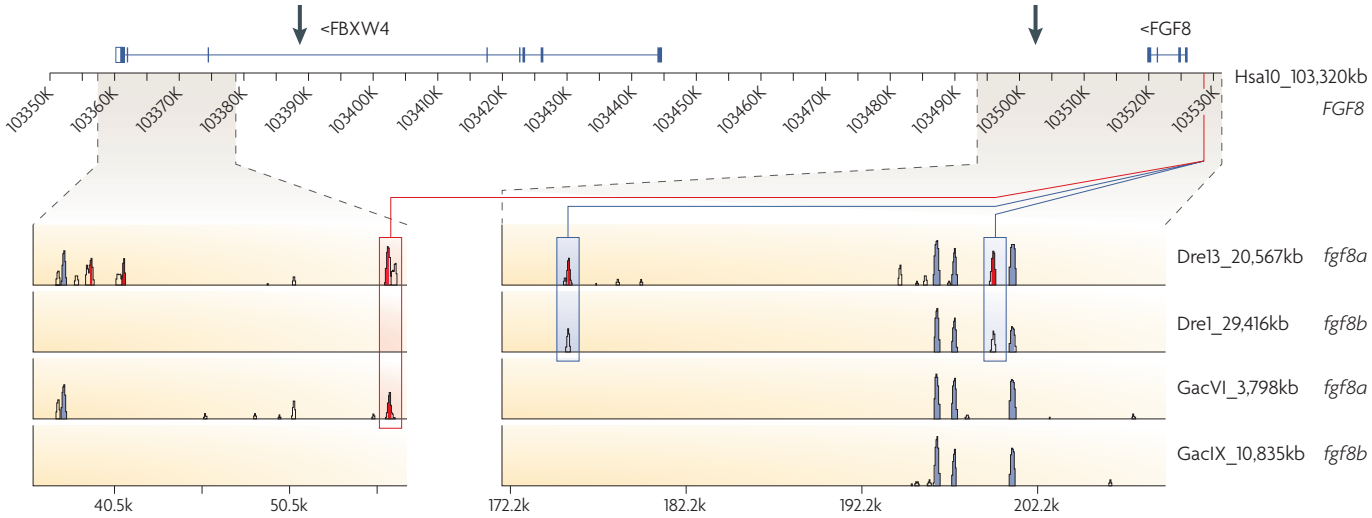

Figure 5 | **Conserved non-coding elements and genomic regulatory blocks.** Sequence comparisons by Vista plot analysis identify CNEs in the intergenic region between *FBXW4* and *FGF8* that are conserved between humans (Hsa), zebrafish (Dre) and sticklebacks (Gac). The CNE indicated by the red line is conserved near *fgf8a*, but not *fgf8b*, in both zebrafish and sticklebacks, suggesting that it partitioned between the two orthologues before these lineages diverged. The CNEs at the tail of the blue lines are conserved by *fgf8a* and *fgf8b* in zebrafish but not in sticklebacks, suggesting lineage-specific subfunction loss. The CNE in *fbxw4* (position of black arrow at top left) apparently helps regulate *fgf8* even though it is within another gene[104]. Orthologues of *FBXW4* and *FGF8* are neighbours in chordates from ascidians to mammals, and the embedded regulatory element in *FBXW4* orthologues might be responsible for preserving this regulatory block.

These studies lend genomic and functional support to the idea that the reciprocal sorting out of CNEs is a common feature of the evolution of duplicated genes derived from genome duplication events.

It will be of particular interest in the future to identify clade-specific CNEs, for example, those that are conserved among perciform fish (including pufferfish, sticklebacks and medakas) but are not in non-perciform fish (including zebrafish), or those that are found in mammals but not birds or amphibia. Clade-specific CNEs become candidates for regulatory elements that programme the developmental novelties that drive evolution.

Conserved syntenies and developmental regulation. Conserved non-coding elements are sometimes located at great distance from the genes they regulate. What are the consequences of these long-range CNEs for genome structure? An informative example comes from the zebrafish co-orthologues of *FGF8*. Becker and colleagues[104] randomly inserted reporters into the zebrafish genome and found four that recapitulate *fgf8a* expression, even though one was in an intron of a neighbouring gene with a different expression pattern (*fbxw4*, which encodes F-box and WD-40 domain protein 4) (FIG. 5). Interestingly, the orthologues of *fgf8a* and *fbxw4* are neighbours not only in humans and zebrafish, but also in the ascidian genome, which diverged from vertebrates before R1 (REF. 78). This result suggests that *fgf8* and *fbxw4* are part of a genomic regulatory block (GRB), the members of which must remain intact to ensure proper gene expression[104]. Human chromosome rearrangements involving *FBXW4* and *FGF8* cause split-hand/foot malformation[105], presumably because they disrupt this GRB. Other human diseases might also result from the disruption of long-range enhancers and, indeed, position-effect human diseases tend to be associated with regions of long-range conserved synteny[106]. These considerations, and the idea that epigenetic mechanisms of global gene regulation might act on large blocks of genes that are located in specific chromosome territories, support the hypothesis that the evolution of genome architecture might be an important factor in the generation of phenotypic variation.

Position-effect human diseases
Diseases associated with chromosome rearrangements that change a gene's position, but do not change the gene's sequence.

Conclusions

Genomics, a descriptive science, has revolutionized our understanding of the history of genome change over time. Comparing the structure of genomes to the evolution of developmental morphologies has transformed our understanding of trait gain and trait loss, and the roles that genome contraction, genome expansion and genome architecture can have in the evolution of developmental mechanisms. These new capabilities unite evolutionary biology, developmental biology and genomics into a new interdisciplinary field. What is now necessary is to turn attention to the genome-wide functional analysis of organisms that are derived from key nodes, a mechanistic science. We must develop technological advances that allow us to turn virtually any species into a 'model organism' for functional studies to appreciate the proximal, developmental causes of morphological change and, eventually, the distal, evolutionary mechanisms of organismal diversification over time.

1. Jacob, F. Evolution and tinkering. *Science* **196**, 1161–1166 (1977).
2. Holland, P. W. Unity or diversity? *Trends Genet.* **12**, 324 (1996).
3. Stern, D. L. Perspective: Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
4. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, Malden, 2004).
5. McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* **37**, 403–408 (1984).
6. Halder, G., Callaerts, P. & Gehring, W. J. Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science* **267**, 1788–1792 (1995).
7. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
8. Jeffroy, O., Brinkmann, H., Delsuc, F. & Philippe, H. Phylogenomics: the beginning of incongruence? *Trends Genet.* **22**, 225–231 (2006).
9. Delsuc, F., Brinkmann, H. & Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nature Rev. Genet.* **6**, 361–375 (2005).
10. Rokas, A. & Carroll, S. B. Bushes in the tree of life. *PLoS Biol.* **4**, e352 (2006).
11. Rodriguez-Ezpeleta, N. *et al.* Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst. Biol.* **56**, 389–399 (2007).
This paper describes methods to alleviate inconsistencies that sometimes arise in phylogenomic analysis.
12. Brusca, R. C. & Brusca, G. J. *Invertebrates* (Sinauer Associates, Sunderland, 2003).
13. Jimenez-Guri, E., Philippe, H., Okamura, B. & Holland, P. W. *Buddenbrockia* is a cnidarian worm. *Science* **317**, 116–118 (2007).
14. Bourlat, S. J. *et al.* Deuterostome phylogeny reveals monophyletic chordates and the new phylum *Xenoturbellida*. *Nature* **444**, 85–88 (2006).
15. Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006).
16. 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.* **8**, 370–377 (2006).
17. Oda, H. *et al.* A novel amphioxus cadherin that localizes to epithelial adherens junctions has an unusual domain organization with implications for chordate phylogeny. *Evol. Dev.* **4**, 426–434 (2002).
18. Jeffery, W. R., Strickler, A. G. & Yamamoto, Y. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. *Nature* **431**, 696–699 (2004).
19. Bassham, S. & Postlethwait, J. H. The evolutionary history of placodes: a molecular genetic investigation of the larvacean urochordate *Oikopleura dioica*. *Development* **132**, 4259–4272 (2005).
20. Mazet, F. *et al.* Molecular evidence from *Ciona intestinalis* for the evolutionary origin of vertebrate sensory placodes. *Dev. Biol.* **282**, 494–508 (2005).
21. Takahashi, T. & Holland, P. W. Amphioxus and ascidian *Dmbx* homeobox genes give clues to the vertebrate origins of midbrain development. *Development* **131**, 3285–3294 (2004).
22. Cañestro, C., Bassham, S. & Postlethwait, J. Development of the central nervous system in the larvacean *Oikopleura dioica* and the evolution of the chordate brain. *Dev. Biol.* **285**, 298–315 (2005).
23. Dufour, H. D. *et al.* Precraniate origin of cranial motoneurons. *Proc. Natl Acad. Sci. USA* **103**, 8727–8732 (2006).
24. Cañestro, C. & Postlethwait, J. H. Development of a chordate anterior–posterior axis without classical retinoic acid signaling. *Dev. Biol.* **305**, 522–538 (2007).
25. Holland, L. Z. Developmental biology: a chordate with a difference. *Nature* **447**, 153–155 (2007).
This Review discusses how evolutionary change in patterning mechanisms can occur without altering basic body plans.
26. Li, J. B. *et al.* Comparative genomics identifies a flagellar and basal body proteome that includes the *BBS5* human disease gene. *Cell* **117**, 541–552 (2004).
27. Avidor-Reiss, T. *et al.* Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell* **117**, 527–539 (2004).
The authors of references 26 and 27 used comparative genomics to identify genes in organisms that have cilia or flagella, but that are absent from organisms that lack these organelles, and then verified the gene set in *D. melanogaster* and *C. elegans*.
28. Pazour, G. J. & Witman, G. B. The vertebrate primary cilium is a sensory organelle. *Curr. Opin. Cell Biol.* **15**, 105–110 (2003).
29. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
30. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396–398 (2007).
31. Urieli-Shoval, S., Gruenbaum, Y., Sedat, J. & Razin, A. The absence of detectable methylated bases in *Drosophila melanogaster* DNA. *FEBS Lett.* **146**, 148–152 (1982).
32. Gutierrez, A. & Sommer, R. J. Evolution of *dnmt-2* and *mbd-2*-like genes in the free-living nematodes *Pristionchus pacificus, Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Nucleic Acids Res.* **32**, 6388–6396 (2004).
33. Regev, A., Lamb, M. & Jablonka, E. The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Mol. Biol. Evol.* **15**, 880–891 (1998).
34. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genet.* **39**, 61–69 (2007).
The described project mapped DNA methylation in the entire *Arabidopsis thaliana* genome and found that moderately transcribed genes are most likely to be methylated, whereas genes at either extreme are least likely, revealing a genome-wide association of methylation and transcription.
35. Suzuki, M. M., Kerr, A. R., De Sousa, D. & Bird, A. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res.* **17**, 625–631 (2007).
36. Robertson, K. D. DNA methylation and human disease. *Nature Rev. Genet.* **6**, 597–610 (2005).
37. Jeltsch, A., Nellen, W. & Lyko, F. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem. Sci.* **31**, 306–308 (2006).
38. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Rev. Genet.* **2**, 292–301 (2001).
39. Mita, K. *et al.* The genome sequence of silkworm, *Bombyx mori*. *DNA Res.* **11**, 27–35 (2004).
40. Marhold, J. *et al.* Conservation of DNA methylation in dipteran insects. *Insect Mol. Biol.* **13**, 117–123 (2004).
41. Wang, Y. *et al.* Functional CpG methylation system in a social insect. *Science* **314**, 645–647 (2006).
The authors revealed for the first time the full set of *Dnmt* genes in a protostome.
42. Smallwood, A., Esteve, P. O., Pradhan, S. & Carey, M. Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev.* **21**, 1169–1178 (2007).
43. Eskeland, R., Eberharter, A. & Imhof, A. HP1 binding to chromatin methylated at H3K9 is enhanced by auxiliary factors. *Mol. Cell. Biol.* **27**, 453–465 (2007).
44. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
45. Seo, H. C. *et al.* Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* **431**, 67–71 (2004).
46. Edvardsen, R. B. *et al.* Hypervariable and highly divergent intron–exon organizations in the chordate *Oikopleura dioica*. *J. Mol. Evol.* **59**, 448–457 (2004).
47. Spada, F. *et al.* Conserved patterns of nuclear compartmentalization are not observed in the chordate *Oikopleura*. *Biol. Cell* **99**, 273–287 (2007).
48. Garcia-Fernandez, J. The genesis and evolution of homeobox gene clusters. *Nature Rev. Genet.* **6**, 881–892 (2005).
49. Duboule, D. The rise and fall of Hox gene clusters. *Development* **134**, 2549–2560 (2007).
This analysis raises the question of whether vertebrate Hox clusters might be 'better organized' than ancestral clusters.
50. Monteiro, A. S. & Ferrier, D. E. Hox genes are not always Colinear. *Int. J. Biol. Sci.* **2**, 95–103 (2006).
51. Krumlauf, R. Hox genes in vertebrate development. *Cell* **78**, 191–201 (1994).
52. Holland, L. Z. & Holland, N. D. Expression of *AmphiHox-1* and *AmphiPax-1* in amphioxus embryos treated with retinoic acid: insights into evolution and patterning of the chordate nerve cord and pharynx. *Development* **122**, 1829–1838 (1996).
53. Bickmore, W. A., Mahy, N. L. & Chambeyron, S. Do higher-order chromatin structure and nuclear reorganization play a role in regulating Hox gene expression during development? *Cold Spring Harb. Symp. Quant. Biol.* **69**, 251–257 (2004).
54. Kmita, M. & Duboule, D. Organizing axes in time and space; 25 years of colinear tinkering. *Science* **301**, 331–333 (2003).
55. Chambeyron, S. & Bickmore, W. A. Chromatin decondensation and nuclear reorganization of the *HoxB* locus upon induction of transcription. *Genes Dev.* **18**, 1119–1130 (2004).
56. Carbone, R. *et al.* Recruitment of the histone methyltransferase SUV39H1 and its role in the oncogenic properties of the leukemia-associated PML-retinoic acid receptor fusion protein. *Mol. Cell. Biol.* **26**, 1288–1296 (2006).
57. Mark, M., Ghyselinck, N. B. & Chambon, P. Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. *Annu. Rev. Pharmacol. Toxicol.* **46**, 451–480 (2006).
58. Ross, S. A., McCaffery, P. J., Drager, U. C. & De Luca, L. M. Retinoids in embryonal development. *Physiol. Rev.* **80**, 1021–1054 (2000).
59. Fujiwara, S. Retinoids and nonvertebrate chordate development. *J. Neurobiol.* **66**, 645–652 (2006).
60. Shimeld, S. M. Retinoic acid, Hox genes and the anterior–posterior axis in chordates. *BioEssays* **18**, 613–616 (1996).
61. Cañestro, C., Postlethwait, J., Gonzalez-Duarte, R. & Albalat, R. Is retinoic acid genetic machinery a chordate innovation? *Evol. Dev.* **8**, 394–406 (2006).
62. Marletaz, F., Holland, L. Z., Laudet, V. & Schubert, M. Retinoic acid signaling and the evolution of chordates. *Int. J. Biol. Sci.* **2**, 38–47 (2006).
63. Simoes-Costa, M. S., Azambuja, A. P. & Xavier-Neto, J. The search for non-chordate retinoic acid signaling: lessons from chordates. *J. Exp. Zoolog. B Mol. Dev. Evol.* 15 Nov 2006 (doi:10.1002/jez.b.21139).
64. Lowe, C. J. *et al.* Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* **113**, 853–865 (2003).
65. Hinman, V. F. & Degnan, B. M. Retinoic acid disrupts anterior ectodermal and endodermal development in ascidian larvae and postlarvae. *Dev. Genes Evol.* **208**, 336–345 (1998).
66. Kawamura, K., Hara, S. & Fujiwara, S. Developmental role of endogenous retinoids in the determination of morphallactic field in budding tunicates. *Development* **117**, 835–845 (1993).
67. Rinkevich, Y., Paz, G., Rinkevich, B. & Reshef, R. Systemic bud induction and retinoic acid signaling underlie whole body regeneration in the urochordate *Botrylloides leachi*. *PLoS Biol.* **5**, e71 (2007).
68. Ikuta, T., Yoshida, N., Satoh, N. & Saiga, H. *Ciona intestinalis* Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc. Natl Acad. Sci. USA* **101**, 15118–15123 (2004).
69. Ohno, S. *Evolution by Gene Duplication* (Springer-Verlag, New York, 1970).
70. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
This work provides the theory behind the preservation of gene duplicates by reciprocal loss of gene subfunctions and provides data to support the model.
71. Hughes, A. L. *Adaptive Evolution of Genes and Genomes* (Oxford Univ. Press, New York, 1999).
72. Lynch, M. & Force, A. The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* **156**, 590–605 (2000).
73. Postlethwait, J., Amores, A., Cresko, W., Singer, A. & Yan, Y. L. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* **20**, 481–490 (2004).
74. Postlethwait, J. H. The zebrafish genome in context: ohnologs gone missing. *J. Exp. Zoolog. B Mol. Dev. Evol.* **308B**, 563–577 (2006).

75. Amores, A. *et al.* Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**, 1711–1714 (1998).
76. Postlethwait, J. *et al.* Vertebrate genome evolution and the zebrafish gene map. *Nature Genet.* **18**, 345–349 (1998).
77. Jaillon, O. *et al.* Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957 (2004).
78. Dehal, P. & Boore, J. L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314 (2005).
Whole-genome analysis of paralogy groups in the human genome is used to test predictions of the idea that two rounds of whole-genome duplication occurred at the base of vertebrate phylogeny.
79. Kasahara, M. *et al.* The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447**, 714–719 (2007).
80. Thisse, B. & Thisse, C. Functions and regulations of fibroblast growth factor signaling during embryonic development. *Dev. Biol.* **287**, 390–402 (2005).
81. Itoh, N. & Ornitz, D. M. Evolution of the *Fgf* and *Fgfr* gene families. *Trends Genet.* **20**, 563–569 (2004).
82. Popovici, C., Roubin, R., Coulier, F. & Birnbaum, D. An evolutionary history of the FGF superfamily. *Bioessays* **27**, 849–857 (2005).
83. Satou, Y., Imai, K. S. & Satoh, N. *Fgf* genes in the basal chordate *Ciona intestinalis*. *Dev. Genes Evol.* **212**, 432–438 (2002).
84. Matus, D. Q., Thomsen, G. H. & Martindale, M. Q. FGF signaling in gastrulation and neural development in *Nematostella vectensis*, an anthozoan cnidarian. *Dev. Genes Evol.* **217**, 137–148 (2007).
85. Kusserow, A. *et al.* Unexpected complexity of the *Wnt* gene family in a sea anemone. *Nature* **433**, 156–160 (2005).
86. Reifers, F., Adams, J., Mason, I. J., Schulte-Merker, S. & Brand, M. Overlapping and distinct functions provided by *fgf17*, a new zebrafish member of the Fgf8/17/18 subgroup of Fgfs. *Mech. Dev.* **99**, 39–49 (2000).
87. Jovelin, R. *et al.* Duplication and divergence of *fgf8* functions in the development and evolution of ray fin fish. *J. Exp. Zoolog. B Mol. Dev. Evol.* 20 August 2007 (doi:10.1002/jez.b.21193).
88. Furthauer, M., Thisse, C. & Thisse, B. A role for FGF-8 in the dorsoventral patterning of the zebrafish gastrula. *Development* **124**, 4253–4264 (1997).
89. Cao, Y. *et al. fgf17b*, a novel member of Fgf family, helps patterning zebrafish embryos. *Dev. Biol.* **271**, 130–143 (2004).
90. Draper, B. W., Stock, D. W. & Kimmel, C. B. Zebrafish *fgf24* functions with *fgf8* to promote posterior mesodermal development. *Development* **130**, 4639–4654 (2003).
91. Abu-Issa, R., Smyth, G., Smoak, I., Yamamura, K. & Meyers, E. N. *Fgf8* is required for pharyngeal arch and cardiovascular development in the mouse. *Development* **129**, 4613–4625 (2002).
92. Yokoi, H. *et al.* Mutant analyses reveal different functions of *fgfr1* in medaka and zebrafish despite conserved ligand-receptor relationships. *Dev. Biol.* **304**, 326–337 (2007).
93. Lewandoski, M., Sun, X. & Martin, G. R. *Fgf8* signalling from the AER is essential for normal limb development. *Nature Genet.* **26**, 460–463 (2000).
94. Moon, A. M. & Capecchi, M. R. *Fgf8* is required for outgrowth and patterning of the limbs. *Nature Genet.* **26**, 455–459 (2000).
95. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
96. Woolfe, A. *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7 (2005).
97. Simpson, P. The stars and stripes of animal bodies: evolution of regulatory elements mediating pigment and bristle patterns in *Drosophila*. *Trends Genet.* **23**, 350–358 (2007).
98. Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L. & McCallion, A. S. Conservation of *RET* regulatory function from human to zebrafish without sequence similarity. *Science* **312**, 276–279 (2006).
99. Langham, R. J. *et al.* Genomic duplication, fractionation and the origin of regulatory novelty. *Genetics* **166**, 935–945 (2004).
100. Woolfe, A. & Elgar, G. Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. *Genome Biol.* **8**, R53 (2007).
The authors compared vertebrate genomes to identify conserved non-coding elements associated with genes that duplicated in the teleost genome duplication event, and showed that they were distributed as predicted by the reciprocal loss of ancestral gene subfunctions.
101. Allende, M. L., Manzanares, M., Tena, J. J., Feijoo, C. G. & Gomez-Skarmeta, J. L. Cracking the genome's second code: enhancer detection by combined phylogenetic footprinting and transgenic fish and frog embryos. *Methods* **39**, 212–219 (2006).
102. McEwen, G. K. *et al.* Ancient duplicated conserved noncoding elements in vertebrates: a genomic and functional analysis. *Genome Res.* **16**, 451–465 (2006).
103. McClintock, J. M., Kheirbek, M. A. & Prince, V. E. Knockdown of duplicated zebrafish *hoxb1* genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. *Development* **129**, 2339–2354 (2002).
104. Kikuta, H. *et al.* Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.* **17**, 545–555 (2007).
Reporter insertions into zebrafish chromosomes identified long-range conserved regulatory elements that might be responsible for the conservation of chromosome segments over vertebrate evolution.
105. Everman, D. B. *et al.* Frequency of genomic rearrangements involving the *SHFM3* locus at chromosome 10q24 in syndromic and non-syndromic split-hand/foot malformation. *Am. J. Med. Genet. A* **140**, 1375–1383 (2006).
106. Goode, D. K., Snell, P., Smith, S. F., Cooke, J. E. & Elgar, G. Highly conserved regulatory elements around the *SHH* gene may contribute to the maintenance of conserved synteny across human chromosome 7q36.3. *Genomics* **86**, 172–181 (2005).

Acknowledgements
We thank for support R01 RR020833, and P01 HD22486 from the National Institutes of Health, and IOB-0719577, from the US National Science Foundation.

DATABASES
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
BBS5 | *EN2* | *fbxw4* | *HOXA1* | *hoxb1* | *PAX2* | *PAX5* | *PAX6* | *PAX8*
OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Bardet–Biedl syndrome | split-hand/foot malformation

FURTHER INFORMATION
***Branchiostoma floridae* genome:** http://genome.jgi-psf.org/Brafl1/Brafl1.home.html
***Ciona intestinalis* genome:** http://genome.jgi-psf.org/Cioin2/Cioin2.home.html
Ensembl: http://www.ensembl.org
John H. Postlethwait's laboratory: http://www.neuro.uoregon.edu/postle/PostleLab.html
***Nematostella vectensis* genome:** http://genome.jgi-psf.org/Nemve1/Nemve1.home.html
***Oikopleura dioica* genome:** http://www.ncbi.nlm.nih.gov/Traces/trace.cgi
***Strongylocentrotus purpuratus* genome:** http://www.ncbi.nlm.nih.gov/projects/genome/guide/sea_urchin/
UTGB medaka: http://medaka.utgenome.org
Vista: http://genome.lbl.gov/vista/index.shtml

ALL LINKS ARE ACTIVE IN THE ONLINE PDF