Selection on Coding Regions Determined *Hox7* Genes Evolution

Mario Ali Fares,†[1] Daniela Bezemer,† Andrés Moya,*† and Ignacio Marín**

*Departamento de Genética, and †Instituto Cavanilles de Biodiversidad and Biología Evolutiva, Universidad de Valencia, Valencia, Spain

The important role of *Hox* genes in determining the regionalization of the body plan of the vertebrates makes them invaluable candidates for evolutionary analyses regarding functional and morphological innovation. Gene duplication and gene loss led to a variable number of *Hox* genes in different vertebrate lineages. The evolutionary forces determining the conservation or loss of *Hox* genes are poorly understood. In this study, we show that variable selective pressures acted on *Hox7* genes in different evolutionary lineages, with episodes of positive selection occurring after gene duplications. Tests for functional divergence in paralogs detected significant differentiation in a region known to modulate HOX7 protein activity. Our results show that both positive and negative selection on coding regions are influencing *Hox7* genes evolution.

Introduction

Vertebrate *Hox* genes encode homeodomain-containing transcription factors involved in determination of cell fates along the anterioposterior body axis (McGinnis and Krumlauf 1992; Krumlauf 1994). Multiple *Hox* genes arose very early in metazoan evolution by successive duplications that generated a cluster of related genes (reviewed in Finnerty and Martindale [1998]). In vertebrates, this cluster, originally containing 13 *Hox* genes (or 15 homeobox-containing genes, including the divergent genes *Mox* and *Evx* [see Pollard and Holland 2000]) was itself duplicated several times (reviewed in Ruddle et al. [1994]; Finnerty and Martindale [1998]). Thus, mammals have four, and some fishes, such as zebrafish, have seven different clusters (Ruddle et al. 1994; Amores et al. 1998; Málaga-Trillo and Meyer 2001; Prince 2002). However, many of the *Hox* genes were lost after these cluster duplications. Thus, there are substantial variations in the *Hox* genes present in distant vertebrate species (Amores et al. 1998; Málaga-Trillo and Meyer 2001; Prince 2002). In particular, in mammals as human or mouse, only three out of the 13 *Hox* ''paralog groups'' (defined as sets of paralogs, each one derived from one of the *Hox* genes in the original cluster) have four genes, one for each cluster (Ruddle et al. 1994).

Multiple forces may be involved in the maintenance or elimination of duplicated genes. The classical model to explain duplicates maintenance postulates that, after a period of redundancy, positive selection generates functional divergence and the emergence of a new function for one of the duplicates (Ohno 1970). However, evidence for positive selection acting on *Hox* genes has not generally been found (e.g., Hughes and Hughes 1993). Moreover, knockout mice phenotypic analyses has provided proof of extensive redundancy among *Hox* genes in a vertebrate (reviewed in St-Jacques and McMahon [1996]). Finally, it has been shown that, keeping intact the *cis*-acting regulatory sequences, the coding regions of some *Hox* genes can substitute for each other efficiently, both in vertebrates and in *Drosophila* (Greer et al. 2000; Hirth et al. 2001). These results have led to the formulation of alternative hypotheses to explain *Hox* genes maintenance (reviewed in Force et al. [1999]; Massingham, Davies, and Liò [2001]). Particularly, acquisition of differential expression patterns by subfunctionalization (Force et al. 1999) and quantitative effects, dependent on the total number of genes within a paralog group (Greer et al. 2000) have been suggested to be the main forces that explain evolutionary conservation of paralogous *Hox* genes.

Even in favorable cases, the determination of the selective regimes acting on coding regions, and most especially the detection of directional, positive selection, is a complex task that requires using both very sensitive methods of detection and the right genes as models. This work is focused on the determination of the impact of natural selection on the coding regions of *Hox7* genes. *Hox* paralog group VII is particularly suitable for this type of study. First, it is in many species the paralog group with the minimum number of genes. Thus, exceptionally, a single *Hox7* gene seems to exist in condrichthyan or actinopterygian fishes (and whether any *Hox7* genes exist in *Fugu rubripes* is unclear [Aparicio et al. 1997; Amores et al. 1998; Snell, Scemama and Stellwag 1999; Kim et al. 2000]). Also, mammals have only two *Hox7* genes (called *Hoxa-7* and *Hoxb-7*). This is a rare feature, shared only with paralog group II. Finally, the tetraploid species *Xenopus laevis* has three group VII genes, one *Hoxa-7* and two *Hoxb-7* genes (Bisbee et al. 1977). This is again an exceptional case where we can check whether positive selection occurred after a recent duplication. Selective forces were analyzed using several advanced methods. First, we explored, using maximum-likelihood methods based on a phylogenetic approach, whether variation on the selective forces acting on the whole set of *Hox7* sequences occurred. Second, we analyzed, using a similar approach, whether selection varied in different vertebrate lineages. Third, we confirmed the results obtained in the second type of analysis by pairwise comparisons among all the *Hox7* genes. Finally, we search for regions under functional differentiation after the *Hoxa-7*/*Hoxb-7* gene duplication. Notably, our results are compatible with the classical view of maintenance of duplicated genes by functional diversification of their protein products (Ohno 1970).

[1] Present address: Department of Genetics, Trinity College, University of Dublin, Dublin, Ireland.

Key words: gene duplication, homeotic genes, homeobox, mutual information.

E-mail: ignacio.marin@uv.es.

Mol. Biol. Evol. 20(12):2104–2112. 2003
DOI: 10.1093/molbev/msg222
Molecular Biology and Evolution, Vol. 20, No. 12,
© Society for Molecular Biology and Evolution 2003; all rights reserved.

Table 1
Results of the Codon-Based Models Designed to Detect Heterogeneity Among Sequences

Model	*l*	AIC	ω	Parameters
M0	−4418.48	8838.96	0.100	K = 1.40; ω = 0.100
M1	−4556.85	9115.70	0.566	K = 1.99; $\omega_0 = 0$ ($p_0 = 0.43$); $\omega_1 = 1$ ($p_1 = 0.57$)
M2	−4321.95	8649.90	0.115	K = 1.42; $\omega_0 = 0$ ($p_0 = 0.41$); $\omega_1 = 1$ ($p_1 = 0.02$); $\omega_2 = 0.172$ ($p_2 = 0.57$)
M3	−4307.34	8624.67	0.112	K = 1.42; $\omega_0 = 0.001$ ($p_0 = 0.33$); $\omega_1 = 0.051$ ($p_1 = 0.30$); $\omega_2 = 0.258$ ($p_2 = 0.37$)
M7	−4307.66	8619.32	0.108	K = 1.40; p = 0.37; q = 2.53
M8	−4307.16	8622.32	0.112	K = 1.40; $p_0 = 0.88$; p = 0.29; q = 2.00; $P_1 = 0.12$; ω = 0.17

Materials and Methods

Sequence Analyses

The full-length nucleotide sequences used in this study were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov/). Accession numbers are as follows: *Homo sapiens* (*Hoxa-7*: XM_011610 and *Hoxb-7*: XM_008559); *Papio hamadryas* (*Hoxa-7*: AC116608 and *Hoxb-7*: AC116664); *Bos taurus* (*Hoxb-7*: AF200721); *Mus musculus* (*Hoxa-7*: NM_010455 and *Hoxb-7*: X06762); *Rattus norvegicus* (*Hoxa-7*: AABR02030143 and *Hoxb-7*: XM_220889); *Xenopus laevis* (*Hoxa-7*: M24752, *Hoxb-7a*: M23916, and *Hoxb-7b*: X06593); *Gallus gallus* (*Hoxa-7*: BU413862 + BI066006 and *Hoxb-7*: AF408695); *Coturnix coturnix* (*Hoxa-7*: M79514); *Heterodontus francisci* (*Hox7*: AF224262; in this study we have considered this gene to be a *Hoxa-7* gene, following Kim et al. [2000]); and *Danio rerio* (*Hoxb-7a*: AL645782). Those were all the full-length sequences available in the databases in May 2003. We however discarded the sequence of the *Hoxa-7* gene of the fish *Morone saxitilis* (accession number AF089743) because it was very different from the rest of sequences, and we were not able to confidently align it outside the highly conserved homeobox.

The nucleotide sequences of these genes were conceptually translated and a multiple-sequence alignment of their encoded proteins was generated using the default parameters of ClustalX version 1.83 (Thompson et al. 1997). The alignments were inspected, and minor changes made when necessary to improve alignment, using GeneDoc version 2.6 (Nicholas and Nicholas 1997). From these final protein alignments, we generated the corresponding nucleotide-sequence alignment to be used for the rest of analyses. Given the high heterogeneity in the conservation of amino acid sequences between the N-terminal and homeobox regions, an analysis of the distribution of amino acid substitutions was performed to obtain an accurate phylogenetic tree. The gamma shape parameter (α) was estimated using the program GAMMA (Gu and Zhang 1997). Thereafter, a protein phylogenetic tree based on gamma-corrected distances among sequences was inferred by the neighbor-joining (NJ) routine (Saitou and Nei 1987) available in MEGA program version 2.1 (Kumar et al. 2001). A total of 1,000 bootstrap replicates were performed under the gamma distribution model using NJ in MEGA to determine the reliability of each node of the phylogenetic tree. The highly supported tree topology obtained was the expected according to our knowledge of the evolution of *Hox* genes and phylogenetic relationships among the analyzed species. For subsequent analyses, those positions containing gaps in any of the sequences were eliminated.

Analyses of Selective Constraints Acting on *Hox7* genes

We used three different methods to determine selective constraints acting on *Hox7* genes: (1) general tests for heterogenous selective constraints among sites, (2) general tests for variable ω ratios, and (3) pairwise comparisons.

Tests for Heterogeneous Selective Constraints Among Amino Acidic Sites in Hox7 Sequences

We followed the strategies described in recent articles (e. g., Yang et al. 2000*b*), based on previous works by Nielsen and Yang (1998). To determine the selective forces acting on *Hox7* genes, six codon-based models (Yang et al. 2000*a*), implemented in version 3.0 of the PAML package (Yang 2000) were analyzed. These models explore whether, given a certain phylogenetic tree, variable selective constraints among amino acid sites are present. For each model, a maximum-likelihood (ML) approach was used to estimate ω, the ratio of nonsynonymous (d_N) to synonymous (d_S) nucleotide substitution rates per codon ($\omega = d_N/d_S$) (Nielsen and Yang 1998; Yang et al. 2000*b*). Then, the basic model of a single ω for the whole set of codons and sequences (usually called M0 [Goldman and Yang 1994]) was compared with more complex models (called M1 to M3, M7, and M8). These models are divided into two groups. M1, M2, and M3 assume discrete distributions. The ''neutral'' model 1 (M1) considers only two ω values ($\omega = 0$ and $\omega = 1$), the ''selection'' model 2 (M2) includes three classes of codons, namely those with $\omega = 0$ and $\omega = 1$ plus a third class of codons with an ω value estimated from the data, and, finally, in the ''discrete'' model 3 (M3), the number of different ω ratios that can be estimated from the data is unconstrained. For these models, p_0, p_1, and p_2 (see table

1) refer to the proportions of the different codon classes. The other models, M7 (β model) and M8 (β + ω model), assume continuous distributions for ω values. M7 does not allow for positively selected sites, whereas M8 takes into account putative positive selection at specific codon sites (Yang et al. 2000*a*). In these models, p and q refer to the parameters of the beta distribution, and p_0 refers to the proportions of sites following the beta distribution. In summary, M2, M3, and M8 may detect positively selected sites, if present.

When two models are nested, they can be compared using the likelihood ratio test (LRT). Between two nested models, twice the log-likelihood difference follows a χ^2 distribution, with a number of degrees of freedom equal to the difference in the number of free parameters between the two models. For nonnested models, the LRT cannot be used to test which model significantly better fits the data. However, their log-likelihood values can be compared by means of the Akaike Information Criterion (Akaike 1974): AIC = −2 (estimated log-likelihood of the model) +2 (number of free parameters in the model). The model that minimizes AIC is then the most appropriate.

Tests for Variable ω Ratios Among Vertebrate Lineages

ML methods were also used to establish whether selection varied among branches of the phylogenetic tree. We compared the M0 model, which assumes a single ω parameter value for the entire tree, with the ''free-ratio'' model, which accepts independent ω parameter values for each branch (Yang 1998) and with a model with four different ω values, which we have called the ''four-ratio'' model. For this last model, we assigned particular ω values for three branches that are related to duplication events. A fourth rate was assigned for all the other branches. Posterior Bayesian probabilities for codons were estimated for the free-ratio model, using the CODEML program from the PAML package.

Pairwise Comparisons of Hox7 Sequences

Maximum-likelihood estimates of d_N and d_S for pairs of sequences were obtained using the discrete models M0 and M3. A maximum of four different types of codons, according to their ω values, were allowed to be estimated by the program, but they were pooled when the program detected only two or three different classes. Number of degrees of freedom was established according to the number of estimated ω parameters. The difference of the log-likelihood values between nested models were tested by the LRT (Huelsenbeck and Crandall 1997). As above, posterior probabilities for codons were estimated in those comparisons were positive selection was detected.

Detection of Functional Divergence after the *Hoxa-7*/*Hoxb-7* Gene Duplication

Wang and Gu (2001) defined two different types of amino acidic divergence after gene duplication. Type I functional divergence refers to changes in functional constraints in one of the genes, resulting in high conservation in the sequences of different species for one paralog, while the other paralog evolves more freely. This type of divergence can be measured following the method developed by Gu (1999): a maximum-likelihood procedure is used to calculate a coefficient of functional divergence (θ), and it is determined whether such coefficient is large enough as to reject the null hypothesis of no functional differentiation. If the null hypothesis is rejected, a posterior probability for functional divergence for each position in the alignment is calculated. To implement this procedure, we used the program DIVERGE version 1.04 (Gu and Vander Velden 2002; available at http://xgu1.zool.iastate.edu/cgi-bin/download.cgi). We established a cutoff value for significance ($P = 0.6$) according to the effect that it had on the θ parameter the elimination of the sets of amino acids with values above a certain posterior probability (see Wang and Gu 2001). We found that the elimination of amino acids with $P > 0.6$, led to a value of θ that is essentially zero.

Type II divergence (Wang and Gu 2001) refers to amino acids that show gene-specific conservation; that is, although conservation is substantial within each paralog in different species, different amino acids are conserved in each of the two paralogs. To detect this type of divergence, we computed the mutual information content of our protein alignment using MatrixPlot (see Gorodkin et al. [1999]). We then established the positions in our alignment that had the maximum mutual information content when compared with those that show gene-specific amino acids (e.g., position 4 in figure 1). To establish whether the amino acidic changes detected by this procedure may have occurred immediately after the *Hoxa-7*/*Hoxb-7* duplication and before the divergence among the analyzed species, we compared the mutual information results with the posterior Bayesian probabilities for codons established according to the ''free-ratio'' model (see above) for the branch connecting *Hoxa-7* genes with *Hoxb-7* genes.

Results

The alignment of the 17 full-length *Hox7* genes analyzed and the phylogenetic tree obtained from that alignment are shown in figures 1 and 2. Given these results, six codon-based models were used to determine selective pressures on the whole set of *Hox7* sequences (table 1). Among the discrete models, M3 fits the data better than M0, M1, or M2, according to the likelihood ratio test (LRT, for nested models) or AIC criteria (for nonnested comparisons: M1/M3, M2/M3). These data suggest that heterogeneous selective pressures along whole *Hox7* sequences exist. However, positive selection on the whole set of *Hox7* genes was not found (ω, the ratio of nonsynonymous to synonymous nucleotide substitution rates per codon, had values smaller than 1 [see table 1]). Among the continuous distribution models, models M7 and M8 were equivalent. The ω value estimated under the simplest model (M7) is 0.108, again implying strong purifying selection. We can conclude that, for the whole data set, the main force acting is negative selection, with evidence for heterogeneous selection rates on different codons. For a comparison of results from table 1, see table 2.

To test for variations in selective regimes among

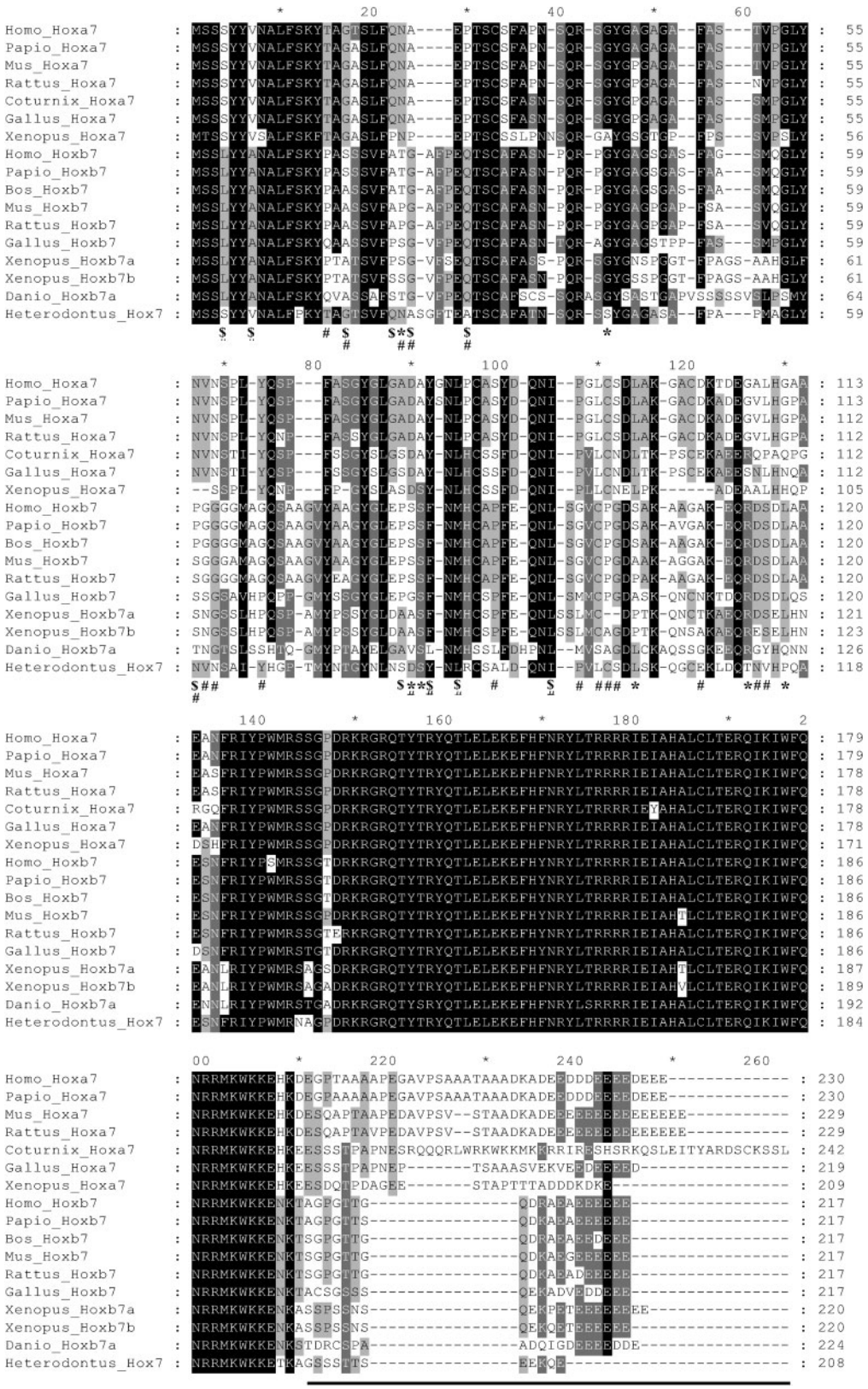
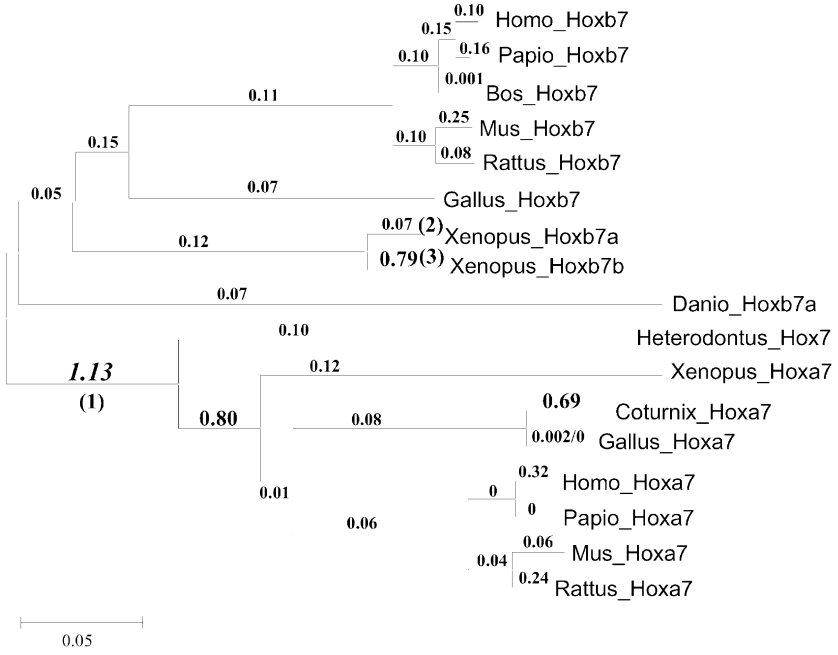


FIG. 1.—Multiple alignment of *Hox7* genes. The highly variable C-terminal region (underlined) was eliminated from subsequent analyses. From the remaining sequences, we also eliminated, when required by the programs used, the positions in which gaps were found in any sequence. Asterisks (*) refer to the positions that had posterior probabilities higher than 0.60 according to the method of detection of type I functional divergence developed by Gu (1999). A number sign (#) indicates positions with maximum mutual information content when compared with those with fixed residues in *Hoxa-7* and *Hoxb-7*. The dollar sign ($) shows positions with posterior Bayesian probabilities of at least 0.9 in the branch connecting *Hoxa-7* and *Hoxb-7* genes.



Model	*l*	ω
M0	-4418.47	0.100
Four-ratio	-4401.03	(1) 0.821, (2) 0.065, (3) 0.806, (others) 0.093
Free-ratio	-4376.63	see tree above

Comparison	2Δ*l*	*d.f.*	*P*
M0 vs. Free-ratio	83.68	30	5.76×10^{-7}
Four-ratio vs. Free-ratio	48.80	27	0.006

FIG. 2.—Neighbor-Joining tree based on the alignment shown in figure 1. Numbers refer to the values of ω for each branch according to the free-ratio model. The branch that leads to the *Gallus Hoxa-7* gene was estimated to have essentially zero nonsynonymous and synonymous nucleotide substitution rates per codon, as indicated. Numbers in parentheses (1 to 3) refer to the branches for which specific values of ω were assigned in the four-ratio model. Results for comparisons among models are shown on the bottom.

evolutionary lineages, we first considered the ''free-ratio'' model, where a different ω value is assigned to each branch of the phylogenetic tree. This model significantly improved over the M0 model, with a single ω (fig. 2). Although in the free-ratio model most branches showed very low values of ω, we found values close to 1 in three cases—the external branches that lead to *Xenopus Hox-b7b* and *Coturnix Hoxa-7* genes and the branch that groups together all tetrapod *Hoxa-7* genes. Moreover, the branch that separates *Hox-b7* genes from *Hoxa-7* genes showed an ω value higher than 1, a significant evidence for positive selection after gene duplication (fig. 2). However, the free-ratio model is parameter rich and thus unlikely to give accurate estimates for all ω values (Yang et al. 2000*b*). We thus decided to test a less complex hypothesis (''four-ratio'' model, see *Materials and Methods* and figure 2, where the branches with specific values of ω are indicated). This model significantly improved over the one-ratio model, but was significantly worse than the free-ratio model (fig. 2). We conclude that the free-ratio model is the one that best fits the data. To obtain additional evidence for positive selection, we performed pairwise sequence comparisons. A total of 47 out of 136 tests were

Table 2
Relevant Comparisons of Results in Table 1

Comparison	2Δ*l*	df	*P*
M0 versus M2	193.06	2	< 0.0001
M0 versus M3	222.28	4	< 0.0001
M1 versus M2	469.80	2	< 0.0001
M7 versus M8	1	2	1

Table 3
Results of Pairwise Comparisons

	Hs HOXA7	Ph HOXA7	Mm HOXA7	Rn HOXA7	Gg HOXA7	Cot HOXA7	Xl HOXA7	Hf HOX7	Hs HOXB7	Ph HOXB7	Bt HOXB7	Mm HOXB7	Rn HOXB7	Gg HOXB7	Xl HOXB7a	Xl HOXB7b	Dr HOXB7a
Hs HOXA7		ns	ns	ns	—	0.0001	—	—	—	—	0.005	0.008	0.011	—	—	—	—
Ph HOXA7			0.028	ns	—	0.0001	—	—	0.001	0.004	0.002	0.004	0.006	—	—	—	—
Mm HOXA7				—	—	0.0001	ns	—	0.002	0.009	0.007	0.003	0.003	—	0.032	—	—
Rn HOXA7					—	0.0001	ns	—	0.002	0.004	0.003	0.002	0.001	—	—	—	—
Gg HOXA7						0.0001	ns	—	0.0001	0.0001	0.0001	0.0001	0.0001	—	—	—	—
Cot HOXA7							0.0001	—	—	—	—	—	—	—	—	—	0.0001
Xl HOXA7								—	0.011	0.011	—	0.021	—	—	—	—	—
Hf HOX7									—	—	—	0.0003	—	—	—	—	0.0001
Hs HOXB7										—	—	—	—	—	0.001	0.003	ns
Ph HOX-B7											ns	—	—	—	0.0002	0.003	ns
Bt HOX-B7												—	—	—	0.0003	0.002	ns
Mm HOX-B7													—	—	0.0001	0.0002	ns
Rn HOX-B7														—	0.0001	0.0004	ns
Gg HOX-B7															—	—	—
Xl HOX-B7a																ns	—
Xl HOX-B7b																	—
Dr HOX-B7a																	

NOTE.—Dashes refer to comparisons in which all values of ω in the M3 model were lower than 1; "ns" refers to those comparisons where ω > 1 values were detected in the M3 model, but the comparison M0 versus M3 showed nonsignificant differences. Numbers refer to *P* values for those M0 versus M3 comparisons in which the M3 model significantly improved over the M0 model and, in addition, ω > 1 values, indicating positive selection, were detected.

significant (summarized in table 3). Interestingly, all 10 comparison involving *Xenopus Hoxb-7a* or *Hoxb-7b* versus mammalian *Hoxb*-7 genes were significant, suggesting again that positive selection may have acted on the *Xenopus* duplicates. Most other significant values were obtained for comparisons involving *Hoxa-7* versus *Hoxb-7* genes. Out of 72 comparisons of that type, evidence for positive selection was obtained in 30 of them (42%). On the contrary, evidence for positive selection for comparison between two *Hoxa-7* or two *Hoxb-7* genes are, once the *Xenopus Hoxb-7* genes are excluded, much more infrequent. None of the 21 comparisons among *Hoxb-7* genes and only seven out of 28 for *Hoxa-7* genes were found to show evidence for positive selection (total: 14%). These results are congruent with the evidence for positive selection after duplication obtained, according to the free-ratio model, for the whole set of sequences.

We detected significant type I functional divergence (i. e., gene-specific differences in constraint; see Gu [1999] and *Materials and Methods*) after the *Hoxa-7*/*Hoxb-7* duplication ($\theta = 0.310 \pm 0.081$; LRT = 14.72; $P < 0.001$). Amino acids responsible for such divergence (with cutoff value $P > 0.6$; see *Materials and Methods*) are shown in figure 1. This figure also contains all positions with type II functional divergence (i. e., gene-specific amino acidic conservation), according to mutual information content value (see *Materials and Methods*). Consideration of the codons that show high posterior Bayesian probabilities of change in the branch connecting *Hoxa-7* and *Hoxb-7* genes suggest that an important fraction (9/23) of type II changes occurred immediately after the duplication of these paralogs and before species divergence (fig. 1). Evidence for type I, type II, or both types of functional divergence was obtained for 29 out of 135 positions in the N-terminal half of the protein (21%), whereas in the very conserved carboxy-terminal part of the protein, which includes the homeodomain, we only detected 2/75 (3%) significant positions. For the pairwise comparisons, almost all significant codons (posterior Bayesian probabilities > 0.90) were also found to be in the N-terminal region (data not shown). Our data thus suggest that changes in the N-terminal region were determinant in *Hox7* paralogs diversification, whereas the rest of the protein, and most especially the homeodomain, has conserved its functions since before these genes become duplicated.

Discussion

As indicated above, most recent hypotheses to explain *Hox* genes maintenance are based on subfunctionalization (Force et al. 1999) or additive effects (Greer et al. 2000). Our data do not contradict those hypotheses, but they show however that positive selection on coding regions, as postulated by the classical model (Ohno 1970), may have been also a significant force shaping *Hox* genes evolution. This type of selection, which leads to diversification of protein function, may have contributed to the maintenance of duplicated *Hox7* genes in multiple lineages.

Interpretation of our results depends on how *Hoxa-7* and *Hoxb-7* genes originated. This question is related to

determining the origin of the multiple *Hox* gene clusters found in vertebrates. So far, several hypotheses have been proposed to explain the evolution of those clusters. First, several authors proposed that mammalian *Hox* gene clusters originated in two successive rounds of duplication, perhaps associated with two full-genome duplications, in such a way that the current four clusters derive from two ancestral clusters, a proto AB cluster and a proto CD cluster. This model is often referred to as [(AB)(CD)] (Schughart, Kappen and Ruddle 1989; Kappen and Ruddle 1993; Zhang and Nei 1996; reviewed in Málaga-Trillo and Meyer [2001]; see also Gu, Wang, and Gu [2002] for a comprehensive analysis on whether genomic duplications actually occurred). However, alternative models, involving three rounds of duplication have been also proposed. Zhang and Nei (1996) indicated two of those scenarios: [(B(A(CD))] or [A(B(CD))]. Bailey et al. (1997) suggested a different one: [D(A(BC))]. Recently, Force, Amores, and Postlethwait (2002), although favoring the two-round hypothesis, proposed two additional alternatives with three steps: [D(C(AB))] and [C(D(AB))]. Largely, the discussion depends on the data used. In this case, sequence similarity of *Hox* genes (Zhang and Nei 1996), similarity of collagen genes, closely linked to each of the *Hox* clusters (Bailey et al. 1997), or parsimony analyses of the patterns of presence/absence of *Hox* genes in different clusters (e. g., Force, Amores, and Posthethwait 2002) yield contradictory results. It was expected that studying organisms that diverged from tetrapods before any of the cluster duplications occurred would sort out which of those alternatives is correct. However, despite considerable advances, the situation is still unclear. Thus, a chondrichthyan, the horn shark *Heterodontus francisci*, has at least two clusters, one of them quite similar to the mammalian A cluster and the other related to the mammalian D cluster (Kim et al. 2000). However, the possibility of this species having up to four clusters has not been excluded (cited in Irvine et al. [2002]). An even more distant relative, the agnathan *Petromyzon marinus*, a sea lamprey, has been recently studied by two groups who found that it contains three or, more likely, four *Hox* clusters (Force, Amores, and Postlethwait 2002; Irvine et al. 2002). Although it seems that part of this complexity arose by cluster duplications specific for agnathan fishes, it still does not fully contribute to establishing the most likely history for the four tetrapod *Hox* clusters.

We think however that, if indeed any of the six hypotheses detailed above are correct, these complications do not greatly affect the interpretation of our results. Fortunately, all them are quite similar from the point of view of the origin of A and B cluster genes. In three of those alternatives, [(AB)(CD)], [D(C(AB))], and [C(D(AB))], *Hoxa-7* and *Hoxb-7* genes would have emerged from an ancestral ''protoa-7/b-7'' gene. In those cases, we can suggest, according to our results, that positive natural selection acted on one or both of those genes after the duplication event, precisely as Ohno's classical model postulates. In the other three cases, [(B(A(CD))], [A(B(CD))], and [D(A(BC))], it is significant that *Hoxa-7* and *Hoxb-7* genes are still separated by a single step (i. e., one duplication event). Thus, no matter when the putative *Hoxc-7* and *Hoxd-7* genes arose and became lost, we still can postulate that strong positive selection may have occurred just after the genes that gave rise to the modern *Hoxa-7* and *Hoxb-7* genes originated by duplication.

In summary, we suggest that, no matter what the precise evolutionary history of *Hox7* genes turns to be, the most likely explanation for our results would be that the duplication that originated the gene lineages that gave rise to the *Hoxa-7* and *Hoxb-7* genes was followed by a period in which those two lineages diversified under positive selection. We also suggest that tetraploidization in the *Xenopus* lineage may have been followed by a similar episode in which positive selection acted on one of the *Hoxb-7* duplicates (most likely *Hoxb-7b* [fig. 2]) or even on both of them (see the positive pairwise comparisons in table 3). After those periods of positive selection–driven processes, strong purifying selection predominated, being in fact the main force in the long term. This may explain why some ω values, as the one for the *Xenopus Hoxb-7b* branch or the one that groups all tetrapod *Hoxa-7* genes, have values in the free-ratio model that are lower than 1 but still several times higher than almost all the other branches (see figure 2). It cannot be excluded that positive selection has contributed also to the substantial divergence of the *Hoxa-7* gene of the fish *Morone* (Snell, Scemama, and Stellwag 1999) that we excluded from our analyses. Additional studies whenever other actynopterygian *Hoxa-7* sequences are available may shed light on this intriguing possibility.

No matter how they originated, that *Hoxa-7* and *Hoxb-7* genes existed before the split of Actinopterygia and Sarcopterygia can be considered an established fact. On one hand, two closely related actinopterygian species (belonging to the *Morone* and *Oreochromis* genera) have been found to contain a *Hoxa-7* gene (Snell, Scemama, and Stellwag 1999; work cited in Málaga-Trillo and Meyer [2001]), while a more distant relative, *Danio rerio*, has a *Hoxb-7* gene (Amores et al. 1998). On the other hand, both tetrapods and the coelacanth *Latimeria* (Koh et al. 2003) have both *Hoxa-7* and *Hoxb-7* genes. A significant problem is then to explain the loss of *Hox7* genes in actinopterygian fish lineages at the same time that we suggest that the two paralogs diversified immediately after the gene duplication that originated them, and thus they may have been already functionally different in the ancestor of all actynopterygian. A possibility is that *Hox7* genes may have become dispensable as a consequence of the additional round of duplication that occurred in actinopterygian fishes. This additional duplication may have happened soon after the Actinopterygia/Sarcopterygia split, if it is confirmed that species that are basal in the actinopterygian tree have more than four *Hox* gene clusters, as suggested by data obtained for the bichir, *Polypterus* (Ledje, Kim, and Ruddle 2002). This hypothesis moreover implies that genes of paralog groups other than group VII must be able in some cases to compensate for lack of *Hox7* gene function. In this context, considering the phenotypes of mouse that are null mutants for *Hox7* genes is significant. Mouse *Hoxa-7* knockout mutants are normal, whereas only a small percentage of *Hoxb-7* null

mutant mice have skeletal abnormalities, and even double mutants are often normal (Chen, Greer, and Capecchi 1998). All these results contrast with the complex patterns of expression detected for both *Hoxa-7* and *Hoxb-7*, which led to predictions of multiple roles for these genes (Mahon, Westphal, and Gruss 1988; Vogels, de Graaff, and Deschamps 1990). Lack of phenotypes in single mutants does not demonstrate identical functions for both paralogs (Nowak et al. 1997). In addition, our results suggest that mammalian *Hox7* genes cannot be redundant, because they both show strong selective constraints. Thus, the subtle phenotypic effect in double mutants suggests that apparent redundancy caused by homeostatic effects due to the action of *Hox* genes that belong to paralog groups other than group VII may be occurring (see Rijli and Chambon [1997] for a discussion).

It is interesting to compare our results to those of Hughes and Hughes (1993). These authors analyzed whether *Xenopus*, mouse, or human *Hox7* genes evolved under directional selection, finding negative results. However, their analyses were based on procedures (Nei and Gojobori 1986) that allow detection of positive selection only when it involves the whole sequence of the gene. With more refined methods, we have been able to detect positive selection acting on a limited number of amino acids and in a phylogenetic context. Interestingly, a recent work presented evidence for positive selection for three duplicated *Hox* genes in zebrafish using a different methodology (Van de Peer et al. 2001). These results suggest that analyzing whether episodes of positive selection on coding regions contribute to explaining the evolution of most or even all *Hox* genes is an attractive research program.

Finally, recent data suggest that changes in coding regions of the homeotic gene *Ultrabithorax* may have been critical in the modifications of body plans found in arthropods and onychophorans (Galant and Carroll 2002; Ronshaugen, McGinnis, and McGinnis 2002). Significant changes map to the C-terminal region of the protein, outside of the homeobox. These results, together with our data, suggest that new functions (or significant modifications of preexisting ones) may be acquired by relatively simple changes affecting particular regulatory regions present in HOX proteins. In particular, it has been shown that the N-terminal region where we have detected most of the significant changes acts as a modulator of the repressive action of the HOXA-7 homeodomain (Schnabel and Abate-Shen 1996), and deletions of any of two parts of it (corresponding to positions 40 to 86 and 88 to132 in figure 1) altered HOXB-7 function on granulocyte differentiation (Yaron et al. 2001). In this context, the data showing that HOXB-7 protein interacts through its N-terminus with the cactus-related protein IκB-α (Chariot et al. 1999) are most interesting. They suggests that the modulatory action of this region may be caused by regulating HOX proteins interactions with certain partners. These partners, as classically shown for cofactors such as Exd/Pbx (reviewed in Mann and Affolter [1998]), determine the ability to bind specific sequences, and thus activate or repress Hox target genes. However, an interesting difference is that Exd/Pbx factors bind to highly conserved regions in HOX proteins, as the YPWM motif that is located just before the homeodomain (positions 139 to 142 in figure 1). Thus, in this case, they could potentially interact with both *Hoxa-7* and *Hoxb-7* gene products. Our results suggest that interactions through regions in the less conserved N-terminal half of HOX proteins may contribute to differentiating the functions of very close relatives, such as the HOXA-7/HOXB-7 pair.

Acknowledgments

I.M. is supported by Fundació *La Caixa* (01/080–00), Generalitat Valenciana (BM-011/2002) and the Spanish Ministry of Science and Technology (MCYT), as part of the *NEUROGENOMICA*/CAGEPEP project, (GEN2001-4851-C06-02).

Literature Cited

Akaike, H. 1974. New look at statistical-model identification tree. IEEE T. Automat. Contr. **AC19**:716–723.

Amores, A., A. Force, Y. L. Yan et al. (13 co-authors). 1998. Zebrafish *hox* clusters and vertebrate genome evolution. Science **282**:1711–1714.

Aparicio, S., S. Hawker, A. Cottage, Y. Mikawa, L. Zuo, B. Venkatesh, E. Chen, R. Krumlauf, and S. Brenner. 1997. Organization of the *Fugu rubripes* Hox clusters: evidence for continuing evolution of vertebrate Hox complexes. Nat. Genet. **16**:79–83.

Bailey, W. J., J. Kim, G. P. Wagner, and F. H. Ruddle. 1997. Phylogenetic reconstruction of vertebrate Hox cluster duplications. Mol. Biol. Evol. **14**:843–853.

Bisbee, C.A., M. A. Baker, A. C. Wilson, H. A. Irandokht, and M. Fischberg. 1977. Albumin phylogeny for clawed frogs (*Xenopus*). Science **195**:785–787.

Chariot, A., F. Princen, J. Gielen, M. P. Merville, G. Franzoso, K. Brown, U. Siebenlist, and V. Bours. 1999. IκB-α enhances transactivation by the HOXB7 homeodomain-containing protein. J. Biol. Chem. **274**:5318–5325.

Chen, F., J. Greer, and M. Capecchi. 1998. Analysis of *Hoxa7*/*Hoxb7* mutants suggests periodicity in the generation of the different sets of vertebrae. Mech. Dev. **77**:49–57.

Finnerty, J. R, and M. Q. Martindale. 1998. The evolution of the Hox cluster: insights from outgroups. Curr. Opin. Genet. Dev. **8**:681–687.

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

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151**:1531–1545.

Galant, R., and S. B. Carroll. 2002. Evolution of a transcriptional repression domain in an insect Hox protein. Nature **415**:910–913.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.

Gorodkin, J., H. H. Staerfeldt, O. Lund, and S. Brunak. 1999. MatrixPlot: visualizing sequence constraints. Bioinformatics **15**:769–770.

Greer, J. M., J. Puetz, K. R. Thomas, and M. R. Capecchi. 2000. Maintenance of functional equivalence during paralogous *Hox* gene evolution. Nature **403**:661–665.

Gu, X. 1999. Statistical methods for testing functional divergence after gene duplication. Mol. Biol. Evol. **16**:1664–1674.

Gu, X., Y. Wang, and J. Gu. 2002. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nat. Genet. **31**:205–209.

Gu, X., and K. Vander Velden. 2002. DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. Bioinformatics **18**:500–501.

Gu, X., and J. Zhang. 1997. A simple method for estimating the parameter of substitution rate variation among sites. Mol. Biol. Evol. **14**:1106–1113.

Hirth, F., T. Loop, B. Egger, D. F. B. Miller, T. C. Kaufman, and H. Reichert. 2001. Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of *Drosophila*. Development **128**:4781–4788.

Huelsenbeck, J. P., and K. A. Crandall. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Syst. **28**:437–466.

Hughes, M. K., and A. L. Hughes. 1993. Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. Mol. Biol. Evol. **10**:1360–1369.

Irvine, S. Q., J. L. Carr, W. J. Bailey, K. Kawasaki, N. Shimizu, C. T. Amemiya, and F. H. Ruddle. 2002. Genomic analysis of Hox clusters in the sea lamprey *Petromyzon marinus*. J. Exp. Zool. **294**:47–62.

Kappen, C., and F. H. Ruddle. 1993. Evolution of a regulatory gene family: HOM/HOX genes. Curr Opin. Genet. Dev. **3**:931–938.

Kim, C.-B., C. Amemiya, W. Bailey, K. Kawasaki, J. Mezey, W. Miller, S. Minoshima, N. Shimizu, G. Wagner, and F. Ruddle. 2000. Hox cluster genomics in the horn shark, *Heterodontus francisci*. Proc. Natl. Acad. Sci. USA **97**:1655–1660.

Koh, E. G., K. Lam, A. Christoffels, M. V. Erdmann, S. Brenner, and B. Venkatesh. 2003. Hox gene clusters in the Indonesian coelacanth, *Latimeria menadoensis*. Proc. Natl. Acad. Sci. USA **100**:1084–1088.

Krumlauf, R. 1994. *Hox* genes in vertebrate development. Cell **78**:191–201.

Kumar, S., K. Tamura, I. B. Jacobsen, and M. Nei. 2001. MEGA: molecular evolutionary genetics analysis. Version 2.1. Distributed by the authors (www.megasoftware.net).

Ledje, C., C. B. Kim, and F. H. Ruddle. 2002. Characterization of Hox genes in the bichir, *Polypterus palmas*. J. Exp. Zool. **294**:107–111.

Mahon, K. A., H. Westphal, and P. Gruss. 1988. Expression of homeobox gene Hox 1.1 during mouse embryogenesis. Development **104**(Suppl.):187–95.

Málaga-Trillo, E., and A. Meyer. 2001. Genome duplications and accelerated evolution of *Hox* genes and cluster architecture in teleost fishes. Am. Zool. **41**:676–686.

Mann, R. S., and M. Affolter. 1998. Hox proteins meet more partners. Curr. Opin. Genet. Dev. **8**:423–429.

Massingham, T., L. J. Davies, and P. Liò. 2001. Analysing gene function after duplication. Bioessays **23**:873–876.

McGinnis, W., and R. Krumlauf. 1992 Homeobox genes and axial patterning. Cell **68**:283–302.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

Nicholas, K. B., and H. B. Nicholas, Jr. 1997. Distributed by the authors (www.cris.com/~ketchup/genedoc.shtml).

Nielsen, R., and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148**:929–938.

Nowak, M. A., M. C. Boerlijst, J. Cooke, and J. Maynard Smith. 1997. Evolution of genetic redundancy. Nature **388**:167–171.

Ohno, S. 1970. Evolution by gene duplication. Springer-Verlag, Berlin.

Pollard, S. L., and P. W. H. Holland. 2000. Evidence for 14 homeobox gene clusters in human genome ancestry. Curr. Biol. **10**:1059–1062.

Prince, V. E. 2002. The Hox paradox: more complex(es) than imagined. Dev. Biol. **249**:1–15.

Rijli, F. M., and P. Chambon. 1997. Genetic interactions of *Hox* genes in limb development: learning from compound mutants. Curr. Opin. Genet. Dev. **7**:481–487.

Ronshaugen, M., M. McGinnis, and W. McGinnis. 2002. Hox protein mutation and macroevolution of the insect body plan. Nature **415**:914–917.

Ruddle, F. H., J. L. Bartels, K. L. Bentley, C. Kappen, M. T. Murtha, and J. W. Pendleton. 1994 Evolution of *Hox* genes. Annu. Rev. Genet. **28**:423–432.

Saitou, N, and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

Schnabel, C. A., and C. Abate-Shen. 1996. Repression by *HoxA7* is mediated by the homeodomain and the modulatory action of its N-terminal-arm residues. Mol. Cell. Biol. **16**:2678–2688.

Schughart, K., C. Kappen, and F. H. Ruddle. 1989. Duplication of large genomic regions during the evolution of vertebrate homeobox genes. Proc. Natl. Acad. Sci. USA **86**:7067–7071.

Snell, E. A., J. L. Scemama, and E. J. Stellwag. 1999. Genomic organization of the Hoxa4-Hoxa10 region from *Morone saxatilis*: implications for *Hox* gene evolution among vertebrates. J. Exp. Zool. (Mol. Dev. Evol.) **285**:41–49.

St-Jacques, B., and A. P. McMahon. 1996. Early mouse development: lessons from gene targeting. Curr. Opin. Genet. Dev. **6**:439–444.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **24**:4876–4882.

Van de Peer, Y., J. S. Taylor, I. Braasch, and A. Meyer. 2001. The ghost of selection past: rates of evolution and functional divergence in anciently duplicated genes. J. Mol. Evol. **53**:436–446.

Vogels, R., W. de Graaff, and J. Deschamps. 1990. Expression of the murine homeobox-containing gene *Hox-2.3* suggests multiple time-dependent and tissue-specific roles during development. Development **110**:1159–1168.

Wang, Y., and X. Gu. 2001. Functional divergence in the caspase gene family and altered functional constraints: statistical analysis and prediction. Genetics **158**:1311–1320.

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

———. 2000. Phylogenetic analysis of maximum likelihood (PAML). Version 3. University College London, England.

Yang, Z., R. Nielsen, N. Goldman, and A. M. K. Pedersen. 2000*a*. Codon substitution models for heterogeneous selection pressures at amino acid sites. Genetics **153**:1077–1089.

———. 2000*b*. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**:431–449.

Yaron, Y., J. K. McAdara, M. Lynch, E. Hughes, and J. C. Gasson. 2001. Identification of novel functional regions important for the activity of HOXB7 in mammalian cells. J. Immunol. **166**:5058–5067.

Zhang, J., and M. Nei. 1996. Evolution of Antennapedia-class homeobox genes. Genetics. **142**:295–303.

William Jeffery, Associate Editor

Accepted July 16, 2003