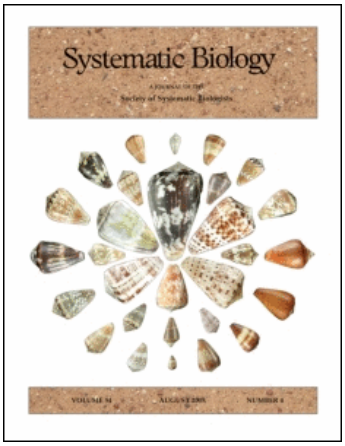
This article was downloaded by: *[Australian National University]*
On: *13 October 2008*
Access details: *Access Details: [subscription number 788777827]*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK




**Systematic Biology**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713658732

# Statistical Tests between Competing Hypotheses of Hox Cluster Evolution

Robert Lanfear [abc]; Lindell Bromham [ab]
[a] Centre for the Study of Evolution, School of Life Sciences, University of Sussex, Brighton, UK [b] Centre for Macroevolution and Macroecology, School of Botany and Zoology, Australian National University, Canberra, Australia [c] Department of Biology, University College London, London, UK

First Published on: 01 October 2008

**To cite this Article** Lanfear, Robert and Bromham, Lindell(2008)'Statistical Tests between Competing Hypotheses of Hox Cluster Evolution',Systematic Biology,57:5,708 — 718
**To link to this Article: DOI:** 10.1080/10635150802430079
**URL:** http://dx.doi.org/10.1080/10635150802430079

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

*Syst. Biol.* 57(5):708–718, 2008
Copyright © Society of Systematic Biologists
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150802430079

# Statistical Tests between Competing Hypotheses of Hox Cluster Evolution

ROBERT LANFEAR[1,2,3] AND LINDELL BROMHAM[1,2]

[1]*Centre for the Study of Evolution, School of Life Sciences, University of Sussex, Brighton, BN1 9QG, UK; E-mail: Rob.Lanfear@anu.edu.au (R.L.)*
[2]*Centre for Macroevolution and Macroecology, School of Botany and Zoology, Australian National University, Canberra, ACT 0200, Australia*
[3]*Department of Biology, University College London, London WC1E 6BT, UK*

*Abstract*.—The Hox genes encode transcription factors that play vital roles in the anterior-posterior patterning of all bilaterian phyla studied to date. Additionally, the gain of Hox genes by duplication has been widely implicated as a driving force in the evolution of animal body plans. Because of this, reconstructing the evolution of the Hox cluster has been the focus of intense research interest. It has been commonly assumed that an ancestral four-gene ProtoHox cluster was duplicated early in animal evolution to give rise to the Hox and ParaHox clusters. However, this hypothesis has recently been called into question, and a number of alternative hypotheses of Hox and ParaHox gene evolution have been proposed. Here, we present the first statistical comparisons of current hypotheses of Hox and ParaHox gene evolution. We use two statistical methods that represent two different approaches to the treatment of phylogenetic uncertainty. In the first method, we estimate the maximum-likelihood tree for each hypothesis and compare these trees to one another using a parametric bootstrapping approach. In the second method, we use Bayesian phylogenetics to estimate the posterior distribution of trees, then we calculate the support for each hypothesis from this distribution. The results of both methods are largely congruent. We find that we are able to reject five out of the eight current hypotheses of Hox and ParaHox gene evolution that we consider. We conclude that the ProtoHox cluster is likely to have contained either three or four genes but that there is insufficient phylogenetic signal in the homeodomains to distinguish between these alternatives. [Bayesian; Homeobox; maximum likelihood; phylogenetics; ProtoHox; ParaHox.]

Downloaded By: [Australian National University] At: 23:40 13 October 2008

The expansion of the Hox cluster has been considered a causal factor in the evolution and diversification of animal body plans (Gellon and McGinnis, 1998; Budd, 1999; Hughes and Kaufman, 2002; Wagner et al., 2003), and because of this, reconstructing the evolution of the Hox cluster has been the focus of intense research interest. It is commonly assumed that an ancestral four-gene ProtoHox cluster duplicated early in animal evolution to give rise to the Hox and ParaHox clusters (Brooke et al., 1998; Kourakis and Martindale, 2000; Ferrier and Holland, 2001; Martinez and Amemiya, 2002; Ferrier and Minguillon, 2003). This hypothesis is based on the observation that Hox and ParaHox genes tend to fall into four main phylogenetic groupings: anterior (*Hox1*, *Hox2*, and the ParaHox gene *Gsx*), group 3 (*Hox3* and the ParaHox gene *Xlox*), central (*Hox4* to *Hox8*), and posterior (*Hox9* to *Hox14* and the ParaHox gene *Cdx*; Brooke et al., 1998; Kourakis and Martindale, 2000; Ferrier and Holland, 2001; Martinez and Amemiya, 2002; Ferrier and Minguillon, 2003). The lack of a ParaHox gene connected to the central (*Hox4* to *Hox8*) grouping was originally attributed to the loss of the central ParaHox gene soon after the ProtoHox cluster duplication event (Brooke et al., 1998).

The four-ProtoHox gene hypothesis has recently been called into question on three grounds. First, although the monophyly of most Hox and ParaHox genes is well supported, the support for some groupings of Hox and ParaHox genes is poor, with bootstrap values ranging from 40% to 70%, well below levels required for confident phylogenetic inference (Garcia-Fernandez, 2005). Second, observations from the recently sequenced cnidarian *Nematostella vectensis* suggest that the Cnidaria may contain only two groups of Hox and ParaHox genes, which has led to the suggestion that the ProtoHox cluster may have contained either two or three genes, rather than four (Finnerty and Martindale, 1999; Ferrier and Holland, 2001; Garcia-Fernandez, 2005). Third, results from recent phylogenetic analyses have challenged the four-ProtoHox gene hypothesis, in particular indicating that the ProtoHox cluster may have contained only two genes (Chourrout et al., 2006), or that both the Hox and ParaHox clusters arose via tandem duplication, without the involvement of whole-cluster duplication (Ryan et al., 2007).

Despite the fact that different phylogenetic analyses of Hox genes have resulted in very different hypotheses about the evolutionary history of the hox cluster, there have been no explicit statistical comparisons of different hypotheses of Hox and ParaHox gene evolution. Although the majority of phylogenetic analyses provide measures of clade support (such as bootstrap percentages or Bayesian posterior probability values), these are usually mapped onto the preferred tree from a given analysis and tell us little about the relative support for one hypothesis over another. In analyses that are limited by such short alignments (typically restricted to the 60 amino acid homeodomain), it seems plausible that there may be insufficient power in the data to allow the confident inference of evolutionary events that occurred over 500 million years ago.

Here we present the first statistical comparisons of current hypotheses of early Hox and ParaHox gene evolution. In total, we consider eight different hypotheses (Fig. 1), each of which suggests a series of events leading to the formation of the pre-bilaterian Hox and ParaHox clusters. The first seven hypotheses represent different assumptions about the number of genes contained in the hypothetical ProtoHox cluster (two, three, or four ProtoHox genes), and these are described below. The first hypothesis we consider is the commonly assumed four-ProtoHox gene hypothesis, which postulates that the ProtoHox cluster contained four genes and then duplicated to form a four-gene Hox cluster and a four-gene

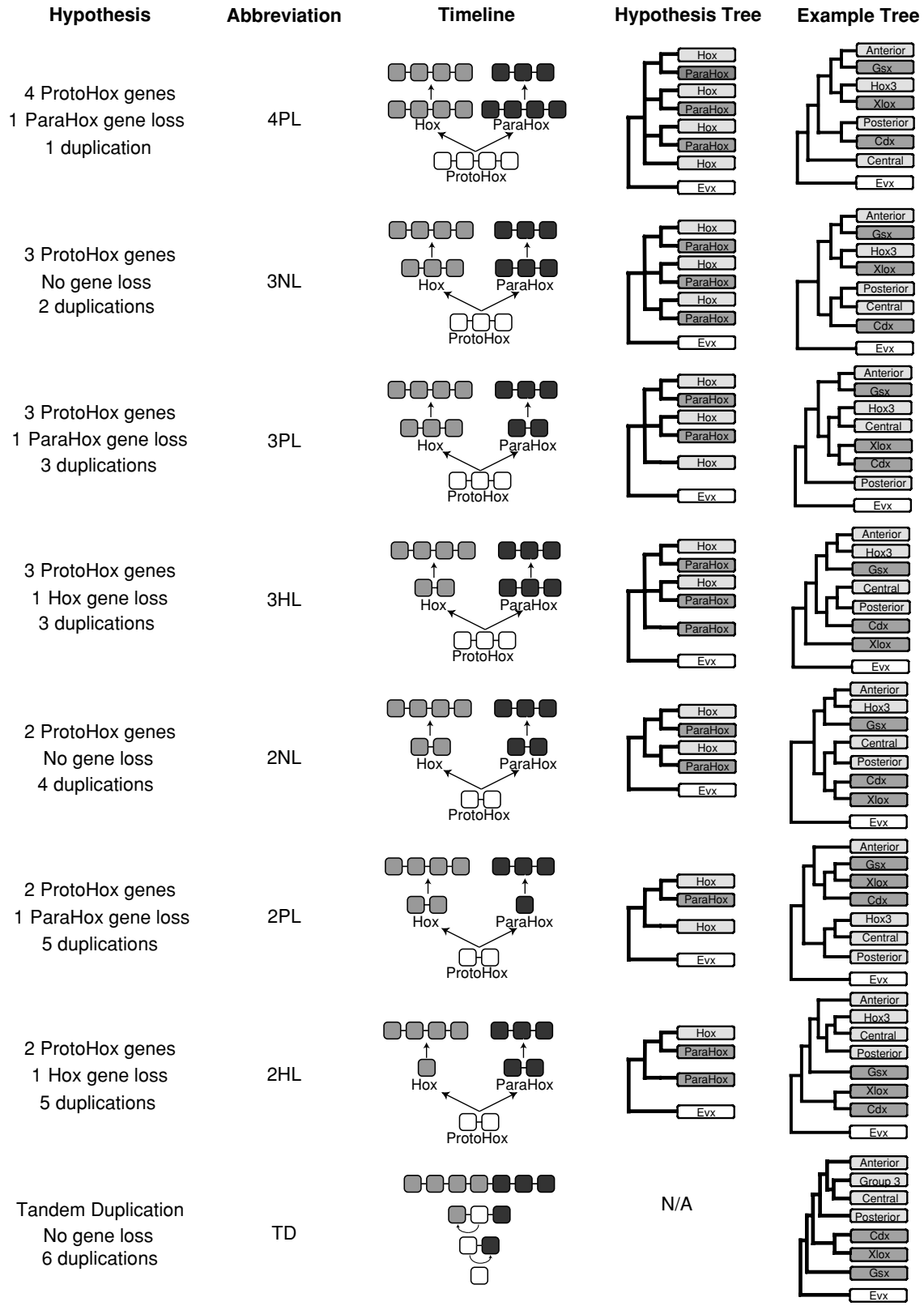



FIGURE 1. Eight different hypotheses of Hox and ParaHox gene evolution. The number of duplications and gene losses posited by each hypothesis is indicated. Each hypothesis is shown as both a timeline, with time running upwards, and as a generalized phylogenetic tree of the expected relationships between Hox and ParaHox genes for that particular hypothesis. An example of a phylogenetic tree consistent with each hypothesis is also shown. For the tandem duplication (TD) hypothesis, both the timeline and the tree represent a particular example of that hypothesis, rather than a generalization, as it is not possible to draw a generalized hypothesis tree for this hypothesis.

Downloaded By: [Australian National University] At: 23:40 13 October 2008

Downloaded By: [Australian National University] At: 23:40 13 October 2008

ParaHox cluster, and that one ParaHox gene was subsequently lost (a ParaHox gene must be lost to accommodate the observation that extant bilaterians contain only three groups of ParaHox genes, not four; 4PL, Fig. 1). Second, we consider three possible versions of the three-ProtoHox gene hypothesis. The three-Protohox gene hypothesis postulates that the ProtoHox cluster contained three genes and that it duplicated to form a three-gene Hox cluster and three-gene ParaHox cluster. The three-gene Hox and ParaHox clusters may have persisted until the beginning of the radiation of the bilaterian phyla with no gene losses (3NL, Fig. 1), a hypothesis that is equally parsimonious to the 4PL hypothesis in terms of the total number of gene loss and duplications events required (Fig. 1). However, it is also possible to imagine two further permutations of the three-ProtoHox gene hypothesis, in which either the ParaHox or the Hox cluster suffer the loss of one gene (Fig. 1, 3PL and 3HL, respectively) before the bilaterian radiation. These hypotheses are equally parsimonious to the 4PL hypothesis in terms of gene loss events, although they are less parsimonious overall as they require more duplication events than the 4PL hypothesis (Fig. 1). Finally, we consider three possible versions of the two-ProtoHox gene hypothesis: that no gene losses occurred post-ProtoHox cluster duplication, that one ParaHox gene was lost or that one Hox gene was lost (Fig. 1, 2NL, 2PL, and 2HL, respectively). All versions of the two-ProtoHox gene hypothesis require more gene loss and duplication events than any other hypotheses considered (Fig. 1). Each hypothesis involving duplications of the ProtoHox cluster can be drawn as a phylogenetic tree, in which the expected phylogenetic relationships between Hox and ParaHox genes arising from each hypothesis are laid out (Fig. 1).

The eighth hypothesis we consider is that the Hox and ParaHox clusters arose by the sequential tandem duplication of one original gene into a single Hox/ParaHox cluster, which was later separated into two distinct clusters (Fig. 1, TD; Ryan et al., 2007). This hypothesis suggests that each gene in the Hox/ParaHox cluster arose as a duplicate of a neighboring gene in the cluster. It cannot be represented as a single "hypothesis tree," but an example of a phylogenetic tree consistent with the TD hypothesis is given in Figure 1. Because this hypothesis postulates that each gene arose as the duplicate of a neighboring gene, it is necessary to know the spatial relationships of the genes in the hypothetical Hox/ParaHox cluster before the TD hypothesis can be tested. The arrangement of genes within both the Hox and ParaHox clusters is conserved in many bilaterian phyla; however, their orientation with respect to one another in the hypothetical Hox/ParaHox cluster cannot be known a priori, and so we allow for two such possibilities in our analyses, shown in Figure 2.

We used an amino acid alignment of 68 bilaterian homeobox genes (50 Hox sequences, 13 ParaHox sequences, and 5 outgroup sequences; see Fig. 3) and two different statistical approaches to explicitly compare these eight hypotheses of Hox and ParaHox gene evolution (Fig. 1). We limit our analysis to the consideration

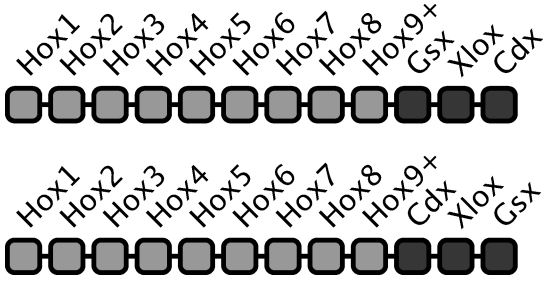



FIGURE 2. Two possible gene orders in the hypothetical linked Hox/ParaHox cluster.

of these eight hypotheses, as other hypotheses, such as those generated by strict interpretation of the maximum-likelihood topology, are far less parsimonious with respect to gene duplication and loss events (Ryan et al., 2007).

We use two methods to test between phylogenetic hypotheses—one maximum likelihood and one Bayesian. These two methods represent fundamentally different approaches to the treatment of phylogenetic uncertainty and to hypothesis testing. In the first method, we assume that the ML tree for a given hypothesis is sufficient to represent that hypothesis. We then test between trees using a frequentist parametric bootstrapping method. In the second method, we use a Bayesian MCMC to estimate the posterior distribution of trees and compare support for different hypotheses by assessing the extent to which each hypothesis is over- or under-represented in the posterior distribution. The Bayesian approach explicitly incorporates uncertainty about the true phylogenetic tree and does not assume that any one tree is representative of any one hypothesis.

## MATERIALS AND METHODS

### *Alignment*

The dataset comprised an amino acid alignment of 68 bilaterian homeodomain sequences: 50 Hox sequences, 13 ParaHox sequences, and 5 *Evx* sequences (Fig. 3; see Appendix 2 at www.systematicbiology.org for accession numbers). We do not use cnidarian sequences in this analysis, as some of our methods require assumptions to be made about gene orthology, and the orthology of the cnidarian homeodomains is not certain (Ryan et al., 2007). *Evx* was used as an outgroup to root phylogenetic trees because it has been convincingly demonstrated to be outside the Hox/ParaHox clade but nonetheless a close relative to it (Minguillon and Garcia-Fernandez, 2003). All sequences were aligned by eye in Se-Al (Rambaut, 1996). Regions outside the homeodomain are extremely divergent between different Hox paralogues, and between Hox and ParaHox genes, and were thus excluded from the alignment.

### *Method 1: SOWH-Like Tests*

In the first statistical method we used, we first calculated the maximum-likelihood (ML) tree for each

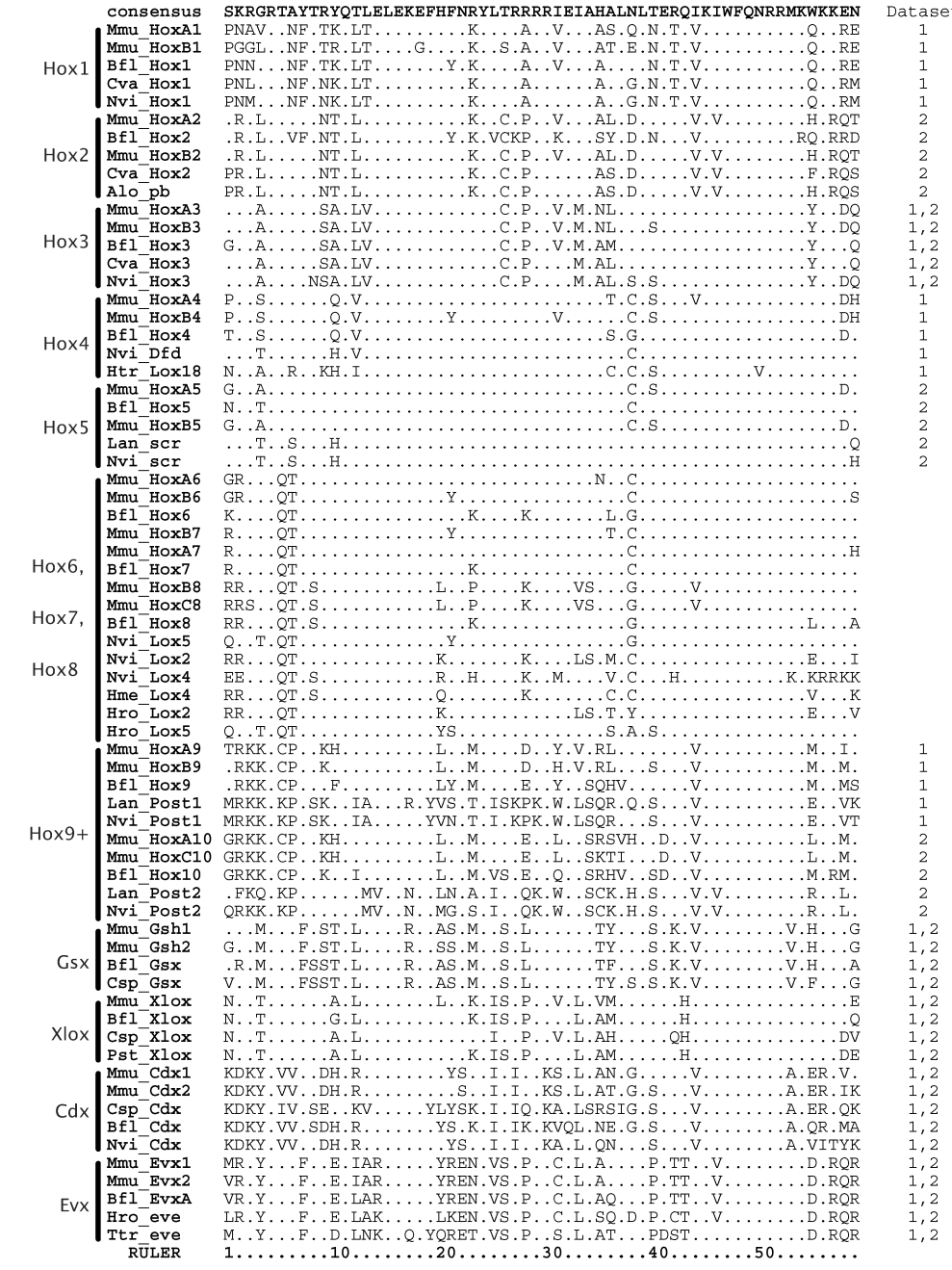



FIGURE 3. Amino acid homeodomain alignment. The sequences included in datasets 1 and 2 (see Materials and Methods) are indicated in the *Dataset* column. Abbreviations: Mmu, *Mus musculus*; Bfl, *Branchiostoma floridae*; Pst, *Phascolion strombi*; Lan, *Lingula anatina*; Csp, *Capitella species*; Cva, *Chaetopterus variopedatus*; Nvi, *Nereis virens*; Htr, *Helobdella triserialis*; Hro, *Helobdella robusta*; Ttr, *Theromyzon trizonare*; Hme, *Hirudo medicinalis*; Alo, *Archegozetes longisetosus*. For accession numbers see Appendix 2.

hypothesis (e.g., $TD_{ML}$ for the tandem duplication hypothesis). We then chose the hypothesis with the best ML tree and compared each of the other seven hypotheses' ML trees to the best ML tree using a parametric bootstrapping technique. This technique is a variant of the SOWH (Swofford-Olsen-Waddell-Hillis; Swofford et al., 1996; Goldman et al., 2000) test, which we refer to as the SOWH-like (SOWHL) test. The SOWHL test uses simulated molecular data to establish a null distribution against which to test the difference in likelihood between two different phylogenetic trees, given the same alignment. That is, it is a test of whether the tree with the higher likelihood score is a significantly better description of the data than the tree with the lower likelihood score. In effect, the test treats the tree with the lower likelihood score as the null hypothesis (Goldman et al., 2000). The a posteriori selection of trees using maximum-likelihood makes this test slightly different from the true

Downloaded By: [Australian National University] At: 23:40 13 October 2008

Downloaded By: [Australian National University] At: 23:40 13 October 2008

SOWH test (sensu Goldman et al., 2000), hence the use of the SOWHL acronym. However, this modification does not violate any assumptions of the test.

For example, if the best ML tree corresponded to the 3HL hypothesis (as is the case in some analyses), we compared the ML trees for all other hypotheses to the ML tree for the 3HL hypothesis ($3HL_{ML}$). For example, to test whether the difference in likelihoods between the $3HL_{ML}$ and $4PL_{ML}$ topologies was significant, we simulated 1000 replicate datasets in PAML (Yang, 1997) along the tree with the inferior likelihood score (the $4PL_{ML}$ topology in this case) using the model and parameters (gamma-shape parameter and branch lengths) estimated from the original data. For each simulated dataset, we then calculated the likelihood of each topology (i.e., $3HL_{ML}$and $4PL_{ML}$) and the difference in likelihood between the two topologies. The significance of the observed difference in likelihood was assessed by comparison with a ranked list of the simulated differences. For instance, if the observed difference in likelihood between the two topologies ranked 30th in the list of 1000 simulated differences, then the $P$ value was calculated as $30/1000 = 0.03$.

For each set of SOWHL tests, we compared the best of the eight hypotheses' ML trees to the ML trees from the other seven hypotheses. This allowed us to ascertain whether the best ML tree is a significantly better description of the data than ML trees from the other seven hypotheses. In the following sections we first describe the data used in the SOWHL tests, subsequently we describe the way in which we generated all trees consistent with each hypothesis, and finally we describe the methods used to calculate likelihood values.

*Data*.—Because the computational burden of the SOWHL tests is extremely high (each test requires 2000 maximum-likelihood computations) it was not feasible to use the entire dataset of 68 homeodomain sequences for each test. Consequently we used a targeted dataset (which we call Data Set 1) of 38 ingroup sequences, comprising *Hox1*, *Hox3*, *Hox4*, *Hox9+*, *Cdx*, *Xlox*, and *Gsx* orthologues (each orthology group is shown in Fig. 3 by vertical bars).

This dataset represents all four clades of Hox and ParaHox genes that have been consistently found in previous analyses (anterior, group 3, central, and posterior). However, we rely on the assumptions that *Hox1* is sufficient to represent the anterior Hox genes, *Hox4* is sufficient to represent the central Hox genes, and that *Hox9* is sufficient to represent the posterior Hox genes. Although these assumptions are not controversial, we further test them by repeating all SOWHL tests with a second dataset (which we call Data Set 2), which replaces the *Hox1*, *Hox4*, and *Hox9* sequences with *Hox2*, *Hox5*, and *Hox10* sequences, respectively.

*Construction of phylogenetic trees*.—In order to calculate the ML tree for each of the eight hypotheses, we first constructed the set of all possible phylogenetic trees consistent with each hypothesis as follows (see Fig. 4). First, we assumed that all orthologues are monophyletic with respect to all other orthologues (Fig. 4A). Second, we assumed that the relationships within each orthology group (e.g., the relationships among the *Hox1* genes from different taxa) match the known phylogenetic relationships among the taxa represented in that group (Fig. 4B). Third, we enumerated all combinations of relationships between orthologous groups that were consistent with each hypothesis (Fig. 4C). For instance, the 2NL hypothesis requires that the Hox and ParaHox genes be organized into two monophyletic sister groups, each containing at least one Hox and one ParaHox orthologue. For

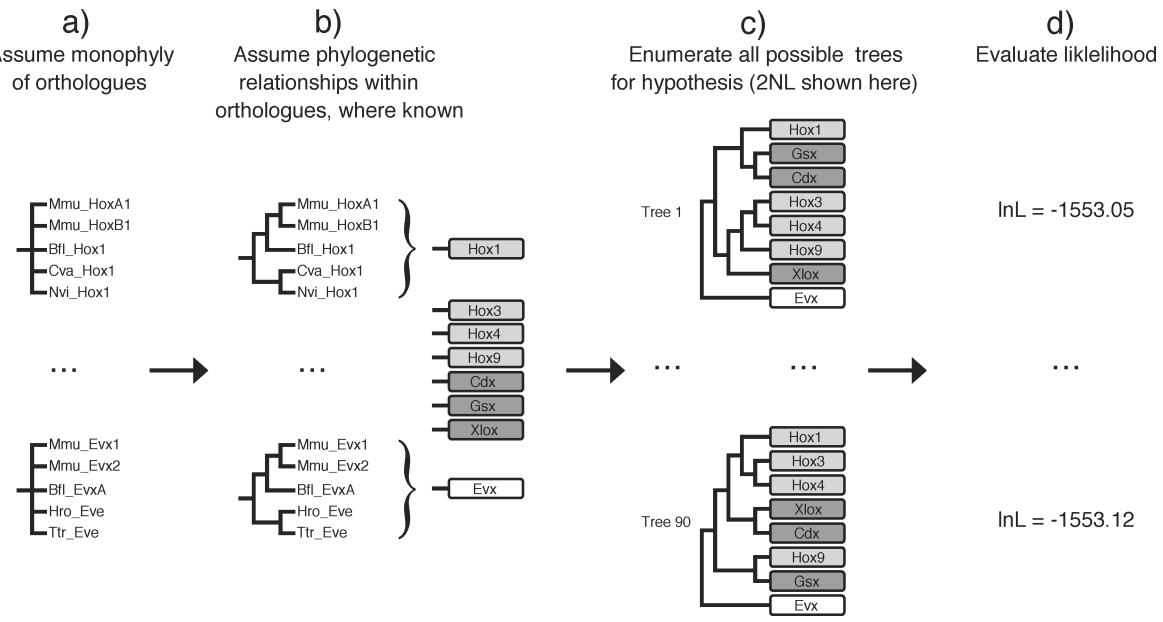



FIGURE 4. Method by which the maximum-likelihood tree is calculated for each hypothesis, in order to perform the SOWH-like tests.

Downloaded By: [Australian National University] At: 23:40 13 October 2008

each hypothesis, the likelihood of all possible trees consistent with that hypothesis was calculated (Fig. 4D) and the tree with the highest likelihood score (e.g., $2NL_{ML}$) was used for further analysis.

For the TD hypothesis, we began by considering all trees in which Gsx is basal, a key requirement of the hypothesis as stated by Ryan et al. (2007). From this set we selected the highest likelihood tree that conformed to the TD hypothesis. A tree was defined as agreeing with the TD hypothesis if the sequence of bifurcations in the tree was consistent with the generation of one of the gene orders shown in Figure 2 by the sequential tandem duplication of genes in a linked cluster. Although other arrangements of the hypothetical Hox/ParaHox cluster are possible, here we consider the two which are consistent with the hypothesis as proposed by Ryan et al. (2007).

*Likelihood calculation*.—Likelihoods of phylogenetic trees were calculated in AAML (part of the PAML [Yang, 1997] package). Because the accuracy of the SOWHL test is influenced by the model of substitution, SOWHL tests were repeated under two commonly used amino acid replacement matrices: WAG (Whelan and Goldman, 2001) and JTT (Jones et al., 1992). All likelihood calculations allowed for gamma-distributed rates across sites, the inclusion of which was determined, along with other aspects of the model, using the second-order Akaike information criterion (Hurvich and Tsai, 1989); see Appendix 1 (www.systematicbiology.org).

### *Method 2: Bayesian Tests*

In this approach we use a Markov chain Monte Carlo (MCMC) method to derive an estimate of the posterior sample of phylogenetic trees, from which the posterior probability of different hypotheses of Hox and ParaHox gene evolution is inferred. We first calculated the posterior distribution of trees for the full alignment of 68 homeodomains. Each tree in the posterior distribution was then classified in terms of the hypothesis (or hypotheses) of Hox and ParaHox gene evolution that it supported. The posterior probability of a given hypothesis was then calculated as the sum of all the posterior probabilities of the trees supporting that hypothesis. This quantity represents the proportion of time the MCMC spent visiting a given hypothesis of Hox and ParaHox gene evolution. The posterior probability was then compared to the prior probability, which represents the amount of time the MCMC would be expected to visit a given hypothesis if all trees were equally likely. We compared support for different hypotheses using the Bayes factor (BF), calculated as follows:

$$BF_{ij} = 2\ln\left(\frac{P(M_i|D)/P(M_j|D)}{P(M_i)/P(M_j)}\right)$$

where $BF_{ij}$ is the Bayes factor of hypothesis $i$ compared to hypothesis $j$, $P(M_i|D)$ is the posterior probability of hypothesis $i$, and $P(M_i)$ is the prior probability of hypothesis $i$. A negative Bayes factor represents support for model $j$, and a positive Bayes factor represents support for model $i$, on the following scale: 0 to 2 shows very weak evidence; 2 to 5 represents positive evidence; 5 to 10 represents strong evidence; and >10 represents very strong evidence (Raftery, 1996). Although this scale is useful and widely used as a rule of thumb for comparing hypotheses in a Bayesian framework, it should be pointed out that we know of no solid theoretical underpinning for its derivation. We therefore present not only Bayes factors (Table 3) but also prior and posterior probabilities of each hypothesis in the results (Table 2). We use the Bayes factor in two ways: to compare the support for hypothesis $i$ versus all alternative hypotheses (i.e., hypothesis $j$ is the hypothesis that $i$ is not true, and hypothesis $j$ therefore contains all possible phylogenetic trees that do not correspond to hypothesis $i$); and to directly compare the support for different hypotheses (i.e., where hypotheses $i$ and $j$ are two different hypotheses of Hox and ParaHox gene evolution).

*Unconstrained Bayesian analysis*.—The full dataset of 68 homeodomains was analyzed in MrBayes version 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Analyses had one cold and one heated chain, and samples were collected at 50-generation intervals. The analysis consisted of two independent runs of 13,000,000 generations each, of which the first 500,000 generations were discarded as burn-in. These parameters were determined from previous runs of MrBayes using the same dataset, which indicated that the MCMC found the global optimum after <100,000 generations, and that the two independent runs converged after around 3,000,000 generations (assessed as the point at which the standard deviation of split frequencies fell below 0.01). The posterior distribution of phylogenetic trees was then estimated from the complete 500,000 samples (250,000 samples from each of two runs). As before, the analysis was repeated using both the WAG (Whelan and Goldman, 2001) and JTT (Jones et al., 1992) amino acid replacement matrices. All analyses were checked for convergence using the program Tracer version 1.3 (Rambaut and Drummond, 2003).

*Tree classification*.—Each tree in the sample of the posterior distribution of trees was first rooted using the outgroup sequences (Evx) and then classified by the minimum possible number of genes in the hypothetical ProtoHox cluster and the number of subsequent gene losses necessary to explain the observed branching pattern in the tree. For instance, if a phylogenetic tree contained three monophyletic pairs of Hox and ParaHox genes, with no Hox or ParaHox genes lying outside these pairs, then it was classified as supporting a three-gene ProtoHox cluster with no gene losses (3NL, Fig. 1).

Because the TD hypothesis makes specific predictions about the relative order of appearance of different Hox and ParaHox genes, it was necessary to make assumptions about the orthology relationships among different Hox and ParaHox sequences from different species (e.g., that all Hox1 sequences are orthologous to one another). To do this, we followed the scheme of de Rosa et al.

Downloaded By: [Australian National University] At: 23:40 13 October 2008

(1999), which has been widely confirmed in subsequent analyses, in assuming that all sequences labeled as *Hox1*, *Hox2*, *Hox3*, *Hox4*, *Hox5*, *Cdx*, *Gsx*, and *Xlox* in Figure 1 are orthologous within those groups (these assumptions are presented graphically as vertical bars in Fig. 3). With the other central and posterior Hox genes (*Hox6* to *Hox8* and *Hox9+*, respectively), the orthology relationships between genes from different taxa are not known with confidence, we therefore make the conservative assumption that all remaining central Hox genes form one large orthology group (*Hox6* to *Hox8*, Fig. 3) and that all of the posterior Hox genes form another orthology group (*Hox9+*, Fig. 3). Thus we infer that a tree supports the TD hypothesis if both of the following conditions are met: all of the 10 orthology groups outlined above must be recovered as monophyletic, and the sequence of bifurcations between these ten orthology groups must be consistent with the generation of one of the two possible orderings of Hox and ParaHox genes in the hypothetical Hox/ParaHox cluster (Fig. 2) by single-gene tandem duplication events. All tree classification was carried out using a Python script, which is available from the authors.

*Calculation of the prior probability of each hypothesis.*—The prior probability of each hypothesis represents the proportion of time the MCMC would be expected to spend visiting that hypothesis if all phylogenetic trees were equally likely; that is, we use a phylogenetically uninformative prior. In this analysis, we have 63 Hox and ParaHox gene sequences (the other 5 sequences are outgroups used to root the trees), and therefore $1.04 \times 10^{103}$ possible rooted phylogenetic trees. The prior probability of a given hypothesis is simply the proportion of all possible trees which are consistent with that hypothesis. Although the use of phylogenetically informative priors might be preferred in some cases (e.g., to reflect our beliefs about the likely monophyly of all Hox1 sequences), we are limited here by the available software to use either uninformative priors or to constrain some parts of the topology to be monophyletic in advance (i.e., a prior of 1). Given an appropriately long burn-in period for the MCMC, the posterior probability of each hypothesis should not be significantly affected by the use of topologically uninformative priors; however, the prior probability of each hypothesis (and thus the Bayes factors used to compare hypotheses) might be affected. For this reason, we present both the posterior probability of each hypothesis (which does not explicitly include information about the prior) and the support for each hypothesis (which does) in Table 2. If differences in the posterior probabilities of hypotheses are comparable to differences in the support for hypotheses (as is the case here, see Table 2), this indicates that the use of topologically uninformative priors has not unduly affected calculation of the prior probabilities of the hypotheses.

Prior probabilities of hypotheses were calculated as follows. First, we calculated all possible ways of organizing the Hox and ParaHox genes into the number of clades required by that hypothesis. For instance, the 2NL hypothesis requires that both the Hox and ParaHox genes be organized into two distinct clades. Second, when any one of these clades contained more than two sequences, we calculated all possible ways of arranging the sequences into rooted subtrees. Third, we calculated all possible ways of resolving any polytomies associated with the hypothesis tree (Fig. 1). Finally, we calculated all possible ways of arranging the required number of clades of Hox and ParaHox genes (see Fig. 1) with respect to one another. The total number of trees consistent with a given hypothesis is simply the product of the four numbers calculated above. All calculations were performed using a Python script, which is available from the authors.

The total number of trees consistent with the tandem duplication hypothesis could not be calculated analytically. Although the combinatorial mathematics exists to calculate numbers of phylogenetic trees consistent with gene duplication scenarios (Gascuel et al., 2003), ours is a special case where we are interested simultaneously in two possible gene orders (shown in Fig. 2). Because of the requirement in the TD hypothesis that all orthology groups be recovered as monophyletic (see above), we estimated the total number of possible trees consistent with the TD hypothesis by calculating the number of such trees in a randomly generated sample 3,500,000 rooted phylogenetic trees of the 10 orthology groups (representing around 10% of the $3.4 \times 10^7$ total possible rooted trees of 10 groups). The number of trees consistent with the TD hypothesis in our random sample of trees was then scaled up to estimate the total number of possible trees consistent with the TD hypothesis. By this method, we estimate that there are 9132 possible TD trees. The prior probability of the TD hypothesis is therefore $9132/(1.0 \times 10^{103}) = 8.8 \times 10^{-100}$.

*Constrained Bayesian analysis: dealing with low prior probabilities.*—The extremely low prior probability of observing the TD hypothesis by chance (the prior probability of the TD hypothesis is $8.8 \times 10^{-100}$) means that it is highly unlikely that trees corresponding to this hypothesis would be observed in our posterior sample of 500,000 trees. This complicates the estimation of the posterior probability for the TD hypothesis. One way to account for this problem would be to calculate the marginal likelihoods of the two hypotheses (in this case, the TD hypotheses, and the hypotheses that TD is not true), using the harmonic mean of the likelihoods of a very large number of visits to each model (Raftery, 1996; and see e.g. Pagel and Meade, 2006). A large number of visits to a given model can be generated by constraining the MCMC to visit solely that model. In our case, this would involve constraining an MCMC to visit trees consistent with the TD hypothesis. Unfortunately, the TD hypothesis cannot be represented as simple constraints trees in the conventional sense, which means that the method is impossible to implement in currently available software. As an alternative to this method, we constrained the MCMC to visit a small sample of the possible trees, which contains all possible trees consistent with the TD hypothesis. The sample is defined by a constraints tree

that assumes the monophyly of all orthology groups which we previously defined as a necessary prerequisite for a tree to support the TD hypothesis. Namely, the following groups are constrained to be monophyletic (see Fig. 1): (*Hox1*), (*Hox2*), (*Hox3*), (*Hox4*), (*Hox5*), (*Hox6*, *Hox7*, *Hox8*), (*Hox9+*), (*Cdx*), (*Xlox*), (*Gsx*). The prior probability of observing a tree consistent with the TD hypothesis in this constrained analysis is $2.6 \times 10^{-4}$, thus we would expect to observe over 130 such trees in our posterior sample of 500,000 trees. In this way, the posterior probability of the TD hypothesis can be calculated with more confidence. Constrained analyses were run in MrBayes3.1.2, as before.

### *Model Testing*

In all cases, the fit of a particular model (either JTT or WAG) to a particular dataset (Data Set 1, Data Set 2, or the complete dataset comprising all sequences shown in Fig. 3) was assessed using the Goldman-Cox test (Goldman, 1993). For each Goldman-Cox test, we first estimated the ML topology for the dataset and model in question using PhyML (Guindon and Gascuel, 2003; Guindon et al., 2005). We then calculated both the unconstrained (i.e., model-free) and constrained (i.e., assuming a particular model) likelihoods for this topology, as well as the free parameters of the model, using PAML (Yang, 1997). We then simulated 200 replicate datasets in PAML (Yang, 1997) along the ML topology, using the model and free parameters estimated from the original data. For each replicate dataset, we recalculated the difference between the unconstrained and constrained likelihood scores. The observed difference was then compared to a ranked list of the simulated differences. If the observed difference falls outside the 95% confidence intervals of the simulated differences, this indicates that the model is a poor fit to the given data and should be rejected.

## RESULTS

### *Method 1: SOWHL Test Results*

In total, three sets of SOWHL tests were performed. The first uses the WAG (Whelan and Goldman, 2001) amino acid replacement matrix and Data Set 1, the second uses the WAG amino acid replacement matrix and Data Set 2, and the third uses the JTT (Jones et al., 1992) amino acid replacement matrix and Data Set 1. Goldman-Cox tests indicate that the assumed model cannot be rejected in any of the three sets of tests ($P = 0.33$ for WAG and Data Set 1, $P = 0.41$ for WAG and Data Set 2, and $P = 0.59$ for JTT and Data Set 1).

All three sets of SOWHL tests indicate that we can reject the $3PL_{ML}$, $2HL_{ML}$, $2NL_{ML}$, and $2PL_{ML}$ topologies in favor of the most likely topology. Of the remaining four topologies ($TD_{ML}$, $4PL_{ML}$, $3NL_{ML}$, $3HL_{ML}$), the first set of SOWHL tests indicates that there is no significant difference between the likelihoods of these topologies, whereas the other two sets of SOWHL tests indicate that the $3HL_{ML}$ topology is significantly preferred over all other topologies (Table 1). If a Bonferroni correction for multiple tests is applied (giving a threshold $P$ value of 0.00714 after correcting for seven independent tests), the $3PL_{ML}$ topology is rejected only when using the JTT matrix (all other results remain unchanged).

TABLE 1. Results of Method 1: SOWHL tests. Each row shows the results of a single SOWHL test. For each hypothesis, the likelihood of the best tree from all possible trees is shown in –lnL. The hypothesis with the best overall tree is highlighted in bold. $\Delta_{observed}$ is the likelihood difference between each tree and the best overall tree. For each set of SOWHL tests, the best overall tree was compared to the ML trees from each of the other seven hypotheses (indicated by $P$ values) but not to itself (indicated by n/a). $P$ values significant at the 5% level are highlighted in bold.

| Model and dataset | Hypothesis ML tree | −lnL | $\Delta_{observed}$ | $P$ |
|---|---|---|---|---|
| WAG Data Set 1 | $4PH_{ML}$ | −1553.2 | 0.6 | 0.127 |
| | $3HL_{ML}$ | −1552.8 | 0.1 | 0.053 |
| | $3NL_{ML}$ | −1553.7 | 1.1 | 0.119 |
| | $3PL_{ML}$ | −1555.9 | 3.2 | **0.011** |
| | $2HL_{ML}$ | −1557.2 | 4.5 | **0.003** |
| | $2NL_{ML}$ | −1556.7 | 4.0 | **0.003** |
| | $2PL_{ML}$ | −1562.1 | 9.4 | **<0.001** |
| | **$TD_{ML}$** | **−1552.7** | **0.0** | n/a |
| WAG Data Set 2 | $4PH_{ML}$ | −1460.6 | 5.1 | **0.002** |
| | **$3HL_{ML}$** | **−1455.5** | **0.0** | n/a |
| | $3NL_{ML}$ | −1462.0 | 6.6 | **<0.001** |
| | $3PL_{ML}$ | −1459.3 | 3.8 | **0.009** |
| | $2HL_{ML}$ | −1463.2 | 7.7 | **<0.001** |
| | $2NL_{ML}$ | −1460.5 | 5.1 | **0.003** |
| | $2PL_{ML}$ | −1466.5 | 11.1 | **<0.001** |
| | $TD_{ML}$ | −1458.6 | 3.1 | **0.002** |
| JTT Data Set 1 | $4PH_{ML}$ | −1543.7 | 1.9 | **0.008** |
| | **$3HL_{ML}$** | **−1541.8** | **0.0** | n/a |
| | $3NL_{ML}$ | −1544.0 | 2.2 | **0.013** |
| | $3PL_{ML}$ | −1545.2 | 3.4 | **0.007** |
| | $2HL_{ML}$ | −1546.5 | 4.6 | **0.003** |
| | $2NL_{ML}$ | −1546.5 | 4.6 | **<0.001** |
| | $2PL_{ML}$ | −1548.5 | 6.7 | **<0.001** |
| | $TD_{ML}$ | −1542.5 | 0.6 | **0.041** |

This indicates that the power of the SOWHL test to distinguish between different hypotheses of Hox and ParaHox gene evolution is somewhat dependent on the data used and the model of evolution assumed. The use of Data Set 1 with the WAG model of amino acid substitution gives the poorest resolution, but changing either the dataset (to Data Set 2) or the model of evolution (to the JTT model) increases the ability of the test to distinguish between models.

It is difficult in this case to find convincing reasons to prefer the use of one dataset over another or the use of one model of amino acid substitution over another. For instance, the WAG model of amino acid substitution might be preferred because it was derived using a maximum-likelihood approach, rather than a maximum parsimony approach as is the case for the JTT model (Whelan and Goldman, 2001). The WAG model is therefore less likely to underestimate parameters in the resultant amino acid substitution rate matrix, and it is likely that this is what underlies the difference in results between the WAG and JTT results shown here. However, even when using the WAG model, the ability of the SOWHL test to resolve different hypotheses of Hox and ParaHox gene evolution is still dependent on the dataset used; furthermore, Goldman-Cox tests indicate that neither the WAG nor

Downloaded By: [Australian National University] At: 23:40 13 October 2008

TABLE 2. Results of Method 2: Bayesian prior and posterior probabilities of the eight hypotheses of Hox and ParaHox gene evolution shown in Figure 1. Results are shown for both the unconstrained and constrained Bayesian analyses. Posterior probabilities were calculated under two different models of amino acid evolution—the JTT and WAG replacement matrices. Abbreviations for hypotheses are as in Figure 1. Support for each hypothesis is indicated as a Bayes factor, in which hypothesis *i* is the hypothesis in the given row, and hypothesis *j* is the hypothesis that *i* is false (i.e., hypothesis *j* includes all possible phylogenetic trees that do not belong to hypothesis *i*). Where the posterior probability is zero, support is calculated assuming that the posterior probability is <(1/500,000); i.e., that less than one tree in the entire posterior sample of trees was concurrent with that hypothesis.

| | | | JTT | | WAG | |
|---|---|---|---|---|---|---|
| Constrained | Hypothesis | Prior | Posterior | Support | Posterior | Support |
| No | 4PL | $2.3 \times 10^{-14}$ | $2.6 \times 10^{-05}$ | 41.7 | $2.3 \times 10^{-04}$ | 46.0 |
| No | 3NL | $3.6 \times 10^{-15}$ | $4.0 \times 10^{-06}$ | 41.7 | $1.1 \times 10^{-04}$ | 48.3 |
| No | 3PL | $7.5 \times 10^{-15}$ | $0.0 \times 10^{+00}$ | <38.8 | $6.0 \times 10^{-06}$ | 41.0 |
| No | 3HL | $7.2 \times 10^{-15}$ | $1.0 \times 10^{-04}$ | 46.7 | $1.3 \times 10^{-04}$ | 47.2 |
| No | 2NL | $1.7 \times 10^{-15}$ | $0.0 \times 10^{+00}$ | <41.8 | $0.0 \times 10^{+00}$ | <41.8 |
| No | 2PL | $1.7 \times 10^{-15}$ | $0.0 \times 10^{+00}$ | <41.8 | $0.0 \times 10^{+00}$ | <41.8 |
| No | 2HL | $1.7 \times 10^{-15}$ | $0.0 \times 10^{+00}$ | <41.8 | $0.0 \times 10^{+00}$ | <41.8 |
| No | TD | $8.8 \times 10^{-100}$ | $0.0 \times 10^{+00}$ | <429.9 | $0.0 \times 10^{+00}$ | <429.9 |
| Yes | TD | $2.7 \times 10^{-04}$ | $1.6 \times 10^{-01}$ | 12.8 | $1.3 \times 10^{-01}$ | 12.3 |

Downloaded By: [Australian National University] At: 23:40 13 October 2008

the JTT model can be rejected for the datasets used. It therefore seems prudent in this case to take the conservative view and to reject only those hypotheses which were consistently rejected in all SOWHL tests; i.e., the 3PL, 2HL, 2NL and 2PL hypotheses.

### *Method 2: Bayesian Test Results*

All results of the Bayesian analyses are shown in Table 2 and the Bayes factors comparing support for different hypotheses are shown in Table 3. Bayes factors comparing different hypotheses were similar in the Bayesian runs which assumed different models of protein evolution (either JTT or WAG), so for brevity Table 3 shows only the results from the runs which assumed the WAG replacement matrix. As with the SOWHL tests, Goldman-Cox tests indicate that neither the WAG nor the JTT model can be rejected when applied to the complete dataset ($P = 0.55$ and $P = 0.64$, respectively).

*Unconstrained Bayesian analysis*.—In the unconstrained Bayesian analyses, the support for each hypothesis is somewhat sensitive to the assumption of a particular model of evolution (compare the "support" for JTT and WAG, Table 2); however, the relative support for different models is largely similar. Bayes factor (BF) values comparing those hypotheses (4PL, 3NL, 3HL, 3PL) that had posterior probabilities greater than zero are shown in Table 3. These BF values show that the 3PL hypothesis can be confidently rejected: all BF values in the 3PL column are negative and less than 5, indicating strong evidence against this hypothesis when compared to each other hypothesis. Comparison of the BF values of the remaining hypotheses indicates that it is not possible to distinguish between the 4PL, 3NL, and 3HL hypotheses, as no single hypothesis shows strong support over the other two.

TABLE 3. Bayes factors comparing different models of Hox and ParaHox gene evolution (unconstrained analysis). Positive values indicate support for the model *i*, whereas negative values indicate support for model *j*. Posterior odds used to calculate Bayes factors were derived from an unconstrained Bayesian MCMC using the WAG amino acid replacement matrix.

| | Model *i* | | | |
|---|---|---|---|---|
| | 4PL | 3NL | 3PL | 3HL |
| Model *j* | | | | |
| 4PL | — | 2.3 | –5.0 | 1.2 |
| 3NL | –2.3 | — | –7.3 | –1.0 |
| 3PL | 5.0 | 7.3 | — | 6.2 |
| 3HL | –1.2 | 1.0 | –6.2 | — |

Trees supporting some hypotheses (namely the 2NL, 2HL, 2PL, and TD hypotheses) were not observed in the sample of the posterior distribution of topologies, regardless of the model of evolution assumed. The posterior probability of these hypotheses is therefore zero. In these cases, a bound was placed on the support level by calculating the support that would have resulted if one instance of each hypothesis had been observed in the posterior sample of topologies (i.e., the bound of the posterior probability is 1/500,000). For instance, in the case of the two-ProtoHox hypotheses (2NL, 2HL, and 2PL), it can be shown that the support must be less than 41.8 in each case (Table 2). This approach can also be used to place bounds on the BF values comparing the two-ProtoHox hypotheses to other hypotheses of Hox and ParaHox gene evolution.

Placing bounds on the BF values for hypotheses with posterior probabilities of zero suggests that there is almost always strong evidence against the two-ProtoHox hypotheses (2NL, 2HL, and 2PL) in favor of alternative hypotheses: in the JTT analysis the 2NL, 2HL, and 2PL hypotheses would be rejected in favor of the 3HL hypothesis (BF > 4.9), and in the WAG analysis they would be rejected in favor of the 4PL (BF > 4.2), 3NL (BF > 6.5), and 3HL (BF > 5.4) hypotheses. We can therefore confidently reject the 2NL, 2HL, and 2PL hypotheses in favor of other hypotheses.

*Constrained Bayesian analysis*.—Unfortunately, placing bounds on the BF value is uninformative for the TD hypothesis, as knowledge of the support value is restricted to the observation that it must be below 429.9, which precludes meaningful comparison with other values

(Table 2). In the constrained Bayesian analysis, the TD hypothesis is shown to have some support. However, the value of the support for the TD hypothesis (12.8 using the JTT model, and 12.3 using the WAG model) in the constrained analysis is far lower than that achieved for many other hypotheses in the unconstrained analyses (4PL, 3NL, 3PL, 3HL, which have support >40). The comparison of these values suggests that the TD hypothesis has considerably less support from the data than any of the 4PL, 3NL, 3PL, or 3HL hypotheses.

In summary, the comparison of support values for all of the Bayesian analyses indicates that five of the eight hypotheses tested can be rejected in favor of the 4PL, 3NL, and 3HL hypotheses. Comparison of the support values for the 4PL, 3NL, and 3HL hypotheses using BF values indicates that it is not possible to reject any of these three hypotheses in favor of any others of the three.

Downloaded By: [Australian National University] At: 23:40 13 October 2008

## DISCUSSION

There are a number of viable hypotheses in the literature concerning the early evolution of the Hox and ParaHox clusters. The results from the ML analysis presented here show that the best overall tree recovered (and the hypothesis of Hox and ParaHox gene evolution that tree supports) was sensitive to small adjustments in both the model of evolution and the dataset used. These observations go some way to explaining the large number of hypotheses of Hox and ParaHox gene evolution currently discussed in the literature.

In this study, we compared the support for different hypotheses of Hox and ParaHox gene evolution directly, using two contrasting statistical methods (one ML and one Bayesian) that account for phylogenetic uncertainty in different ways. In the ML approach, we generated all possible trees consistent with eight different hypotheses of Hox and ParaHox gene evolution (Fig. 1) and assessed the likelihood of each of these trees. We then compared hypotheses by comparing the ML trees from different hypotheses using a parametric bootstrap. In the Bayesian analysis, we assessed the degree to which each hypothesis was over- or underrepresented in the posterior distribution of trees, and compared hypotheses using Bayes factors.

The results of both the ML and Bayesian analyses are largely congruent. In the ML analyses, results of the statistical tests between topologies were consistent regardless of assumptions made. In particular, the results indicate that the 3PL, 2HL, 2NL, and 2PL hypotheses are consistently rejected regardless of the model of evolution assumed or the dataset used. Similarly, in the Bayesian analysis, the results allow us to reject the 3PL, 2HL, 2NL, 2PL, and TD hypotheses in favor of the 4PL, 3HL, and 3NL hypotheses.

The most conservative interpretation of these results is to reject only those hypotheses that were consistently rejected in all tests; i.e., the 3PL, 2HL, 2NL, and 2PL hypotheses. In this case, we would conclude that there is sufficient phylogenetic signal in the homeodomain to reject the all hypotheses in which the ProtoHox cluster contained two genes and the hypothesis that the ProtoHox cluster contained three genes with a subsequent loss of one ParaHox gene. However, the case of the TD hypothesis is difficult—it was rejected in two out of three SOWHL tests, and it was rejected in the Bayesian analysis. We believe that the failure to reject the TD hypothesis in one of the SOWHL tests is due to the inherent assumption of the SOWHL test that one particular topology (in this case, the ML topology) is sufficient to represent a given hypothesis. A number of observations point to this not being the case for our data. The results from the Bayesian analyses (not shown) indicate that the tree with the highest posterior probability in any analysis was visited a total of nine times in 500,000 samples of the MCMC and that the majority of trees in sample of the posterior distribution were visited only once. These results indicate that there is no overwhelming support for any one tree in the analysis, and therefore that a given ML tree is unlikely to be a fair representation of a given hypothesis. That is, the phylogenetic uncertainty in the data may well preclude the assumption of any one particular tree as a reasonable representation of the true state of affairs. The Bayesian analysis does not rest on the assumption that one particular tree is the best or representative tree for a given hypothesis. Rather, it explicitly allows all possible trees for each hypothesis to contribute to the support for that hypothesis, essentially integrating over the phylogenetic uncertainty in the data. We therefore suggest that the rejection of the TD hypothesis in two of the three SOWHL tests and in the Bayesian analysis is sufficient evidence on which to reject the TD hypothesis.

We conclude that despite the very limited sequence data available to assess hypotheses of Hox and ParaHox gene evolution, it is nevertheless possible to make inferences about evolutionary processes that occurred over 500 million years ago. We are able to reject five out of eight hypotheses of Hox and ParaHox gene evolution, and conclude that the ProtoHox cluster did indeed exist and that it contained either three or four genes before it duplicated to form the Hox and ParaHox clusters.

The inability to distinguish between three different hypotheses of Hox and ParaHox gene evolution (4PL, 3NL, and 3HL) might be remedied by the use of a larger dataset of Hox and ParaHox genes than is used here; however, we think this is unlikely. Rather, it seems plausible that the phylogenetic signal contained in the sixty amino acid homeodomains is simply insufficient to distinguish between certain similar hypotheses of Hox and ParaHox gene evolution, largely as a result of the degradation of the phylogenetic signal over 500 million years of evolutionary time. That is, the inability to distinguish between certain hypotheses is probably not down to a lack of power but rather down to a lack of reliable signal. One approach to distinguishing the three remaining hypotheses is to consider the number of gene losses and gene duplications that they require. In this case, one might prefer the 4PL and 3NL hypotheses (which require two loss/duplication events each) to the 3HL hypothesis (which requires four loss/duplication events; see Fig. 1). Another possible approach to extract more of the

remaining phylogenetic signal from the homeodomains would be to develop evolutionary models that are specifically tailored to the homeodomain. The models used here were estimated using extremely large datasets of proteins and therefore represent a rather coarse-grained view of the probability of one amino acid being replaced by another over evolutionary time. Although elegant approaches exist for the inference of tailored models of amino acid replacement (Lartillot and Philippe, 2004) their application here is again hampered by the very limited sequence data available.

## ACKNOWLEDGMENTS

We thank Mike Hendy for help with combinatorial mathematics, Brett Calcott for help with programming, and John Welch and Jessica Thomas for comments on an earlier draft of the manuscript. R.L. is funded by a BBSRC doctoral research grant.

Downloaded By: [Australian National University] At: 23:40 13 October 2008

## REFERENCES

Brooke, N. M., J. Garcia-Fernandez, and P. W. H. Holland. 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature 392:920–922.

Budd, G. E. 1999. Does evolution in body patterning genes drive morphological change—Or vice versa? Bioessays 21:326–332.

Chourrout, D., F. Delsuc, P. Chourrout, R. B. Edvardsen, F. Rentzsch, E. Renfer, M. F. Jensen, B. Zhu, P. de Jong, R. E. Steele, and U. Technau. 2006. Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. Nature 442:684–687.

de Rosa, R., J. K. Grenier, T. Andreeva, C. E. Cook, A. Adoutte, M. Akam, S. B. Carroll, and G. Balavoine. 1999. Hox genes in brachiopods and priapulids and protostome evolution. Nature 399:772–776.

Ferrier, D. E. K., and P. W. H. Holland. 2001. Ancient origin of the Hox gene cluster. Nat. Rev. Genet. 2:33–38.

Ferrier, D. E. K., and C. Minguillon. 2003. Evolution of the Hox/ParaHox gene clusters. Int. J. Dev. Biol. 47:605–611.

Finnerty, J. R., and M. Q. Martindale. 1999. Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. Evol. Dev. 1:16–23.

Garcia-Fernandez, J. 2005. Hox, ParaHox, ProtoHox: Facts and guesses. Heredity 94:145–152.

Gascuel, O., M. D. Hendy, A. Jean-Marie, and R. McLachlan. 2003. The combinatorics of tandem duplication trees. Syst. Biol. 52:110–118.

Gellon, G., and W. McGinnis. 1998. Shaping animal body plans in development and evolution by modulation of Hox expression patterns. Bioessays 20:116–125.

Goldman, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36:182–198.

Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49:652–670.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Guindon, S., F. Lethiec, P. Duroux, and O. Gascuel. 2005. PHYML online—A Web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33:W557–W559.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Hughes, C. L., and T. C. Kaufman. 2002. Hox genes and the evolution of the arthropod body plan. Evol. Dev. 4:459–499.

Hurvich, C. M., and C. L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika 76:297–307.

Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8:275–282.

Kourakis, M. J., and M. Q. Martindale. 2000. Combined-method phylogenetic analysis of Hox and paraHox genes of the metazoa. J. Exp. Zool. 288:175–191.

Lartillot, N., and H. Philippe. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21:1095–1109.

Martinez, P., and C. T. Amemiya. 2002. Genomics of the HOX gene cluster. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 133:571–580.

Minguillon, C., and J. Garcia-Fernandez. 2003. Genesis and evolution of the Evx and Mox genes and the extended Hox and ParaHox gene clusters. Genome Biol. 4:R12.

Pagel, M., and A. Meade. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. Am. Nat. 167:808–825.

Raftery, A. E. 1996. Hypothesis testing and model selection. Pages 163–188 *in* Markov chain Monte Carlo in practice. (W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds.). Chapman & Hall, London.

Rambaut, A. 1996. SE-AL: Sequence alignment editor. Oxford, Evolutionary Biology Group, University of Oxford.

Rambaut, A., and A. J. Drummond. 2003. Tracer version 1.3. Oxford, Evolutionary Biology Group, University of Oxford.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Ryan, J. F., M. E. Mazza, K. Pang, D. Q. Matus, A. D. Baxevanis, M. Q. Martindale, and J. R. Finnerty. 2007. Pre-Bilaterian origins of the Hox cluster and the Hox code: Evidence from the Sea Anemone, Nematostella vectensis. PLoS ONE 2:e153.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. Pages 407–514 *in* Molecular systematics (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinnauer, Sunderland, Massachussetts.

Wagner, G. P., C. Amemiya, and F. Ruddle. 2003. Hox cluster duplications and the opportunity for evolutionary novelties. Proc. Natl. Acad. Sci. USA 100:14603–14606.

Whelan, S., and N. Goldman. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18:691–699.

Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555–556.

*First submitted 10 January 2008; reviews returned 5 March 2008; final acceptance 6 June 2008*
*Associate Editor: Tim Collins*