



Evolution of homeobox genes

Peter W. H. Holland*

Many homeobox genes encode transcription factors with regulatory roles in animal and plant development. Homeobox genes are found in almost all eukaryotes, and have diversified into 11 gene classes and over 100 gene families in animal evolution, and 10 to 14 gene classes in plants. The largest group in animals is the ANTP class which includes the well-known Hox genes, plus other genes implicated in development including ParaHox (*Cdx*, *Xlox*, *Gsx*), *Evx*, *Dlx*, *En*, *NK4*, *NK3*, *Msx*, and *Nanog*. Genomic data suggest that the ANTP class diversified by extensive tandem duplication to generate a large array of genes, including an NK gene cluster and a hypothetical ProtoHox gene cluster that duplicated to generate Hox and ParaHox genes. Expression and functional data suggest that NK, Hox, and ParaHox gene clusters acquired distinct roles in patterning the mesoderm, nervous system, and gut. The PRD class is also diverse and includes *Pax2/5/8*, *Pax3/7*, *Pax4/6*, *Gsc*, *Hesx*, *Otx*, *Otp*, and *Pitx* genes. PRD genes are not generally arranged in ancient genomic clusters, although the Dux, Obox, and Rhox gene clusters arose in mammalian evolution as did several non-clustered PRD genes. Tandem duplication and genome duplication expanded the number of homeobox genes, possibly contributing to the evolution of developmental complexity, but homeobox gene loss must not be ignored. Evolutionary changes to homeobox gene expression have also been documented, including Hox gene expression patterns shifting in concert with segmental diversification in vertebrates and crustaceans, and deletion of a *Pitx1* gene enhancer in pelvic-reduced sticklebacks.

© 2012 Wiley Periodicals, Inc.

How to cite this article:

WIREs Dev Biol 2013, 2:31–45. doi: 10.1002/wdev.78

INTRODUCTION

The study of animal developmental biology, whether focused on morphogenesis, patterning, or differentiation, on limbs, brain, guts, muscle, blood, skeletons, or immune systems, rarely seems to escape the reach of homeobox genes. These genes are defined by presence of a (variable) DNA sequence, the homeobox, that if translated encodes a peptide motif with a recognizable helix-loop-helix-turn-helix structure: the homeodomain.¹ Most homeobox genes encode transcription factors that act through sequence-specific DNA-binding (mediated by the homeodomain) and interaction with protein cofactors to regulate expression of other genes, thereby effecting changes in cell behavior or activity. It is not always realized, however, just how large

and diverse is the homeobox gene superclass. The human, mouse, *Xenopus*, and zebrafish genomes, for example, each have well over 200 homeobox genes, and insects, nematodes, and amphioxus each have around 100 homeobox genes^{2,3} (Figure 1; data from HomeoDB2). Some flowering plants also have large numbers of homeobox genes, over 100 in *Arabidopsis thaliana*, although most of these genes are very distinct from those found in animals.⁴ Fungi and unicellular eukaryotes have far smaller numbers, typically less than 10. Fungal homeobox genes are also implicated in cellular or developmental decisions^{1,4} for example, the mating type loci of baker's yeast *Saccharomyces cerevisiae* are homeobox genes.^{5,6}

SUPERCLASSES, CLASSES, AND GENE FAMILIES

The term 'gene family' is sometimes used in a loose way to refer to a group of related genes. The problem with this usage is that it gives no impression of the

Conflict of interest: The author declares that he has no conflicts of interest.

*Correspondence to: peter.holland@zoo.ox.ac.uk

Department of Zoology, University of Oxford, OX1 3PS, UK

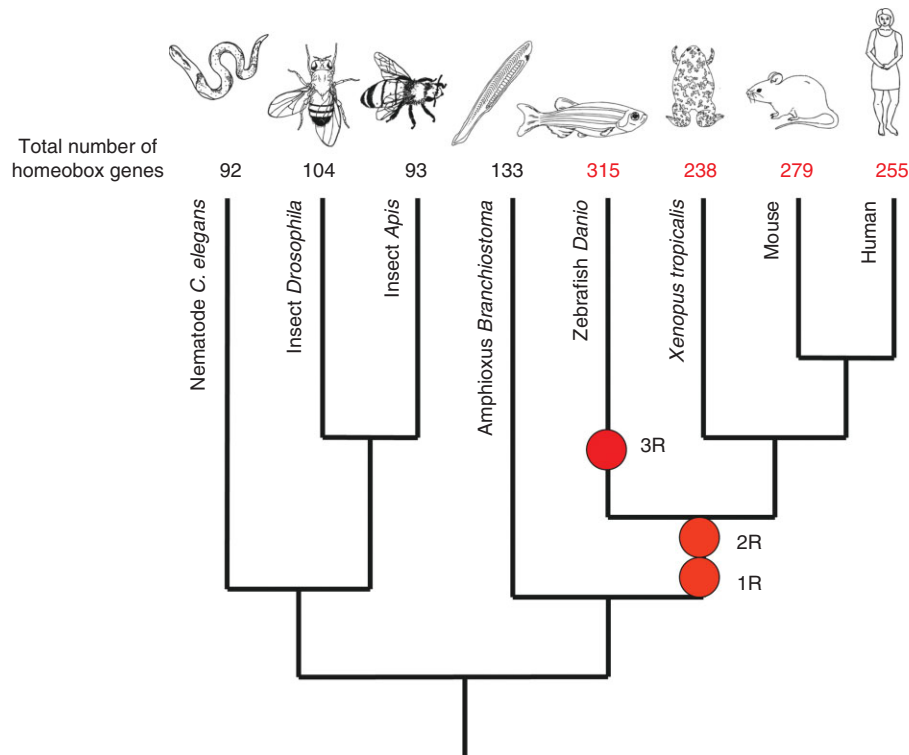


FIGURE 1 | Evolutionary tree showing total number of homeobox genes present in the genomes of selected animal species. The precise numbers are liable to change slightly with each release of a revised genome assembly; these data from HomeoDB2 accessed December 2011³ 1R, 2R, and 3R denote whole genome duplication events. (Animal diagrams by Tatiana Solovieva)

nested nature of evolutionary relationships. To give greater clarity, three levels of grouping are often used in the classification of homeobox genes: superclass, class, and gene family. Sometimes intermediate terms, such as subclass and subfamily are added, but these are not used consistently.

Every gene with a homeobox sequence is a member of the ‘homeobox gene superclass’. It does not matter how divergent is the homeobox sequence, nor what its role is, because classification reflects shared evolutionary history and not a particular biochemical function or developmental role. For example, the *Drosophila* gene *bicoid* encodes a homeodomain protein that binds RNA as well as DNA.^{7,8} Similarly, the human genes *CERS2* to *CERS6* have divergent homeoboxes, but encode transmembrane proteins and so their homeodomains are unlikely to act as DNA-binding motifs.^{9,10} Genes in the Pax-258 gene family (such as human *PAX2*, *PAX5*, and *PAX8*) have only a partial homeobox.

Below the level of superclass, a ‘class’ denotes a set of genes that share additional motifs or a set of genes that clearly fall together in an evolutionary tree. Thus, the ZF class denotes homeobox genes that also encode zinc-finger motifs, LIM class homeobox genes encode homeodomains plus LIM

domains, and the CUT and SINE classes have their own particular motifs. The CERS class encodes a small group of transmembrane proteins with divergent homeodomains, while the HNF class also has exceedingly divergent and extended homeobox sequences. The PROS class, represented by just a single gene in *Drosophila* and two in humans, has a different, extended homeodomain. The TALE (Three Amino Acid Loop Extension) class genes encode proteins, with three extra amino acids, between alpha helices 1 and 2 of the homeodomain.¹¹ The POU class has a conserved approximately 75 amino acid region immediately N-terminal to the 60 amino acid homeodomain.

In summary, the homeobox genes of animal genomes can be divided into 11 classes: ANTP, PRD, TALE, POU, CERS, PROS, ZF, LIM, HNF, CUT, and SINE.¹² A few animal genes have been found that cannot be placed readily into these categories. Homeobox genes of plants have been classified into 14 classes: HD-ZIP I, HD-ZIP II, HD-ZIP III, HD-ZIP IV, KNOX, BEL, PLINC, WOX, DDT, PHD, NDX, LD, PINTOX, and SAWADEE.⁴ There is an argument for merging the first four plant classes into one HD-ZIP class, while KNOX and BEL are both subtypes of TALE gene so they could be merged as

one gene class. Since TALE genes are found in animals and plants, there is an argument for erecting a higher level of classification for these genes.

Classes are subdivided into gene families, with over 100 homeobox gene families described. In animals, homeobox genes are placed in the same gene family if they all descend from a single gene in the long-extinct common ancestor of *Drosophila* (an ecdysozoan protostome) and human (a deuterostome vertebrate), the animal sometimes referred to as the 'urbilaterian'.¹² This definition works reasonably, but it is not perfect. Under this rule, it is relatively simple to define many homeobox gene families, such as the *Msx*, *En*, *Emx*, *Otx*, *Tlx*, *Gsc*, *Otp*, *Cdx*, and *Gsx* gene families. For example, the *Drosophila* genes *en* and *inv*, plus human *EN1* and *EN2*, fall into the *En* gene family, because they all descend from a single gene that underwent independent gene duplication in the two evolutionary lineages.¹³ Similarly, the *Drosophila* gene *otd*, plus the human genes *OTX1*, *OTX2*, and *CRX*, fall in the *Otx* gene family, regardless of the fact that the name *CRX* is a misnomer.¹⁴ In contrast, Hox genes do not comprise a single gene family, since the common ancestor of bilaterian animals already had several Hox genes.^{15,16} One problem is that in a few cases it is difficult to determine the exact evolutionary history of a set of genes. Thus, the exact number of 'middle' Hox genes present in the common ancestor of bilaterian animals is hard to resolve, as is the number of ancestral *Dlx* genes.^{15,17} Compromise is needed in these cases. A second problem concerns apparent 'orphan' genes found in one evolutionary lineage but not others, yet whose origin is obscure, such as *Nanog* in vertebrates.¹⁸ A third problem is that this definition is not easily extended to non-bilaterian animals, such as sea anemones and sponges, nor of course to non-animals. At least for sea anemones the bilaterian classifications work reasonably well in most cases.¹⁹

EARLY EVOLUTION OF HOMEBOX GENES

Homeobox genes have been found in most eukaryote genomes examined, but not in bacteria or Archaea. However, immediately after discovery of the homeobox, it was noted that the encoded homeodomain peptide motif has a similar fold to the bacterial helix-turn-helix proteins.²⁰ It seems likely, therefore, that the homeobox genes evolved from ancestral helix-turn-helix genes. The earliest steps of homeobox gene diversification, however, are difficult to reconstruct. TALE class homeobox genes are present in animals, plants, and unicellular eukaryotes such as green algae and the amoeboid

opisthokont *Capsaspora*.^{4,21} Furthermore, plant and animal TALE class homeodomain proteins have sufficient sequence similarity to be confident that these have a common origin.^{4,21–23} There are also 'typical' 60-amino acid homeodomains in plants, animals, and other eukaryotes, although in this case there is less evidence for homology between gene classes.^{4,5} On balance, ancestral eukaryotes most likely had one (or more) TALE and one (or more) non-TALE homeobox genes. Subsequent gene duplication followed independent routes in animal and plant evolution. The homeodomain superfamily has not been maintained in every eukaryotic lineage, and some parasitic groups and intracellular symbionts have lost homeobox genes from their genome.⁵

THE ANTP CLASS: EVOLUTION OF HOX GENES, PARAHOX GENES, AND THEIR RELATIVES

The ANTP class is the largest class of animal homeobox genes, with around 100 genes in the human genome and similar numbers in other vertebrates.^{3,12} There are no ANTP class genes outside the Metazoa. ANTP class genes are divided into almost 50 gene families (37 in human), the majority of which date back to single genes in the ancestral urbilaterian.^{2,24} These include several families of Hox genes, three ParaHox gene families, various NK (and NK-related) homeobox gene families, the *Dlx* genes, and many more. When, and how, did the great diversity of ANTP class homeobox gene families arise?

There is no doubt the extensive tandem gene duplication of Hox genes to form a Hox gene cluster of perhaps 8 to 10 genes, had already occurred by the time of the common ancestor of flies and vertebrates.¹⁵ But is there a Hox gene cluster in more divergent, non-bilaterian animals? Answering this question turned out to be more complex than expected. One problem is that several other homeobox genes, not part of the Hox gene clusters, have homeobox sequences that are very similar to those of Hox genes. These are the *Gsx*, *Xlox* (=Pdx), *Cdx*, *Mox*, and *Evx* gene families, and less conclusively the *Mnx*, *Ro*, *Gbx*, and *En* gene families. Indeed, the homeodomains encoded by *Gsx*, *Xlox*, and *Cdx* are more similar to Hox genes than some Hox genes are to each other. This conundrum was clarified when it was discovered that *Gsx*, *Xlox*, and *Cdx* genes are arranged into a distinct gene cluster in amphioxus, and indeed in several vertebrate species.²⁵ The name given to this cluster, ParaHox, reflects a proposed 'sister' relationship with the Hox gene cluster, with the two gene clusters (Hox and ParaHox) hypothesized to

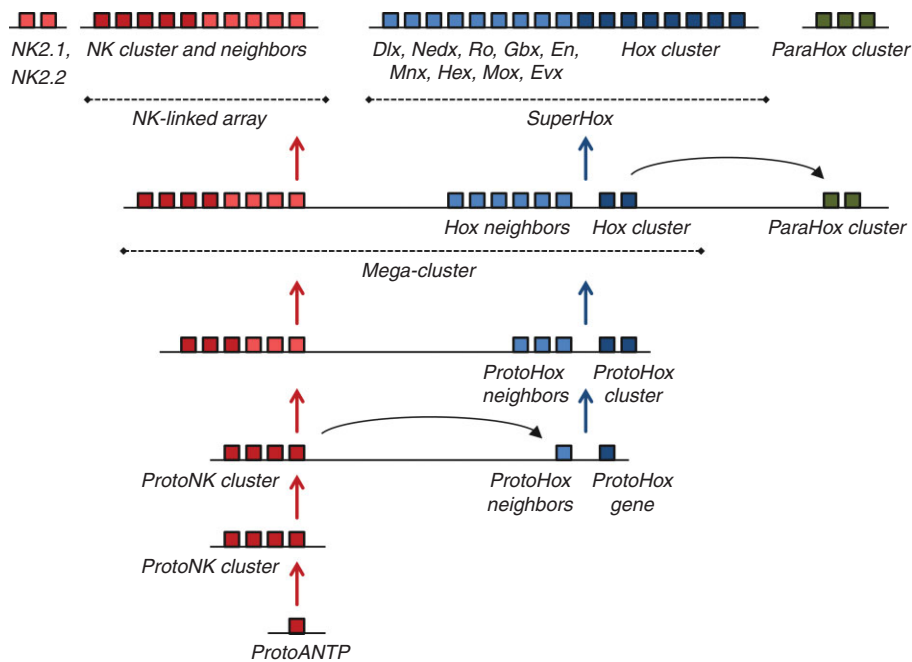


FIGURE 2 | Possible evolutionary history of the ANTP class homeobox genes, involving extensive tandem duplication to generate NK genes (red) and later ProtoHox genes (blue). Duplication of a hypothetical ProtoHox gene cluster (shown here as two genes, though the precise number is unresolved) gave distinct Hox and ParaHox gene clusters (blue and green, respectively).

have been generated by duplication from an ancestral ‘ProtoHox’ gene cluster (Figure 2).

There has been much debate and discussion as to the number of genes in the hypothetical ProtoHox cluster (one, two, three, or four?). These arguments center on the topology and robustness of phylogenetic trees built from homeodomain proteins.^{25–27} The fundamental problem with all these arguments, however, is that assumptions made by molecular phylogenetic algorithms might have been seriously compromised during the evolution of these genes. For example, if following gene duplication one daughter gene diverges in sequence radically, while the other retains essentially the ancestral sequence, this is rarely revealed by molecular phylogenetics and thus trees may not reflect reality. In addition, tandem duplication could generate genes of apparently ‘chimaeric’ nature, which would further compromise precise phylogenetic reconstruction.²⁸ Regardless of details, Hox and ParaHox genes are evolutionarily closely related.

Solving the origin of the Hox gene cluster is dependent on dating of the Hox/ParaHox duplication. Following several years of confusion, the situation in the four non-bilateria phyla (Cnidaria, Porifera, Ctenophora, and Placozoa) is now becoming clearer. Cnidarians, such as sea anemones and *Hydra*, have several Hox-like homeobox genes and careful genome synteny analyses in *Nematostella vectensis*

have shown that these include a small cluster of definitive Hox genes^{19,29} and a separate pair of ParaHox genes.³⁰ The two ParaHox genes of *Nematostella* had been interpreted as an orthologue of Gsx (called *Anthox-2* or *Cnox-2*) and a putative Gsx/Xlox precursor *NvHD065*,²⁶ but the latter interpretation is challenged by finding of a true Xlox gene in hydrozoan cnidarians.³¹ The data indicate that distinct ParaHox and Hox gene clusters were present in the common ancestor of cnidarians and bilaterians. As stressed by Kamm and Schierwater^{32,33}, this does not necessarily imply that the functions are the same in cnidarians and bilaterians. In contrast to cnidarians, no Hox-like genes have been found in the genomes of either poriferans (sponges) or ctenophores (comb jellies),^{34,35} implying that some animals build their bodies without Hox genes. Assuming that animals in these phyla have not lost these genes in evolution, Hox genes originated quite early in animal diversification, though not right at the base. The placozoan *Trichoplax adharens* has what might be an intermediate condition. In this animal just a single Hox-like gene, *Trox-2*, has been found, most similar in sequence to bilaterian Gsx and cnidarian *Cnox-2* genes.^{36–38} Whether *Trox-2* represents a relict ‘ProtoHox’ gene is a tantalizing possibility. *Trox-2* has a striking expression pattern in a ring of dividing cells close to the periphery of the animal, and functional assays suggest it plays a role in growth of the animal.³⁹

In many bilaterian genomes, including those of mouse, human, amphioxus, and annelids, Hox genes are found close chromosomally to other ANTP class genes, notably the *Evx* (*eve*) and *Dlx* (*distalless*) genes, plus in some species *Ro*, *Mnx*, *En*, *Mox*, *Hex*, *Nedx*, and/or *Gbx*.⁴⁰ Other ANTP class genes also form chromosomal arrays, with the largest being the so-called 'NK homeobox gene cluster' or 'MetaHox gene cluster',^{41–43} including genes of the NK4 (= *Csx* or *tin*), NK3 (= *bap*), NK1, *Tlx*, *Lbx*, *Msx*, NK5 (= *Hmx*), NK6, and NK7 genes. These linked arrays of genes, the Hox-linked array and the NK-linked array, though variable between species and not always retained in tight clusters, are remnants of the evolution of the ANTP class genes by extensive tandem gene duplication (Figure 2). Furthermore, these events can be dated to very early in pre-bilaterian animal evolution because sponges and comb jellies, although lacking Hox/ParaHox genes, have many genes typical of the NK-linked array, as does *Trichoplax*.^{34,36,38} A picture emerges of extensive expansion of the ANTP class of homeobox genes by gene duplication early in animal evolution. This expansion spawned the NK-linked genes, then later the Hox-linked genes (including true Hox genes), and the ParaHox genes (Figure 2).

In extant bilaterian genomes, such as those of amphioxus and polychaetes, it is usual to find all these ANTP class genes at four major chromosomal sites: one containing the Hox-linked genes, one comprising the ParaHox genes (*Gsx*, *Xlox*, and *Cdx*), one including the NK-linked genes, and a fourth location comprising a pair of unusual NK genes denoted *NK2.1* and *NK2.2*.^{44–46} (Figure 2). In many species these four genomic locations have been disrupted by gene loss and rearrangement, while in vertebrates each genomic region has been quadruplicated. The ancestral set of homeobox genes that gave rise to the Hox gene cluster and its chromosomal neighbors has been called the 'SuperHox' gene cluster,⁴⁰ while the hypothetical ancestral array of ANTP class genes including Hox- and NK-linked arrays has been dubbed the 'Mega-cluster'.^{45–47}

Hox, ParaHox, NK: The Germ-Layer Hypothesis

Studies on the evolution of ANTP class genes have suggested an intriguing link with the evolution of body plans. Hox genes have been implicated in specifying position along the anteroposterior axis of many bilaterian animals. In vertebrates, mutant analysis and gene expression patterns suggest that these roles apply to the central nervous system, the peripheral nervous system, the notochord, the vertebral column, and the visceral mesoderm surrounding organ

systems.^{48–51} In short, the principal roles of vertebrate Hox genes are in ectoderm and in mesoderm. A similar ectoderm/mesoderm patterning role is seen in *Drosophila*^{52,53} apart from a specialized role of one gene in one midgut endodermal cell type.⁵⁴ In many other groups of animals, including amphioxus, ascidians, annelids, hemichordates, and molluscs, the principal tissues displaying anteroposterior-nested Hox gene expression are only ectodermal, particularly neural.^{55,56} The ancestral role of Hox genes, at least in bilaterians, was most likely to specify or encode positional information along the anteroposterior axis of *ectoderm*.^{55–57} In some groups, such as vertebrates and insects, a mesodermal role was added later.

In contrast the ParaHox gene cluster, the proposed sister of the Hox cluster, seems to be predominantly involved in *endoderm* patterning, or to be embryologically correct *gut*-patterning since in many animals the extreme anterior and posterior parts of the gut tube are not strictly endodermal.^{25,57,58} The most 'posterior' of the three ParaHox genes, *Cdx*, marks the anus region of many animals, including vertebrates, amphioxus, annelids, molluscs, and (for its zygotic role) *Drosophila*. The 'central' ParaHox gene, *Xlox* or *Pdx*, is expressed and functional in the endoderm of the central gut, around the pancreas/duodenal region of vertebrates,^{59,60} the homologous region in amphioxus and in a central gut region of at least some molluscs and annelids. The gene has been lost from insects and nematodes. Are these two genes revealing an underlying 'rule' about ParaHox gene activity? If so, then one might expect the 'anterior' ParaHox gene, *Gsx*, to be expressed in the mouth. The situation is not so simple, however, since in *Drosophila* the *Gsx* gene (called *ind*) is expressed along the ventral nerve cord, in vertebrates the duplicated *Gsx* genes have roles in the brain and the amphioxus ortholog is expressed in the cerebral vesicle.²⁵

Gsx may not have a 'gut-patterning' role in flies, amphioxus or vertebrates, but this does not exclude the possibility that the *ancestral* role of *Gsx* was to specify the mouth region, because gene functions can be modified in evolution. It is worth noting that evolutionary scenarios based on comparative embryology suggest that vertebrates (along with amphioxus, ascidians, echinoderms, and hemichordates) have lost the ancestral mouth region, evolving a 'second mouth' not homologous to the mouth of other animals, nor homologous to the ancestral bilaterian mouth (though alternatives to this hypothesis have been proposed⁶¹). This leaves open the possibility that the ancestral role of the three ParaHox genes in bilaterian animals was to specify mouth, midgut, and anus, and that in deuterostome

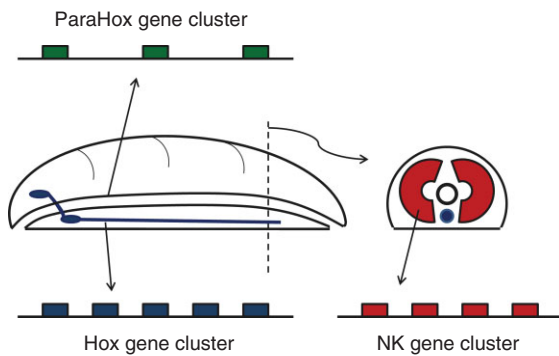


FIGURE 3 | Hypothetical role of Hox, ParaHox, and NK homeobox gene clusters in patterning of neural, gut, and mesodermal tissues in an early bilaterian animal. Anterior to the left.

animals (such as vertebrates and amphioxus) the role of *Gsx* in the mouth was lost when an ‘new mouth’ evolved.⁵⁸ Analysis of ParaHox gene expression in molluscs and annelids (which are protostomes, not deuterostomes) reveals clear *Gsx* expression in the developing mouth region, as predicted by this hypothesis.^{62,63}

In summary, in the earliest bilaterian animals the Hox genes probably had primary roles in specifying anteroposterior position along the centralized nerve cord while the ParaHox genes may have played a similar role in the developing gut (Figure 3). These may not have been their only roles, for example localized neural expression is also commonly seen for ParaHox genes. Nerve cord and gut are not the only two systems for which patterning is needed. The long-extinct early bilaterian animals, including the common ancestor of humans, fish, flies, worms, snails, and millions more species, were able to exploit the world actively in three-dimensions; with directed locomotion, an anterior brain, sense organs, and a forward-facing mouth. To effect active locomotion, whether it be burrowing, crawling, or swimming, mesodermal patterning was also required. Intriguingly, the third of the ANTP class gene clusters discussed above, the NK homeobox gene cluster, includes several genes with predominant expression in mesoderm.⁴² These include *Msx*, *NK4* (=C*sox* or *tin*), and *NK3* (=b*ap*) genes. Further comparative studies are needed, but it is a possibility that these genes played an ancient role in mesodermal patterning in the first bilaterian animals. This may not have been concerned with anteroposterior patterning of the mesoderm, and was perhaps a role in specifying different mesodermal functional domains, perhaps around the mediolateral axis of the animal.

This hypothesis proposes that early bilaterian animals used three sets of ANTP class homeobox genes to pattern three major embryonic systems: the nerve

cord, the gut, and the mesoderm, roughly (but not precisely) equivalent to three germ layers (Figure 3). These would not be the only genes involved, and evidence has been published in support of similarly distinct roles for different sets of Fox class transcription factors.^{64,65} It does not propose that the origin of distinct Hox, ParaHox, and NK sets of genes was the cause of the evolution of new tissues and structures, only that these genes were recruited to new tissues and body structures as they evolved. It also makes no statement about the role of these genes in non-bilaterian animals. For example, cnidarians, such as the sea anemone *Nematostella vectensis*, have distinct sets of genes homologous to Hox, ParaHox, and NK genes, but these animals do not have a central nerve cord, distinct mouth and anus, or extensive mesodermal differentiation. These genes may play different roles in cnidarians, as has been argued by others.³² Furthermore, the sponge *Amphimedon queenslandica* has clear NK genes, though not Hox or ParaHox genes, but this does not mean that sponges are mesodermal animals. Gene functions change in evolution.

THE PRD CLASS: METAZOAN DIVERSITY AND MAMMALIAN ADDITIONS

The PRD homeobox class, named after the *Drosophila* gene *paired*, includes the ‘Pax genes’ (genes possessing a ‘paired box’) that also contain a homeobox sequence, including the Pax2/5/8, Pax3/7, Pax4/6, and *eyg* gene families, but not the Pax1/9 family. The *poxn* genes lack a homeobox in bilaterians, although a cnidarian homolog possesses one.⁶⁶ In addition to Pax genes, I also include in the PRD class many related genes with a homeobox but not a paired box: approximately 40 gene families in bilaterian animals (31 families in human¹²). Examples include *Gsc* (*gooseoid*; implicated in gastrulation and mesoderm patterning), *Hesx* (involved in forebrain and pituitary development), *Otx* (an early anterior marker in flies and vertebrates), *Otp* (another brain development marker), *Dmbx* (a mid- and hindbrain expressed gene), and *Pitx* (including *Pitx1* involved in fin/limb development and *Pitx2* implicated in left-right axis patterning). Other authors divide this class into PRD and PRD-like, the former having a paired box and the latter most likely being the ancestral form of these genes.¹

PRD genes are rarely found in ancient conserved clusters, although some arrays of PRD class genes are present in mammals, notably the *Obox*, *Rhox*, and *Dux* loci. The mouse *Obox* (oocyte-specific homeobox) genes are expressed in germ cells and implicated in reproductive biology, but

they are not present in humans and may have evolved specifically in rodents.^{67,68} They comprise a large gene cluster at mouse chromosome 7 A1-2, including 6 intron-containing homeobox genes and 28 intronless loci, plus probable pseudogenes on another chromosome.^{67,68} Since this gene cluster arose relatively recently in evolution, it is unclear if clustering is functionally significant and retained by natural selection. The same applies to the RhoX (reproductive homeobox) genes expressed during embryogenesis and gametogenesis.⁶⁹ These constitute an enormous gene cluster in mouse, with 36 loci clustered together at chromosome X A3.3. RhoX genes are present in other mammals; the human genome contains three RhoX loci at chromosome Xq24.

Another PRD class homeobox gene family to have expanded massively during mammalian evolution is the Dux (double homeobox) gene family, which typically contains two homeobox sequences per gene. Four intron-containing Dux genes can be found in placental mammals, named *Duxa*, *Duxb*, *Duxc*, and *Duxbl*, although not every species has all four due to gene losses in some evolutionary lineages of mammals.^{70,71} There are large numbers of intronless Dux loci in some mammals, for example, the human genome has at least 36 Dux loci, although this number is variable and probably an underestimate, because Dux homeobox sequences have become part of a 3.3 kb tandemly repeated element in human euchromatin and heterochromatin.⁷² Some of these repeated units are functional; contraction in the number of Dux repeats in the D4Z4 tandem array at human chromosome 4q35 is associated with a condition known as facioscapulohumeral muscular dystrophy, probably acting through alteration of local chromatin structure.⁷³ The precursor to the Dux repeats must have arisen by retroposition from a transcript of one of the intron-containing Dux genes, most likely *Duxc*, early in placental mammal evolution.⁷⁰ As for the origin of the Dux gene family itself, this can be dated to the base of placental mammals, since a potential precursor gene—with just a single homeobox sequence—has been identified in the genomes of marsupial mammals (wallaby, opossum), a monotreme (platypus) and two non-mammalian amniotes (chick and anole lizard).⁷¹

Another intriguing feature of PRD class gene evolution is the evolution of 'new' PRD class loci during vertebrate evolution. For example, homeobox genes in the Dprx, Tprx, Leutx, and Argfx gene families have only been found in placental mammals to date.^{12,68,71,74} The evolutionary origins of these genes are not fully resolved, but evidence is accumulating that they may have originated by tandem duplication

from the Crx homeobox gene (a member of the Otx gene family), followed by extensive sequence divergence.

THE EFFECT OF GENOME DUPLICATION

The expansion of the ANTP and the PRD class genes can be attributed primarily to tandem gene duplication, with occasional segmental duplication. In contrast, whole genome duplication is rare in animal evolution, although it has also been relevant to homeobox gene diversification. There is now strong evidence for two complete genome duplications early in the evolution of vertebrates,^{75–77} referred to as the 2R event, and one additional whole genome duplication in teleost fish, referred to as the 3R^{78,79} (Figure 1).

Studies of homeobox genes provided an important strand of evidence in early discussions concerning the existence and timing of the 2R event.^{76,80} For example, the discovery of a single Hox gene cluster in amphioxus^{76,81} (Figure 4), first suggested that duplications post-dated the divergence of vertebrates from cephalochordates, contrary to the early speculations of Ohno.⁸² As other homeobox genes were cloned from amphioxus, these gave a remarkably consistent picture: one gene in amphioxus was very often related to two, three, or four homologs in vertebrates.^{83,84} The complete genome sequencing of an amphioxus species, *Branchiostoma floridae*,⁷⁷ showed that this pattern extends to most of the homeobox superclass, the implication being that there has been a high retention rate of homeobox genes after the 2R genome duplications. Recent surveys of homeobox genes in completed genome sequences estimate 255 functional homeobox genes in human, 279 in the mouse and 238 in the amphibian *Xenopus tropicalis*, but only 133 in amphioxus and 104 in *Drosophila melanogaster*³ (Figure 1).

The consequences of the teleost fish-specific 3R event on homeobox gene evolution may be subtly different from those of the 2R events. The zebrafish genome does have more homeobox genes than other well-studied vertebrates (315 putatively functional homeobox genes), although numbers for non-teleost (pre-duplication) ray-finned fish are not yet known. When Hox genes in particular are examined, there is evidence of duplication from four clusters to eight in teleost fish, although significant Hox gene loss followed the 3R event. Thus, zebrafish has 49 Hox genes, two pufferfish each have 45, stickleback has 48, and a tilapia has 46.⁸⁵ These figures are not massively larger than the 39 Hox genes of mouse or human, suggesting

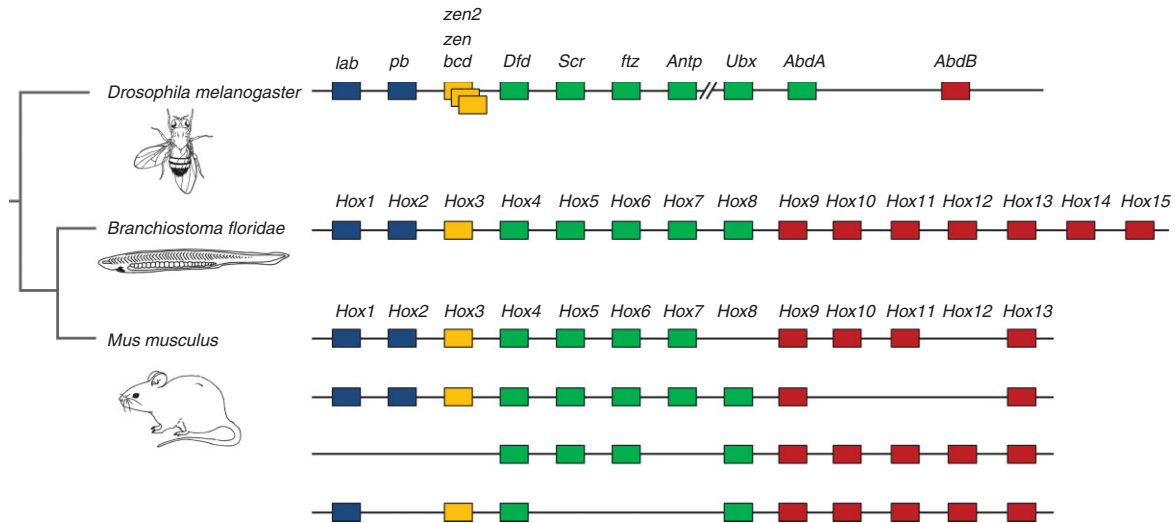


FIGURE 4 | Single Hox gene clusters in an insect (*Drosophila melanogaster*) and amphioxus (*Branchiostoma floridae*) are homologous to four gene clusters in human or mouse (*Mus musculus*). Color coding denotes division of Hox genes in 'anterior', 'group 3', 'middle', and 'posterior' groups. The group 3 gene in *D. melanogaster* has triplicated and diverged to give *zen*, *zen2*, and *bcd*. The Hox genes *zen*, *zen2*, *bcd*, and *ftz* have diverged in function and do not have homeotic roles.

that around 45 might represent a limit of Hox gene complexity attainable (or retainable) in teleost fish.

USE IT OR LOSE IT

Attempts to relate genetic change to phenotypic change in evolution should not overlook gene loss. Inactivation of a gene could be causal in altering a genetic pathway or process, or it could be a secondary consequence (mutations accumulating in a gene that is no longer needed). Until the advent of whole genome sequencing, it was almost impossible to deduce if loss of a gene had occurred in evolution, because even if cloning attempts had failed this could always be for technical reasons or unusual sequence divergence.

One example is the apparent absence of the *abdA* Hox gene in cirripede barnacles,^{86,87} as assessed by PCR and library screening. This is interesting because cirripede barnacles have extremely reduced abdomens, precisely the domain where *abdA* genes are deployed in other arthropods. Similarly, it has been proposed that hagfish lack the Pdx family of ParaHox genes,⁸⁸ a family of genes normally involved in pancreas development. Hagfish may have lost a defined pancreatic organ in evolution. In both these cases, circumstantial evidence for gene loss is strong, but complete genome sequencing will be needed to verify these proposals. In these cases a link between gene function and morphology seems clear, but other cases warn us that such correlations will not always be straightforward. For example, a survey of Hox genes in a range of nematode species revealed extensive Hox

gene loss in some clades, with *Caenorhabditis elegans* having around half the ancestral Hox genes, and *Trichella spiralis*, *Brugia malayi*, and *Ascaris suum* having lost only three or four.⁸⁹ Linking these losses to changes in developmental mechanisms is not yet possible.

With complete genome sequences now available for many animals, it is possible to examine global patterns of gene loss or retention. For example, comparing the homeobox gene complements between the three chordate subphyla (Tunicata, Cephalochordata, and Vertebrata) and outgroups reveals an interesting picture. Vertebrates (including humans) lost 7 ancient homeobox genes, tunicates (including the ascidian *Ciona* and the larvacean *Oikopleura*) lost 25 ancestral genes, and cephalochordates (amphioxus, *Branchiostoma*) lost none.^{24,90,91} It is not clear why there is such a dramatic difference, although it can be argued that the cephalochordate (amphioxus) body plan is the least derived compared to the inferred ancestral morphology, so there is a parallel between extent of morphological divergence and gene loss.

A final example of differential gene loss concerns a dramatic difference between primates and rodents. The human and mouse genomes have different numbers of homeobox genes because of both gene duplications (discussed above) and gene losses. Comparisons between the genomes of different mammals revealed that the *Leutx*, *Ventx*, *Argfx*, *Dprx*, *Shox*, *Rax2*, *LOC647589*, *Tprx1*, and *Nanognb* homeobox genes—all of which are present in a diversity of placental mammals—were lost in rodent evolution but not in primate evolution.⁶⁸

EVOLUTIONARY CHANGES IN HOMEBOX GENE EXPRESSION

Several well-studied mutations in *Drosophila* Hox genes, including the original dominant *Antp* mutation, are associated with Hox genes being expressed at the wrong time or in the wrong place.^{92,93} Similarly, ectopic expression of mouse Hox genes can cause homeotic phenotypes or other developmental abnormalities.⁹⁴ These and other studies indicate that the correct function of Hox genes is critically dependent on their time and place of expression. The same applies to many other types of homeobox gene. It is perhaps not surprising, therefore, that differences in homeobox gene expression between species can be associated with phenotypic differences.

One of the first examples reported concerns the expression of vertebrate Hox genes.^{95,96} Different species of vertebrate animals have different numbers of vertebrae in their skeletons specialized for particular regions of the body. For example, the mouse has 7 cervical or neck vertebrae, 13 thoracic or rib-bearing vertebrae, 6 lumbar, 4 sacral, and many caudals, whereas chick has 14 cervical, 7 thoracic, 12–13 lumbosacral, and 5 coccygeal (vestigial tail) vertebrae. Goose is different again. The implication is that during vertebrate evolution, developmental pathways have been altered such that a particular somite will differentiate into one shape (e.g., a neck vertebra) in one species, but a different shape (e.g., a rib-bearing vertebra) in another species. Comparison of Hox gene expression patterns between embryos of different vertebrates showed that anterior expression boundaries varied between species in concert with morphological change.^{95,96} To give just one example, *Hoxc6* has an anterior expression boundary around somite 12 in mouse, somite 19 in chick, and somite 22 in goose; in each case this corresponds to the first thoracic vertebra⁹⁵ (Figure 5). It would be easy to dismiss this as simply correlation, but this would miss the point. We know, from mutational and transgenic studies, that moving Hox gene boundaries will cause somites to differentiate in altered ways; Hox genes are causative factors in ensuring development occurs correctly in relation to anteroposterior position along the body axis. Hence, it is very likely that naturally-occurring mutations that shifted Hox gene expression boundaries were the cause of shifting axial identities in the evolution of vertebrate diversity.

A similar link between Hox gene expression and morphology has been reported across crustacean diversity. The embryonic anterior expression boundary of the *Ubx* gene (detected using a cross-reactive antibody to both *Ubx* and *abd-A* proteins) predicts the segment in the body where there is a

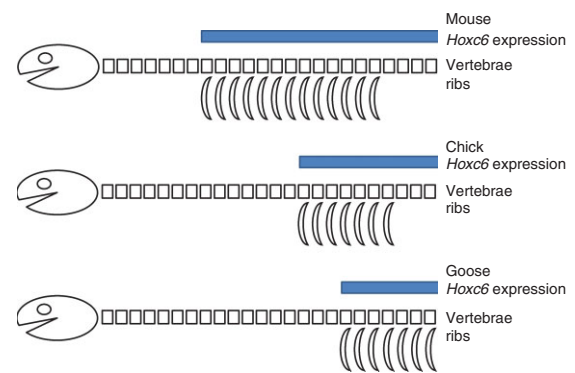


FIGURE 5 | Schematic diagram demonstrating correlation between the anterior expression boundary of the *Hoxc6* gene (blue domain) and the cervical–thoracic boundary in different vertebrate species. The gene is expressed at earlier developmental stages, in embryonic somites, before formation of definitive vertebrae. Occipital vertebrae, which contribute to the base of the skull, are not shown. (Based on data from Ref 95)

transition between two functionally and anatomically distinct types of appendage⁹⁷ (Figure 6). Anterior to this boundary, the appendages on segments are used for feeding, while posterior to this boundary the appendages are adapted for locomotion. The feeding/locomotion transition can occur anywhere from the first to the fourth thoracic segment depending on species. The same researchers also detected cases of intermediate morphology marked by lower or patchy expression of the Hox gene. More recently, a causal relationship between *Ubx* expression and this morphological boundary has been confirmed. Use of siRNA to interfere with *Ubx* gene function in the crustacean *Paryphale hawaiensis* demonstrated that reduction in *Ubx* causes transformation of ‘locomotory segments’ into ‘feeding segments’, or at least into segments bearing appendages with feeding morphology.⁹⁸ Similarly, ectopic expression of *Parhyale Ubx* causes feeding to locomotory transformations.⁹⁹ Shifting Hox gene boundaries correlate with, and most likely have contributed to, anatomical changes to body organization during evolution.

In the vertebrate and crustacean examples above, Hox gene expression changes were conserved, but the precise mutational changes responsible were not identified. An opposite situation has been reported in baleen whales, where a regulatory DNA sequence next to a Hox gene has been found to be mutated but the precise effect on gene expression is not known.¹⁰⁰ Engineering the mutation into the homologous mouse sequence affects Hox gene expression in a transgenic mouse assay, but the *in vivo* effect on development in an animal as experimentally intractable as a whale is not known. A neat example where mutation, effect

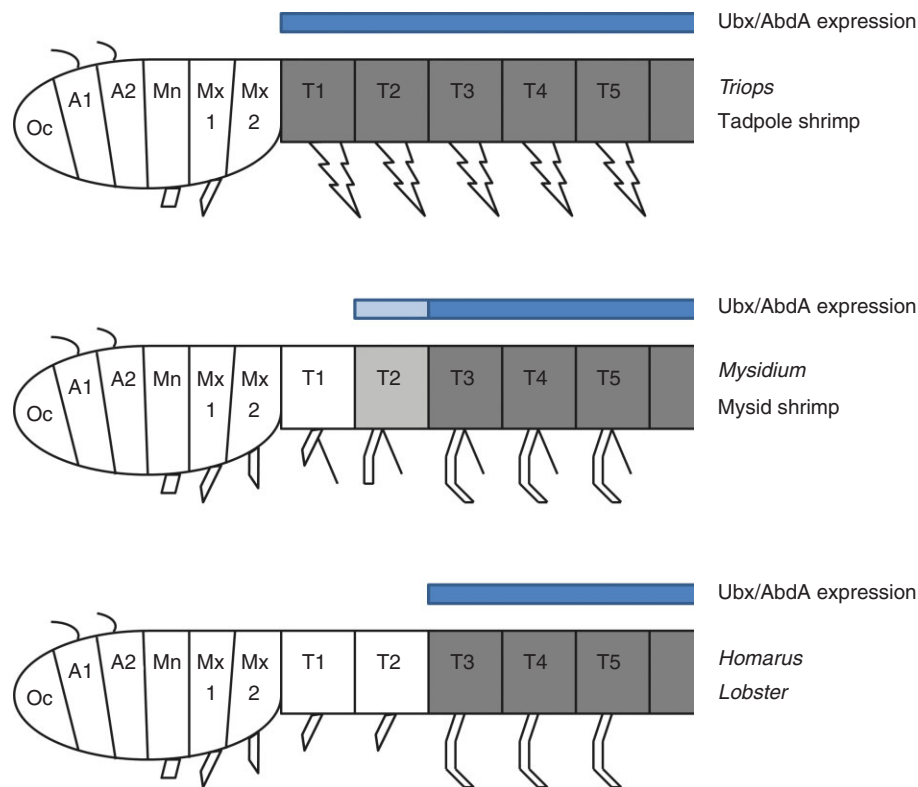


FIGURE 6 | Schematic diagram showing segmental specialization in three crustacean species in relation to the the expression of Ubx protein, detected using an antibody recognizing Ubx and AbdA (blue domain). The six head segments shown are Oc, ocular; A1, first antennal; A2, second antennal, Mn, mandibular; Mx1, first maxillary; Mx2, second maxillary. Only the first five trunk segments are shown (T1–T5). Segments Mn, Mx1, and Mx2 usually bear feeding appendages, and trunk segments bear locomotory appendages. In *Mysidium* T1, and *Homarus* T1 and T2, trunk segments are modified as feeding appendages, in association with absence of Ubx protein expression. *Mysidium* T2 has an intermediate state. (Based on data from Ref 97)

on gene expression and phenotypic consequence are all known concerns one of the PRD class homeobox genes of vertebrates, *Pitx1*. Three-spined sticklebacks *Gasterosteus aculeatus* show considerable variation in the development of their pelvic fins, with some marine or migratory fish developing large and spiny pelvic fins, and others from freshwater having greatly reduced pelvic skeletons. The *Pitx1* homeobox gene, implicated in the development of limbs, thymus, olfactory pits and sensory neuromasts, differs in expression between embryos of these fish, being undetectable specifically in the developing pelvic region of the freshwater forms.¹⁰¹ This difference has been traced to deletion mutations in the enhancer of the *Pitx1* gene that modify expression without inactivating the gene¹⁰² (Figure 7).

CONCLUSION

The homeobox gene superclass is diverse. Although found throughout eukaryotes, it was particularly during animal evolution that homeobox genes diversified

to take up a multiple of developmental roles. In animals, homeobox genes can be classified into at least 11 classes, including the large ANTP and PRD gene classes, each in turn divided into gene families. Within the ANTP class are three sets of homeobox genes that are ancestrally arranged into gene clusters—Hox, ParaHox, and NK—and each of these is a remnant from a swathe of ANTP tandem gene duplication that occurred in early animal evolution. Evidence is accumulating that Hox, ParaHox, and NK homeobox genes may have ancestrally patterned three major embryonic systems—the nervous system, the gut and the mesoderm. This hypothesis needs further testing. In the PRD class, there is little evidence for ancient gene clusters, although more recent tandem duplications have generated some massive arrays of PRD class genes in mammals, notably the Obox, Rux, and Dux clusters. Genome duplication is also a force in homeobox gene evolution, with two whole genome duplications in early vertebrate evolution driving an increase in total homeobox gene count and one additional genome duplication in teleost fish leading to a moderate increase.

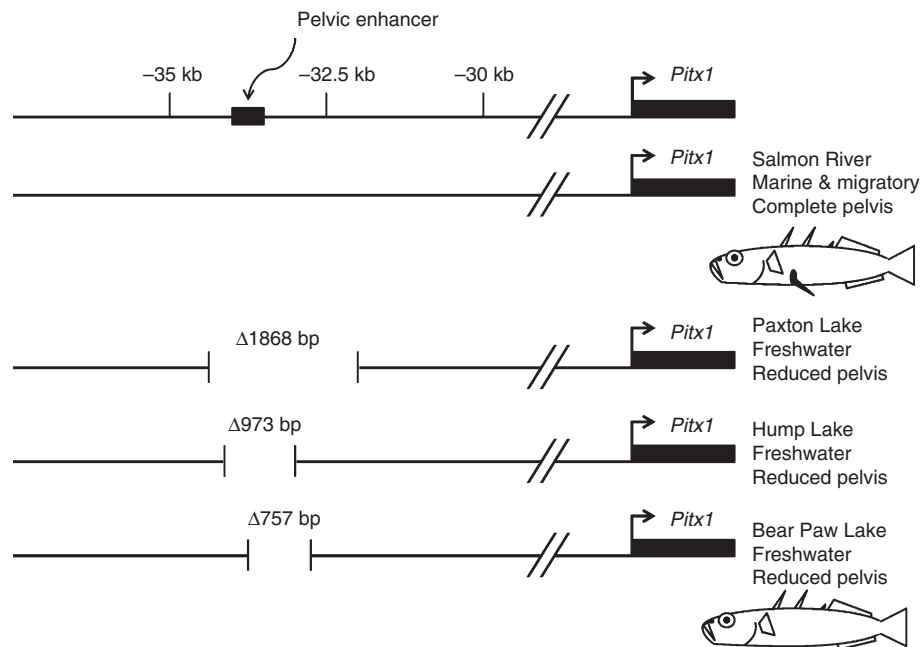


FIGURE 7 | Deletion of a *Pitx1* 'pelvic enhancer' in pelvic-reduced sticklebacks. An enhancer region driving *Pitx1* gene expression specifically in the pelvic region is located around 34 kb 5' of the *Pitx1* transcriptional start site (top line). A population of three-spined sticklebacks from Salmon River has a complete pelvic region (shaded black) and the *cis*-regulatory region is complete. Isolated freshwater populations (three examples shown) can have dramatically reduced pelvic regions, associated with deletion mutations covering or overlapping the characterized pelvic enhancer. (Based on data from Ref 102)

New homeobox genes generated by gene duplication may provide opportunities for natural selection to adapt genes to novel roles, but gene loss might be an equally potent force in evolution. Some examples are known where loss of individual homeobox gene correlates with morphological simplification, but on the whole the patterns of homeobox gene loss generate more questions than answers. Gene gain and loss are not the only dynamics to consider, as evidence accumulates for an important role for gene expression change in evolution. Examples from

vertebrates and crustaceans have revealed shifting Hox gene expression boundaries in concert with changes to anatomy, but it remains a challenge to find the actual mutations responsible. Where mutations affecting homeobox gene expression have been identified and linked to phenotypic change, such as deletions in an enhancer for *Pitx1* in pelvic-reduced fish, this has taken a herculean effort. More such examples are needed if the evolution of homeobox genes is to be fully related to our understanding of the evolution of development and the diversification of body form.

ACKNOWLEDGMENTS

The author's research is funded from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant [268513]11.

REFERENCES

- Burglin TR. Homeodomain subtypes and functional diversity. *Subcell Biochem* 2011, 52:95–122.
- Zhong Y-F, Butts T, Holland PWH. HomeoDB: a database of homeobox gene diversity. *Evol Dev* 2008, 10:516–518.
- Zhong Y-F, Holland PWH. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol Dev* 2011, 13:567–568.
- Mukherjee K, Brocchieri L, Burglin TR. A comprehensive classification and evolutionary analysis of plant homeobox genes. *Mol Biol Evol* 2009, 26:2775–2794.
- Derelle R, Lopez P, Le Guyader H, Manuel M. Homeodomain proteins belong to the ancestral

- molecular toolkit of eukaryotes. *Evol Dev* 2007, 9: 212–219.
6. Shepherd JCW, McGinnis W, Carrasco AE, Derober-
tis EM, Gehring WJ. Fly and frog homoeo domains
show homologies with yeast mating type regulatory
proteins. *Nature* 1984, 310:70–71.
 7. Dave V, Zhao C, Yang F, Tung CS, Ma J. Repro-
grammable recognition codes in Bicoid homeodomain-
DNA interaction. *Mol Cell Biol* 2000, 20:7673–7684.
 8. Rivera-Pomar R, Niessing D, Schmidt-Ott U, Gehring
WJ, Jackle H. RNA binding and translational suppres-
sion by *bicoid*. *Nature* 1996, 379:746–749.
 9. Mizutani Y, Kihara A, Igarashi Y. Mammalian Lass6
and its related family members regulate synthesis of
specific ceramides. *Biochem J* 2005, 390:263–271.
 10. Pewzner-Jung Y, Ben-Dor S, Futerman AH. When
do lasses (longevity assurance genes) become CerS
(ceramide synthases)? Insights into the regula-
tion of ceramide synthesis. *J Biol Chem* 2006,
281:25001–25005.
 11. Mukherjee K, Burglin TR. Comprehensive analysis of
animal TALE homeobox genes: new conserved motifs
and cases of accelerated evolution. *J Mol Evol* 2007,
65:137–153.
 12. Holland PWH, Booth HAF, Bruford EA. Classifica-
tion and nomenclature of all human homeobox genes.
BMC Biol 2007, 5:47.
 13. Joyner A, Hanks M. The engrailed genes: evolution of
function. *Semin Dev Bio* 1991, 2:435:445.
 14. Plouhinec JL, Sauka-Spengler T, Germot A, Le Mentec
C, Cabana T, Harrison G, Pieau C, Sire JY, Veron
G, Mazan S. The mammalian *Crx* genes are highly
divergent representatives of the *Otx5* gene family, a
gnathostome orthology class of *orthodenticle*-related
homeogenes involved in the differentiation of retinal
photoreceptors and circadian entrainment. *Mol Biol
Evol* 2003, 20:513–521.
 15. de Rosa R, Grenier JK, Andreeva T, Cook CE, Adoutte
A, Akam M, Carroll SB, Balavoine G. Hox genes in
brachiopods and priapulids and protostome evolution.
Nature 1999, 399:772–776.
 16. Graham A, Papalopulu N, Krumlauf R. The murine
and *Drosophila* homeobox gene complexes have com-
mon features of organization and expression. *Cell*
1989, 57:367–378.
 17. Stock DW, Ellies DL, Zhao ZY, Ekker M, Rud-
dle FH, Weiss KM. The evolution of the vertebrate
Dlx gene family. *Proc Natl Acad Sci USA* 1996,
93:10858–10863.
 18. Canon S, Herranz C, Manzanares M. Germ cell
restricted expression of chick *nanog*. *Dev Dyn* 2006,
235:2889–2894.
 19. Ryan JF, Burton PM, Mazza ME, Kwong GK, Mul-
likin JC, Finnerty JR. The cnidarian-bilaterian ancestor
possessed at least 56 homeoboxes: evidence from the
starlet sea anemone, *Nematostella vectensis*. *Genome
Biol* 2006, 7:R64.
 20. Laughon A, Scott MP. Sequence of a *Drosophila*
segmentation gene–protein–structure homology with
DNA binding proteins. *Nature* 1984, 310:25–31.
 21. Sebe-Pedros A, de Mendoza A, Lang BF, Degnan BM,
Ruiz-Trillo I. Unexpected repertoire of metazoan tran-
scription factors in the unicellular holozoan *Capsas-
pora owczarzaki*. *Mol Biol Evol* 2011, 28:1241–1254.
 22. Burglin TR. Analysis of TALE superclass homeobox
genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a
novel domain conserved between plants and animals.
Nucl Acids Res 1997, 25:4173–4180.
 23. Burglin TR. The PBC domain contains a MEINOX
domain: coevolution of Hox and TALE homeobox
genes? *Dev Genes Evol* 1998, 208:113–116.
 24. Takatori N, Butts T, Candiani S, Pestarino M, Fer-
rier D, Saiga H, Holland PWH. Comprehensive survey
and classification of homeobox genes in the genome of
amphioxus, *Branchiostoma floridae*. *Dev Genes Evol*
2008, 218:579–590.
 25. Brooke NM, Garcia-Fernandez J, Holland PWH. The
ParaHox gene cluster is an evolutionary sister of the
Hox gene cluster. *Nature* 1998, 392:920–922.
 26. Chourrout D, Delsuc F, Chourrout P, Edvardsen
RB, Rentsch F, Renfer E, Jensen MF, Zhu B, de
Jong P, Steele RE, et al. Minimal ProtoHox cluster
inferred from bilaterian and cnidarian Hox comple-
ments. *Nature* 2006, 442:684–687.
 27. Garcia-Fernandez J. Hox, ParaHox, ProtoHox: facts
and guesses. *Heredity* 2004, 94:145–152.
 28. Gehring WJ, Kloter U, Suga H. Evolution of the Hox
Gene complex from an evolutionary ground state.
Curr Top Dev Biol 2009, 88:35–61.
 29. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chap-
man J, Salamov A, Terry A, Shapiro H, Lindquist E,
Kapitonov VV, et al. Sea anemone genome reveals
ancestral eumetazoan gene repertoire and genomic
organization. *Science* 2007, 317:86–94.
 30. Hui JHL, Holland PWH, Ferrier DEK. Do cnidarians
have a ParaHox cluster? Analysis of synteny around a
Nematostella homeobox gene cluster. *Evol Dev* 2008,
10:725–730.
 31. Quiquand M, Yanze N, Schmich J, Schmid V, Galliot
B, Piraino S. More constraint on ParaHox than Hox
gene families in early metazoan evolution. *Dev Biol*
2009, 328:173–187.
 32. Kamm K, Schierwater B, Jakob W, Dellaporta SL,
Miller DJ. Axial patterning and diversification in the
Cnidaria predate the Hox system. *Curr Biol* 2006,
16:920–926.
 33. Schierwater B, Kamm K. The early evolution of
Hox genes: a battle of belief? *Adv Exp Med Biol*
2010:81–90.
 34. Larroux C, Fahey B, Degnan SM, Adamski M,
Rokhsar DS, Degnan BM. The NK homeobox gene

- cluster predates the origin of Hox genes. *Curr Biol* 2007, 17:706–710.
35. Ryan J, Pang K, Program NCS, Mullikin J, Martindale M, Baxevanis A. The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the Para-Hoxzoa. *EvoDevo* 2010, 1:9.
 36. Monteiro AS, Schierwater B, Dellaporta SL, Holland PWH. A low diversity of ANTP class homeobox genes in Placozoa. *Evol Dev* 2006, 8:174–182.
 37. Schierwater B, Kuhn K. Homology of hox genes and the zootype concept in early metazoan evolution. *Mol Phylogenet Evol* 1998, 9:375–381.
 38. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, et al. The *Trichoplax* genome and the nature of placozoans. *Nature* 2008, 454:955–960.
 39. Jakob W, Sagasser S, Dellaporta S, Holland PWH, Kuhn K, Schierwater B. The Trox-2 Hox/ParaHox gene of *Trichoplax* (Placozoa) marks an epithelial boundary. *Dev Genes Evol* 2004, 214:170–175.
 40. Butts T, Holland PWH, Ferrier DEK. The urbilaterian super-Hox cluster. *Trends Genet* 2008, 24:259–262.
 41. Coulier F, Popovici C, Villet R, Birnbaum D. MetaHox gene clusters. *J Exp Zool* 2000, 288:345–351.
 42. Jagla K, Bellard M, Frasch M. A cluster of *Drosophila* homeobox genes involved in mesoderm differentiation programs. *BioEssays* 2001, 23:125–133.
 43. Luke GN, Castro LF, McLay K, Bird C, Coulson A, Holland PWH. Dispersal of NK homeobox gene clusters in amphioxus and humans. *Proc Natl Acad Sci USA* 2003, 100:5292–5295.
 44. Castro LFC, Holland PWH. Chromosomal mapping of ANTP class homeobox genes in amphioxus: piecing together ancestral genomes. *Evol Dev* 2003, 5:459–465.
 45. Pollard SL, Holland PWH. Evidence for 14 homeobox gene clusters in human genome ancestry. *Curr Biol* 2000, 10:1059–1062.
 46. Hui JHL, McDougall C, Monteiro AS, Holland PWH, Arendt D, Balavoine G, Ferrier DEK. Extensive chordate and annelid macrosynteny reveals ancestral homeobox gene organization. *Mol Biol Evol* 2012, 29:157–165.
 47. Garcia-Fernandez J. The genesis and evolution of homeobox gene clusters. *Nat Rev Genet* 2005, 6:881–892.
 48. Holland PWH, Hogan BLM. Expression of homeobox genes during mouse development: a review. *Genes Dev* 1988, 2:773–782.
 49. Mallo M, Wellik DM, Deschamps J. Hox genes and regional patterning of the vertebrate body plan. *Dev Biol* 2010, 344:7–15.
 50. Prince VE, Price AL, Ho RK. Hox gene expression reveals regionalization along the anteroposterior axis of the zebrafish notochord. *Dev Genes Evol* 1998, 208:517–522.
 51. Krumlauf R. Mouse Hox genetic functions. *Curr Opin Genet Dev* 1993, 3:621–625.
 52. Andrew DJ, Scott MP. Downstream of the homeotic genes. *New Biol* 1992, 4:5–15.
 53. Enriquez J, Boukhatmi H, Dubois L, Philippakis AA, Bulyk ML, Michelson AM, Crozatier M, Vincent A. Multi-step control of muscle diversity by Hox proteins in the *Drosophila* embryo. *Development* 2010, 137:457–466.
 54. Hoppler S, Bienz M. Specification of a single cell type by a *Drosophila* homeotic gene. *Cell* 1994, 76:689–702.
 55. Holland PWH, Garcia-Fernandez J. Hox genes and chordate evolution. *Dev Biol* 1996, 173:382–395.
 56. Fröblius AC, Matus DQ, Seaver EC. Genomic organization and expression demonstrate spatial and temporal Hox gene colinearity in the lophotrochozoan *Capitella* sp. I. *PLoS ONE* 2008, 3:e4004.
 57. Holland PWH. Major transitions in animal evolution: A developmental genetic perspective. *Am Zool* 1998, 38:829–842.
 58. Holland PWH. Beyond the Hox: how widespread is homeobox gene clustering? *J Anat* 2001, 199:13–23.
 59. Ashizawa S, Brunicardi FC, Wang XP. PDX-1 and the pancreas. *Pancreas* 2004, 28:109–120.
 60. Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, Magnuson MA, Hogan BLM, Wright CVE. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* 1996, 122:983–995.
 61. Arendt D, Technau U, Wittbrodt J. Evolution of the bilaterian larval foregut. *Nature* 2001, 409:81–85.
 62. Hui JHL, Raible F, Korchagina N, Dray N, Samain S, Magdelenat G, Jubin C, Segurens B, Balavoine G, Arendt D, et al. Features of the ancestral bilaterian inferred from *Platynereis dumerilii* ParaHox genes. *BMC Biol* 2009, 7:43.
 63. Samadi L, Steiner G. Conservation of ParaHox genes' function in patterning of the digestive tract of the marine gastropod *Gibbula varia*. *BMC Dev Biol* 2010, 10:74.
 64. Mazet F, Amemiya CT, Shimeld SM. An ancient Fox gene cluster in bilaterian animals. *Curr Biol* 2006, 16:R314–R316.
 65. Shimeld SM, Boyle MJ, Brunet T, Luke GN, Seaver EC. Clustered Fox genes in lophotrochozoans and the evolution of the bilaterian Fox gene cluster. *Dev Biol* 2010, 340:234–248.
 66. Suga H, Tschopp P, Graziussi DF, Stierwald M, Schmid V, Gehring WJ. Flexibly deployed Pax genes in eye development at the early evolution of animals demonstrated by studies on a hydrozoan jellyfish. *Proc Natl Acad Sci* 2010, 107:14263–14268.

67. Rajkovic A, Yan CN, Yan W, Klysik M, Matzuk MM. Obox, a family of homeobox genes preferentially expressed in germ cells. *Genomics* 2002, 79:711–717.
68. Zhong Y-F, Holland PWH. The dynamics of vertebrate homeobox gene evolution: gain and loss of genes in mouse and human lineages. *BMC Evol Biol* 2011, 11:169.
69. MacLean JA, Wilkinson MF. The RhoX genes. *Reproduction* 2010, 140:195–213.
70. Clapp J, Mitchell LM, Bolland DJ, Fantes J, Corcoran AE, Scoffing PJ, Armour JAL, Hewitt JE. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulo-humeral muscular dystrophy. *Am J Hum Genet* 2007, 81:264–279.
71. Leidenroth A, Hewitt JE. A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC Evol Biol* 2010, 10:364.
72. Gabriels J, Beckers MC, Ding H, De Vriese A, Plaisance S, van der Maarel SM, Padberg GW, Frants RR, Hewitt JE, Collen D, et al. Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* 1999, 236:25–32.
73. Vandutekom JCT, Wijmenga C, Vantienhoven EAE, Gruter AM, Frants RR, Hewitt JE, Padberg GW, Vanommen GJB, Hofker MH. FSHD associated DNA rearrangements are due to deletions of integral copies of A3.2 KB tandemly repeated unit. *Hum Mol Genet* 1993, 2:2037–2042.
74. Booth HAF, Holland PWH. Annotation, nomenclature and evolution of four novel homeobox genes expressed in the human germ line. *Gene* 2007, 387:7–14.
75. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. *PLOS Biol* 2005, 3:1700–1708.
76. Holland PWH, Garcia-Fernandez J, Williams NA, Sidow A. Gene duplications and the origins of vertebrate development. *Dev Suppl* 1994, 1994:125–133.
77. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 2008, 453:1064–1071.
78. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 2004, 431:946–957.
79. Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res* 2003, 13:382–390.
80. Lundin LG. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* 1993, 16:1–19.
81. Garcia-Fernandez J, Holland PWH. Archetypal organization of the amphioxus Hox gene cluster. *Nature* 1994, 370:563–566.
82. Ohno S. *Evolution by gene duplication*. New York: Springer-Verlag; 1970.
83. Holland PWH. Molecular biology of lancelets: insights into development and evolution. *Isr J Zool* 1996, 42:S247–S272.
84. Holland PWH. More genes in vertebrates? *J Struct Funct Genomics* 2003, 3:75–84.
85. Hoegg S, Boore JL, Kuehl JV, Meyer A. Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 2007, 8:317.
86. Geant E, Mouchel-Vielh E, Coutanceau J-P, Ozouf-Costaz C, Deutsch JS. Are cirripedia hopeful monsters? Cytogenetic approach and evidence for a Hox gene cluster in the cirripede crustacean *Sacculina carcini*. *Dev Genes Evol* 2006, 216:443–449.
87. Mouchel-Vielh E, Rigolot C, Gibert JM, Deutsch JS. Molecules and the body plan: the Hox genes of Cirripedes (Crustacea). *Mol Phylogenet Evol* 1998, 9:382–389.
88. Furlong RF, Younger R, Kasahara M, Reinhardt R, Thorndyke M, Holland P. A degenerate ParaHox gene cluster in a degenerate vertebrate. *Mol Biol Evol* 2007, 24:2681–2686.
89. Aboobaker AA, Blaxter ML. Hox gene loss during dynamic evolution of the nematode cluster. *Curr Bio* 2003, 13:37–40.
90. Holland LZ, Albalat R, Azumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ, et al. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res* 2008, 18:1100–1111.
91. Holland PWH. From genomes to morphology: a view from amphioxus. *Acta Zool* 2010, 91:81–86.
92. Frischer LE, Hagen FS, Garber RL. An inversion that disrupts the *Antennapedia* gene causes abnormal structure and localization of RNAs. *Cell* 1986, 47:1017–1023.
93. Schneuwly S, Kuroiwa A, Gehring WJ. Molecular analysis of the dominant homeotic Antennapedia phenotype. *EMBO J* 1987, 6:201–206.
94. Kessel M, Balling R, Gruss P. Variations of cervical vertebrate after expression of a Hox-1.1 transgene in mice. *Cell* 1990, 61:301–308.
95. Burke AC, Nelson CE, Morgan BA, Tabin C. Hox genes and the evolution of vertebrate axial morphology. *Development* 1995, 121:333–346.

96. Gaunt SJ. Conservation in the hox code during morphological evolution. *Int J Dev Biology* 1994, 38:549–552.
97. Averof M, Patel NH. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* 1997, 388:682–686.
98. Liubicich DM, Serano JM, Pavlopoulos A, Kontarakis Z, Protas ME, Kwan E, Chatterjee S, Tran KD, Averof M, Patel NH. Knockdown of *Parhyale Ultrabithorax* recapitulates evolutionary changes in crustacean appendage morphology. *Proc Natl Acad Sci USA* 2009, 106:13892–13896.
99. Pavlopoulos A, Kontarakis Z, Liubicich DM, Serano JM, Akam M, Patel NH, Averof M. Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. *Proc Natl Acad Sci USA* 2009, 106:13897–13902.
100. Shashikant CS, Kim CB, Borbely MA, Wang WCH, Ruddle FH. Comparative studies on mammalian *Hoxc8* early enhancer sequence reveal a baleen whale-specific deletion of a cis-acting element. *Proc Natl Acad Sci USA* 1998, 95:15446–15451.
101. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, Kingsley DM. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 2004, 428:717–723.
102. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 2010, 327:302–305.

APPENDIX

TABLE 1 | A Homeobox Glossary

| | |
|--------------------------|--|
| Homeobox | DNA sequence motif usually translated into homeodomain. Found in many genes, not just Hox genes. Present in eukaryotes. |
| Homeodomain | Peptide motif encoded by homeobox, capable of folding into three alpha helices with certain conserved residues. Often 60 amino acids though many are longer. |
| Homeotic gene | A gene which when mutated causes transformation of one body structure into the likeness of another. Most Hox genes are homeotic, though not all homeotic mutations are in Hox genes. |
| Hox genes | Genes orthologous to genes of the ANT-C and BX-C gene complexes of <i>Drosophila</i> and the Hoxa, Hoxb, Hoxc, Hoxd gene clusters of mammals. Often but not always in gene clusters. All Hox genes possess a homeobox, but not all homeobox genes are Hox genes. Present in most but not all animal phyla. Implicated in anteroposterior positional specification in bilaterian animals. |
| Hox-linked gene array | A set of homeobox genes mapping chromosomally close to Hox genes, but not necessarily as one tightly-linked gene cluster. Depending on species may include Hox, Evx, Dlx, Ro, Mnx, En, Mox, Hex, Nedx, and/or Gbx. |
| ParaHox genes | A set of three homeobox gene families (Gsx, Xlox/Pdx, and Cdx) with high sequence similarity to Hox genes. Clustered in some animal species. Implicated in gut and neural patterning. |
| ProtoHox gene | Hypothetical ancestral gene that gave rise to Hox and ParaHox genes, possibly through the intermediate step of a ProtoHox gene cluster. Pre-dated the cnidarian-bilaterian divergence. |
| Mega-cluster | Hypothetical ancestral gene cluster containing precursors of NK-linked gene array and Hox-linked gene array. |
| MetaHox gene cluster | An alternative name for the NK homeobox gene cluster. |
| NK homeobox gene cluster | A tightly-linked cluster of homeobox genes comprising some but not necessarily all of NK1, NK3, NK4, Tlx, Lbx, and Msx. Gene cluster present in insects, but secondarily dispersed in vertebrates and amphioxus. |
| NK-linked array | A set of homeobox genes mapping chromosomally close to NK cluster genes, but not necessarily as one tightly-linked gene cluster. Depending on species may include NK1, NK3, NK4, Tlx, Lbx, Msx, NK5, Emx, and/or NK6. |
| SuperHox gene cluster | Hypothetical ancestral gene cluster in bilaterian ancestor containing precursors of Hox-linked array only, before their dispersal. Proposed to have comprised Hox, Evx, Dlx, Ro, Mnx, En, Mox, Hex, Nedx, and Gbx. |