THE ROYAL SOCIETY

Received 7 June 2004
Accepted 25 June 2004
Published online 20 October 2004

Adaptive evolution of *HoxA-11* and *HoxA-13* at the origin of the uterus in mammals

Vincent J. Lynch[1*], Jutta J. Roth[1,2], Kazuhiko Takahashi[1†], Casey W. Dunn[1], Daisuke F. Nonaka[1¶], Geffrey F. Stopper[1] and Günter P. Wagner[1]

[1]*Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, CT 06551, USA*
[2]*Department of Genetics and General Biology, University of Salzburg, Hellbrunnerstrasse 34, 5020 Salzburg, Austria*

The evolution of morphological characters is mediated by the evolution of developmental genes. Evolutionary changes can either affect *cis*-regulatory elements, leading to differences in their temporal and spatial regulation, or affect the coding region. Although there is ample evidence for the importance of *cis*-regulatory evolution, it has only recently been shown that transcription factors do not remain functionally equivalent during evolution. These results suggest that the evolution of transcription factors may play an active role in the evolution of development. To test this idea we investigated the molecular evolution of two genes essential for the development and function of the mammalian female reproductive organs, *HoxA-11* and *HoxA-13*. We predicted that if coding-region evolution plays an active role in developmental evolution, then these genes should have experienced adaptive evolution at the origin of the mammalian female reproductive system. We report the sequences of *HoxA-11* from basal mammalian and amniote taxa and analyse *HoxA-11* and *HoxA-13* for signatures of adaptive molecular evolution. The data demonstrate that these genes were under strong positive (directional) selection in the stem lineage of therian and eutherian mammals, coincident with the evolution of the uterus and vagina. These results support the idea that adaptive evolution of transcription factors can be an integral part in the evolution of novel structures.

Keywords: *Hox* genes; adaptive evolution; natural selection; uterus; origin of mammals

1. INTRODUCTION

Hox genes are a family of highly conserved homeodomain-containing transcription factors that are homologous to genes in the *Drosophila* Antp and Ubx complexes (Shashikant *et al.* 1991; Ruddle *et al.* 1994). The primary function of *Hox* genes in bilaterians is morphogenesis and patterning the body axis (McGinnis & Krumlauf 1992). In mammals, 13 different paralogous groups of *Hox* genes have been identified in four clusters (HoxA–HoxD), with two to four members of each paralogue group. The *Hox* clusters are thought to have arisen from tandem duplications of a single ancestral gene and subsequent cluster duplication (Kappen *et al.* 1989). Divergence among *Hox* genes and their functions is thought to be a source of many of the differences in animal body plans (Holland 1992).

In vertebrates, *Hox* genes have acquired several secondary functions in addition to their role in morphogenesis and pattern formation, including the development of limbs (Haack & Gruss 1993) and urogenital organs (Hsieh-Li *et al.* 1995; Taylor *et al.* 1997; reviewed in Kobayashi & Behringer 2003). In mammals, expression of the Abd-B related *HoxA* genes *HoxA-9*, *HoxA-10*, *HoxA-11* and *HoxA-13* along the paramesonephric (Müllerian) duct is essential for the development and function of the female reproductive tract (Taylor *et al.* 1997). These genes continue to be expressed in adults (figure 1*a*) and two, *HoxA-10* and *HoxA-11*, are required for successful implantation of mammalian blastocysts (Hsieh-Li *et al.* 1995; Satokata *et al.* 1995). In addition to expression in developing and adult vagina, *HoxA-13* is essential for the formation of the umbilical arteries (Stadler *et al.* 2001) and *HoxA-13* homozygous null mice die *in utero* from stenosis of the umbilical arteries (Warot *et al.* 1997).

The female reproductive system of monotremes (egg-laying mammals) and other amniotes is radically different from marsupial and placental mammals (therians; figure 1*b*). In therians, for example, eggshell formation no longer occurs, the uterus nourishes the developing embryo through a vascular placenta and a novel structure, the vagina, has evolved (Romer & Parsons 1997). These structural and functional changes occurred relatively recently compared to other major morphological innovations, such as the fin–limb transition, which might have occurred too long ago to test for adaptive molecular evolution using current methods. Remarkably, positive selection has been detected in extremely ancient events, including in *xanthine dehydrogenase* genes in the common ancestor of plants and animals (Rodriguez-Trelles *et al.* 2003), *phytochrome A* genes in early angiosperms (Mathews *et al.* 2003) and in the *Troponin C* gene family at the base of the vertebrates (Bielawski & Yang 2003), demonstrating that directional selection can be detected in ancient gene duplication events given optimal sequence divergence.

Hox genes are composed of two exons: exon 1 encodes an amino-terminal domain, which may regulate functional specificity through cofactor associations and exon 2 encodes the highly conserved homeodomain. The con-

* Author for correspondence (vincent.j.lynch@yale.edu).
† Present address: National Institute for Basic Biology, Okazaki National Research Institutes, 38 Nishigonaka, Myodaijicho, Okazaki, Aichi 444-8585, Japan.
¶ Present address: Department of Medicine, UCSF Prostate Cancer Program, 2340 Sutter Street, Box 0128, San Francisco, CA 94115, USA.

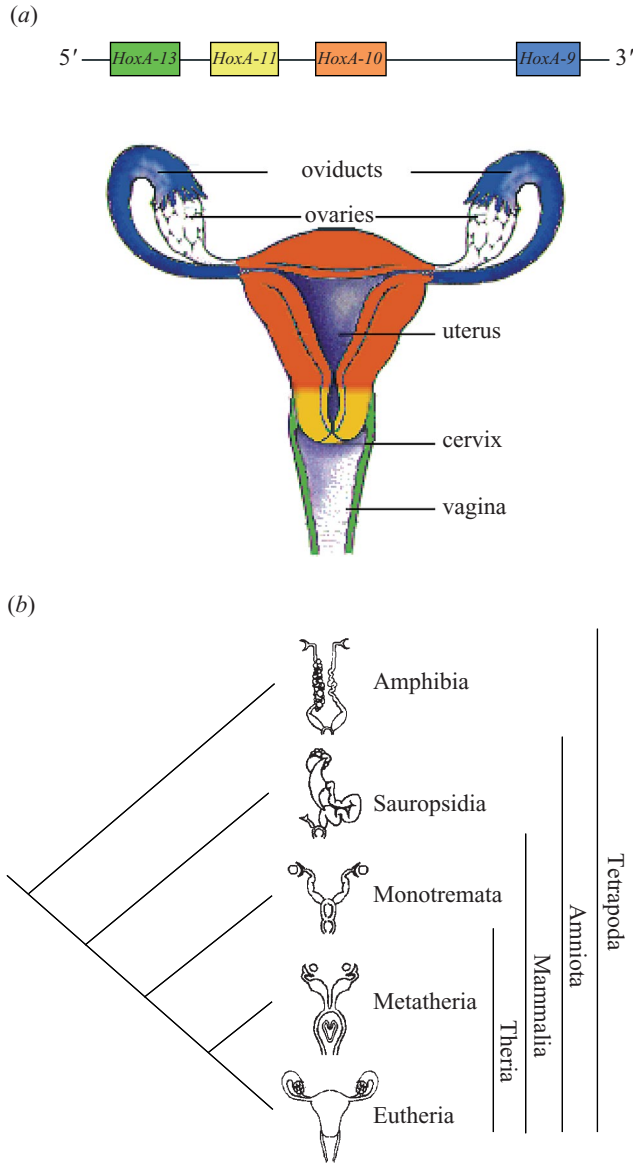

Figure 1. *Hox* gene expression pattern and the evolution of the female reproductive tract. (*a*) *HoxA-13* to *HoxA-9* are located at the 5′ end of the *HoxA* cluster and are expressed in the same regions in the adult as in the embryo: *HoxA-13* (green), *HoxA-11* (yellow), *HoxA-10* (orange) and *HoxA-9* (blue). (*b*) Tetrapod phylogeny showing representative female reproductive systems from each group (amphibian ovaries shown only on the left).

served function of the homeodomain in DNA-binding makes it an unlikely target of adaptive evolution, however, the amino terminal domain's role in functional specificity make it a probable target of adaptive evolution associated with recruitment into novel developmental pathways and cell biological functions.

Several methods have been developed to test for adaptive molecular evolution. These methods determine the strength and direction of selection by estimating the non-synonymous-to-synonymous substitution rate ratio ($d_N/d_S = \omega$), with $\omega = 1$, less than 1 and greater than 1 indicating neutral evolution, purifying selection and directional selection, respectively.

Using the d_N/d_S ratio as an indicator of positive selection, Messier & Stewart (1997) developed a method for detecting episodic adaptive evolution that incorporated ancestral sequence reconstruction to identify ancestral lineages with elevated d_N/d_S. This method has been modified by Yang (1998) into a maximum-likelihood framework that averages over all possible ancestral sequences at each interior node in the tree and accounts for the transition–transversion rate ratio (κ), the structure of the genetic code and different base frequencies at codon positions (Goldman & Yang 1994). These codon-based likelihood models allow d_N/d_S to vary among lineages and can be used to construct likelihood-ratio tests that examine whether lineages have experienced episodes of positive selection. The codon-substitution model has been further refined to detect molecular adaptation at specific amino acid sites along specific lineages (Yang & Nielsen 2002).

The origin of the mammalian female reproductive tract provides an excellent opportunity to investigate the molecular evolution of developmental genes recently involved in the evolution of novel structures, including the uterus, vagina and placenta. We examined *HoxA-11* and *HoxA-13* genes from mammals and other species using codon-based maximum-likelihood models of evolution to test the idea that adaptive evolution of transcription factor proteins is involved in the origin of novel characters.

2. MATERIAL AND METHODS

(a) *Polymerase chain reaction and sequencing*

Most *Hox* genes are composed of two exons, separated by an intron of variable length. Exon 1 encodes the amino-terminal domain and exon 2 encodes the highly conserved homeodomain. We amplified exon 1 of *Hox-A-11* by polymerase chain reaction (PCR) from the genomic DNA of cow (*Bos taurus*), armadillo (*Dasypus* sp.), hyrax (*Procavia capensis*), opossum (*Didelphis virginiana*), platypus (*Ornithorynchus anatinus*), echidna (*Tachyglossus aculeatus*) and tortoise (*Geochelonenigra porteri*) using amniote-specific degenerate primers (*HoxA-11*-forward: 5′-ATGGATTTTGATGAGCGTGKTCCYTGYT-3′; *HoxA-11*-reverse: 5′-GCMCTCGCCACGTGATC-3′). PCR products were cloned into the pGEM-T vector (Promega). Cloned PCR products were sequenced in both directions by dideoxy chain termination using Big Dye chemistry and an automated sequencer. Multiple colonies (at least four) were sequenced for each species.

The human (*Homo sapiens*, NM 005523), rat (*Rattus norvegicus*, NW 046692), mouse (*Mus musculus*, NM 010450), chicken (*Gallus gallus*, AF327372) and *Xenopus tropicalis* (AF287140) *HoxA-11* sequences were from GenBank. Chris Amemiya provided the coelacanth (*Latimeria menadoensis*) *HoxA-11* sequence. The *Xenopus laevis* (AJ31473), chicken (AY030050), opossum (*Monodelphis domestica*, AF083097), human (U82827) and mouse (*Mus musculus*, U59322) *HoxA-13* sequences were from GenBank. To have taxon sampling comparable to *HoxA-11* we attempted to sequence the monotreme *HoxA-13* gene but were unsuccessful, a difficulty also noted by other researchers (Mortlock *et al.* 2000).

(b) *Sequence analysis*

HoxA-11 and *HoxA-13* amino acid sequences were initially aligned in ClustalW (Higgins *et al.* 1994) and manually adjusted in Se-Al (Rambaut 1996). Alignment gaps, regions of ambiguous alignment and exon 2 were excluded from the analysis, resulting in a *HoxA-11* dataset 185 codons long and a *HoxA-13* dataset 114 codons long. As an initial exploration of the data we estimated the

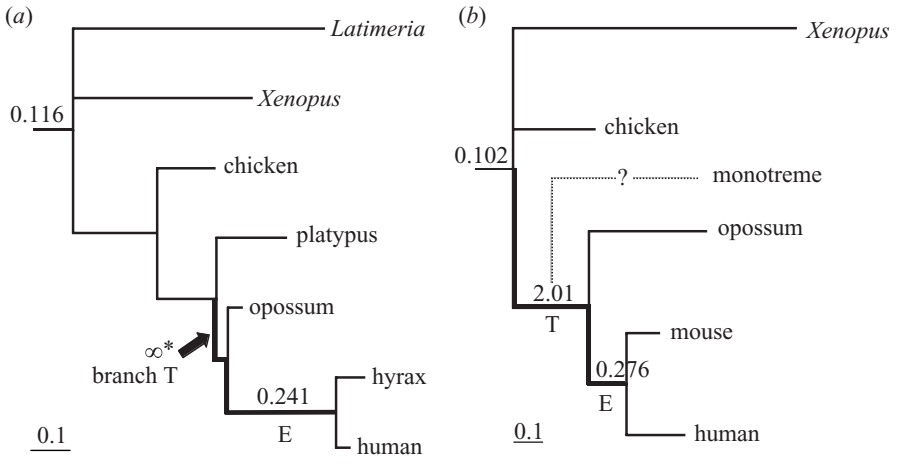

Figure 2. Phylogeny of the species in this study. The maximum-likelihood estimate of the non-synonymous-to-synonymous substitution rate ratio is given above the branch ($\omega = d_N/d_S$) for *HoxA-11* (*a*) and *HoxA-13* (*b*). The three-ratios model was used to estimate d_N and d_S for both genes. The stem lineages of therians (T) and eutherians (E) are labelled. The estimate of the background ratio (ω_0) is shown above the root of each tree. Branches inferred to be under positive selection are shown in bold and are drawn in proportion to the estimate of their length, defined as the number of subsitutions per codon. *An estimate of ω for this branch could not be made because $d_S = 0$; however, 5.2 non-synonymous changes occurred along this branch.

average d_N and d_S and the d_N/d_S ratio using the methods of Nei & Gojobori (1986) and Li (1993).

We used codon-based maximum-likelihood models of coding sequence evolution (Goldman & Yang 1994) implemented in CODEML in the PAML package of programs (v. 3.14; Yang 1997) to test for lineages and amino acid sites under positive selection. We used the generally accepted sarcopterygian phylogeny for the input tree; because *HoxA-11* and *HoxA-13* are single copy genes in tetrapods the true gene tree is the species tree. In addition, the branching order of the major tetrapod groups is well supported by numerous molecular and morphological studies (Killian *et al.* 2001; Phillips & Penny 2003; Woodburne *et al.* 2003). Initial branch lengths for the input trees were determined under the one-ratio model (M0) for *HoxA-11* and *HoxA-13*; these branch lengths were used as starting values for more complex models. To conserve computation time and examine the effects of taxon sampling on our results we ran multiple analyses on the *HoxA-11* dataset including additional and alternate taxa in each analysis. This included replacing human and hyrax with armadillo and mouse, replacing chicken with tortoise, and including armadillo and cow in the eutherian clade.

We used three models to test for positive selection in the stem lineage of therians and the stem lineage of eutherians (branches T and E in figure 2). The simplest model (the one-ratio model) estimates the same d_N/d_S ratio for all branches in the tree and is the null model that more complex models are tested against. To test for positive selection we used two- and three-ratios models that estimate an independent d_N/d_S ratio for the lineages of interest (foreground lineages) and all other branches (background lineages). These three models are nested and compared using the likelihood-ratio test statistic to examine if lineages of interest have d_N/d_S ratios that are significantly different from other lineages. Twice log-likelihood difference between models, $2\Delta\ell = 2(\ell_1 - \ell_0)$, is compared with a χ^2 distribution with the degrees of freedom equal to the number of parameter differences in the models. If a lineage of interest has a $d_N/d_S > 1$ and the likelihood ratio test is significant, then the neutral model of evolution is violated and positive selection is indicated.

A limitation of lineage-specific models is that they do not allow for variable d_N/d_S among sites and can only detect positive selection if the average d_N is greater than the average d_S. Thus, if adaptive evolution affects only a few amino acids while most sites are conserved, such as in the acquisition of a novel/secondary function that preserves the ancestral function, lineage-specific models may lack the power to detect selection (Yang & Nielsen 2002). Recently, codon-substitution models have been developed that can detect positively selected sites along specific lineages (Yang & Nielsen 2002). These branch-site models account for rate variation among sites and lineages and are powerful tools for detecting episodic adaptive evolution. Branch-site model A assumes the background branches (those not under directional selection) have $\omega_0 = 0$ and $\omega_1 = 1$ and is an extension of the site-specific 'neutral' model M1, which assumes two site classes with $\omega_0 = 0$ and $\omega_1 = 1$ in all lineages. In model B, ω_0 and ω_1 are estimated from the data as free parameters. Model B is an extension of the site-specific model M3 (discrete) with two site classes. In both models A and B the foreground lineage (the lineage suspected of being under directional selection), specified *a priori*, has an additional ω parameter (ω_2) that may be greater than one.

Similar to the lineage-specific models, a likelihood ratio test is used to determine if the alternate models (A and B) are significantly better than the null models (M1 and M3 with two site classes). After likelihood estimates of parameters are obtained, an empirical Bayes approach is used to infer the probability that an amino acid falls into a particular site class. Sites with a posterior probability of greater than 0.5 are reported here. We assessed the effect of positively selected sites on likelihood estimates of ω by selectively removing positive sites with posterior probabilities under a certain threshold value (see tables 1 and 2) and reanalysing the data with both branch-specific and branch-site models.

To ensure convergence of the maximum-likelihood estimates generated under the branch-specific and branch-site models we altered the starting values of the d_N/d_S ratio (initial $\omega = 0.1$, 0.4, 1.0 and 2), the transition : transversion ratio (initial $\kappa = 0$, 0.5, 2 and 4) and the equilibrium codon frequencies ($f = f3 \times 4$ and $f61$) in the substitution model and reanalysed the data. Additionally, we also examined the effects of ancestral sequence reconstruction on our results by comparing reconstructions generated by maximum likelihood and maximum parsimony using BASEML and PAMP programs in the PAML package (v. 3.14; Yang 1997).

Table 1. Maximum-likelihood parameter estimates for *HoxA-11*.
(p is the number of free parameters for the ω ratios, sig. is the χ^2-significance value of the model when compared with the one-ratio or 'neutral' models (see §2b). Parameters indicating positive selection are presented in bold type; those in parentheses are not free parameters but are presented for clarity. Positively selected sites are identified using the human *HoxA-11* sequence as the reference followed by the posterior probability (PP) that each site falls into the site class $\omega > 1$. E refers to the stem lineage of eutherians.)

model	p	λ	sig.	estimates of parameters	positive sites
M0: one ratio	1	−2135.94		$\omega_0 = 0.133$	n.a.
branch-specific					
two ratios	2	−2134.19	$p = 0.06$	$\omega_0 = 0.127$, $\omega_T = \infty$ **(5.9/0)**	n.a.
three ratios	3	−2132.56	$p = 0.03$	$\omega_0 = 0.116$, $\omega_T = \infty$ **(5.2/0)**, $\omega_E = 0.241$	n.a.
'neutral'					
M1: neutral	1	−2132.28		$p_0 = 0.603$, $\omega_0 = 0$, ($p_1 = 0.395$), $\omega_1 = 1$	not allowed
M3: discrete ($k = 2$)	3	−2090.72		$p_0 = 0.748$, $\omega_0 = 0.032$, ($p_1 = 0.252$), $\omega_1 = 0.508$	none
branch-site					
model A (E)	3	−2122.20	$p \ll 0.01$	$p_0 = 0.585$, $p_1 = 0.324$, ($p_2 + p_3 = 0.091$), $\omega_2 =$ **2.62**	58,228 ($0.79 \leqslant PP \leqslant 0.85$); 70,80 ($0.90 \leqslant PP \leqslant 0.95$) 70, 82,170 ($0.95 \leqslant PP \leqslant 0.99$)
model B (E)	5	−2087.71	$p = 0.05$	$p_0 = 0.716$, $\omega_0 = 0.028$, $p_1 = 0.233$, $\omega_1 = 0.511$, ($p_2 + p_3 = 0.051$), $\omega_2 =$ **2.67**	70,80,227,228 ($0.51 \leqslant PP \leqslant 0.60$); 58 (PP = 0.77); 82 (PP = 0.97)

Table 2. Maximum-likelihood parameter estimates for *HoxA-13*.
(p is the number of free parameters for the ω ratios, sig. is the χ^2-significance value of the model when compared with the one-ratio or 'neutral' models (see § 2b; n.s., not significant). Parameters indicating positive selection are presented in bold type, those in parentheses are not free parameters but are presented for clarity. Positively selected sites are identified using the human *HoxA-13* sequence as the reference followed by the posterior probability (PP) that each site falls into the site class $\omega > 1$. T refers to the stem lineage of therians. E refers to the stem lineage of eutherians.)

model	p	λ	sig.	estimates of parameters	positive sites
M0: one ratio	1	−1210.76		$\omega_0 = 0.153$	n.a.
branch-specific					
two ratios	2	−1206.08	$p < 0.01$	$\omega_0 = 0.114$, $\omega_T =$ **2.1**	n.a.
three ratios	3	−1205.38	$p < 0.01$	$\omega_0 = 0.102$, $\omega_T =$ **2.01**, $\omega_E = 0.276$	n.a.
'neutral'					
M1: neutral	1	−1219.27		$p_0 = 0.538$, $\omega_0 = 0$, ($p_1 = 0.462$), $\omega_1 = 1$	not allowed
M3: discrete ($k = 2$)	3	−1200.27		$p_0 = 0.563$, $\omega_0 = 0.023$, ($p_1 = 0.436$), $\omega_1 = 0.359$	none
branch-site					
model A (T)	3	−1215.43	$p = 0.02$	$p_0 = 0.523$, $p_1 = 0.362$, ($p_2 + p_3 = 0.115$), $\omega_2 =$ **5.7**	36 (PP = 0.67); 1,16,19,60,105 ($0.75 \leqslant PP \leqslant 0.85$); 33,48 (PP = 0.94)
model B (T)	5	−1194.5	$p < 0.01$	$p_0 = 0.576$, $\omega_0 = 0.031$ $p_1 = 0.273$, $\omega_1 = 0.327$, ($p_2 + p_3 = 0.151$), $\omega_2 =$ **14.59**	36,101,114 ($0.60 \leqslant PP \leqslant 0.70$); 68,106 ($0.71 \leqslant PP \leqslant 0.80$); 74,185,210,222 ($0.81 \leqslant PP \leqslant 0.90$); 23,71,100, 108,117,160 ($0.91 \leqslant PP \leqslant 0.99$)
model A (E)	3	−1217.36	n.s.	$p_0 = 0.537$, $p_1 = 0.433$, ($p_2 + p_3 = 0.03$), $\omega_2 =$ **30.22**	51,64 ($0.60 \leqslant PP \leqslant 0.65$); 84 (PP = 0.96)
model B (E)	5	−1197.35	$p = 0.05$	$p_0 = 0.463$, $\omega_0 = 0.005$, $p_1 = 0.488$, $\omega_1 = 0.283$, ($p_2 + p_3 = 0.049$), $\omega_2 =$ **56.69**	166 (PP = 0.50); 108,150 (PP = 0.82); 13,84 (PP = 0.98)

3. RESULTS

(a) *Preliminary data analysis*

We analysed the evolution of *HoxA-11*, which is expressed in the lower uterus, uterine cervix and endometrial cells (Satokata *et al.* 1995; Taylor *et al.* 1997; figure 1*a*), using three estimators of d_N, d_S and the d_N/d_S ratio: the methods of Nei & Gojobori (1986: NG86) and Li (1993: Li93) and codon-based maximum-likelihood models of coding sequence evolution. NG86 and Li93 gave very similar estimates of d_N and d_S and here we report only those estimates for NG86. The average d_N and d_S for *HoxA-11* were 0.106 and 1.49, respectively, whereas the d_N/d_S ratio ranged from 0.037 to 0.226. The average d_N and d_S for *HoxA-13* were 0.155 and 1.43, respectively, whereas the d_N/d_S ratio ranged from 0.031 to 0.188. These results indicate that *HoxA-11* and *HoxA-13* are generally under strong purifying selection. Finally, ancestral sequence reconstructions were robust to reconstruction method and likelihood and parsimony generated the same sequences. The sequences generated in this study have been deposited in GenBank under the accession numbers AY677108–AY677113.

(b) *Adaptive evolution of* HoxA-11 *in the stem lineage of placental mammals*

Pairwise estimators of the d_N and d_S, such as NG86 and Li93, are best suited for detecting strong episodes of positive selection between species pairs, but are poor at identifying episodic adaptive evolution and do not account for phylogenetic structure. Codon-based maximum-likelihood models of evolution can account for phylogenetic structure and are valuable tools for hypothesis testing in molecular evolution. The one-ratio model is the simplest, estimating the same d_N/d_S ratio (ω) for all branches in the phylogeny (Goldman & Yang 1994; Yang 1997). The estimate of ω under this model, 0.133, is an average over all codons and branches and highlights the dominant role of purifying selection on *HoxA-11*. To test for directional selection we used a two-ratios model that estimated the ω of the stem lineage of therians (ω_T) separately from all other branches (ω_0) and three-ratios model that estimated ω_T and the ω of the stem lineage of eutherians (ω_E) separately from each other and all other branches (ω_0). The two-ratios and three-ratios models fit the data marginally better and significantly better than the one-ratio model (table 1), respectively, suggesting an episode of directional selection in the stem lineage of therians followed by purifying selection or weak directional selection in the stem lineage of eutherians (table 1 and figure 2). Results from different *HoxA-11* datasets were similar to above and were robust to taxon sampling and starting parameters. Here, we report on the simplest dataset (shown in figure 2) for clarity.

We used branch-site models (Yang & Nielsen 2002) to determine which amino acid sites were under positive selection in the stem lineages of therians and eutherians. We proposed that specific amino acid changes may have been important in the adaptation of *HoxA-11* (and *HoxA-13*) to new cofactor associations in derived cell types and structures, such as in endometrial cells, the lower uterus and cervix. Cofactor associations are important for the function of Hox proteins (Mann & Morata 2000) and interactions with cofactors may be critical for the binding specificity of Hox proteins.

Parameter estimates under branch-site models A and B with the stem lineage of therians as the foreground (branch T in figure 2) suggested that all codon sites were positively selected ($\omega = 999$) with a posterior probability of 1, these spurious results may be related to short branch length. Interestingly, Suzuki & Nei (2004) have recently reported similar results when using a star phylogeny, suggesting the failure to estimate reliable rate parameters may be due to short internal branch lengths or poorly resolved phylogenies. In the stem lineage of eutherians (branch E in figure 2), parameter estimates under model A suggest that 58% of sites were highly conserved ($\omega_0 = 0$), 32% were nearly neutral ($\omega_1 = 1$) and 10% (seven sites) were under directional selection ($\omega_2 = 2.62$) (electronic Appendix A, figure 3). Parameter estimates under model B are similar to model A and suggest that 72% of sites were highly constrained ($\omega_0 = 0.03$), 23% were weakly constrained ($\omega_1 = 0.51$) and 5% (six sites) were under directional selection along this branch ($\omega_2 = 2.67$) (electronic Appendix A, figure 3). Branch-site models A and B fit the data much better than the neutral models M1 and M3 (table 1). These results are consistent with results of the lineage-specific models indicating that *HoxA-11* was under strong directional selection during the evolution of early placental mammals.

Recently, branch-site models have been found to have a high rate of type-I error under some circumstances (Zhang 2004). We examined the sensitivity of our results to the presence of positively selected sites by sequentially removing sites with the highest posterior probabilities and reanalysing the data. We found that removing sites in the class $\omega > 1$ with a posterior probability greater than 0.90, identified from model A, reduced the ω of branch E to the background ratio and reduced models A and B to the neutral models, indicating that these sites are the main reason for the increased substitution rates in branch E. Estimates for branch T and the background ratio were unaffected, showing that positive selection acting on sites with a posterior probability greater than 0.9 in branch E is the main reason for the increased substitutions rate.

(c) *Adaptive evolution of* HoxA-13 *in the stem lineage of placental mammals*

We explored whether directional selection was also acting on *HoxA-13* during the evolution of early mammals. *HoxA-13* is expressed in the developing and adult vagina and in the umbilical arteries, novel structures that evolved in therians that have no homologues in other groups (Starck 1975). Unfortunately, we were unsuccessful in amplifying the monotreme *HoxA-13* gene, a difficulty also noted by other researchers (Mortlock *et al.* 2000). Thus, there is a long branch that corresponds to the divergence of sauropsids and therian mammals. The branch-specific one-ratio (M0) model estimates ω to be 0.153. This low d_N/d_S ratio indicates that *HoxA-13* is generally under relatively strong purifying selection, similar to *HoxA-11* and other *Hox* genes (Fares *et al.* 2003). The two- and three-ratio models, which allow ω values of the therian and eutherian stem lineages (branches T and E in figure 2) to vary from the background ratio and each other, identified an episode of positive selection in the stem lineage of therians ($\omega_T = 2.10$) and an elevated d_N/d_S in the stem lineage of eutherians ($\omega_E = 0.276$) relative to the background

ratio ($\omega_0 = 0.102$). Both the two- and three-ratio models are significantly better than the one-ratio model (table 2).

Branch-site models A and B with the stem lineage of therians as the foreground identified 8 and 14 sites under directional selection, respectively, and fit the data much better than the neutral models (table 2 and electronic Appendix A, figure 3). Model A with the stem lineage of eutherians as the foreground identified three sites under directional selection ($\omega_2 = 30.22$) (electronic Appendix A, figure 3). This model, however, is no better than the neutral model M1 (table 2). Model B with the stem lineage of eutherians as the foreground lineage identified five sites under directional selection ($\omega_2 = 56.69$; electronic Appendix A, figure 3), this model is better than the neutral model at the 5% significance level (table 2). Hence we conclude there is evidence for strong positive selection on *HoxA-13* in the stem lineage of therian mammals, but there is less evidence for directional selection on *HoxA-13* in the stem lineage of eutherians than for *HoxA-11*. We note that the elevated substitution rate observed in the lineage-specific model and the positively selected amino acids identified in the branch-site models could have occurred at any time during a *ca*. 200 Myr interval between the divergence of sauropsids (birds and reptiles) and therian mammals (Kumar & Hedges 1998).

We also examined the sensitivity of the *HoxA-13* results to the presence of positively selected sites by following the same procedure of site removal as for *HoxA-11* (described in §3b). We found that removing sites in the class $\omega > 1$ with a posterior probability greater than 0.75 reduced the value of ω of branch T to 0.889 and branch E to 0.335. Although these estimates of ω are greater than the background ratio (0.103), neither the two-ratios nor the three-ratios models are statistically distinguishable from neutral models. Similar results are obtained with branch-sites models A and B, which still identify a class of sites with $\omega > 1$ (consisting of sites from the complete dataset that were not removed) but are no better than neutral models. Interestingly, these results indicate that sites with a posterior probability less than 0.75 of being in the site class $\omega > 1$ still contribute to the increased substitution rates in branch-specific models.

4. DISCUSSION

The evolution of mammals is associated with radical changes in their reproductive biology, particularly the structure and function of the female reproductive organs. These changes include the evolution of the uterus, cervix, vagina, placenta and specialized cell types associated with each of those structures. The results presented above show that the Abd-B related genes of the HoxA cluster, *HoxA-11* and *HoxA-13,* were under strong directional selection in the stem lineages of therian and eutherian mammals. The known functions for these genes are body axis development (Kessel & Gruss 1990), limb development (Haack & Gruss 1993), blood cell differentiation (van Oostveen *et al*. 1999), female reproductive system development (Taylor *et al.* 1997) and formation of umbilical arteries (Warot *et al.* 1997; Stadler *et al.* 2001). Out of these, only the function in the mammalian female reproductive organs and umbilicus originated coincident with the inferred selective episode reported here. Thus, the adaptive changes in these Hox proteins were most probably caused by their recruitment into novel developmental and cell biological functions associated with the evolution of the placenta, the uterus, endometrial cells and the vagina in mammals. Although positive selection has been identified in many genes (reviewed in Yang & Bielawski 2000), only recently has positive selection been identified in transcription factor genes (Martinez-Castilla & Alvarez-Buylla 2003; Fares *et al.* 2003). To our knowledge, this is the first case in which a specific evolutionary change in development has been demonstrated to be coincident with adaptive molecular evolution of development control genes.

In recent years, there has been an accumulation of experimental evidence that transcription factors do not remain functionally equivalent during evolution (Ranganayakulu *et al.* 1998; Grenier & Carroll 2000; Galant & Carroll 2002; Ronshaugen *et al.* 2002). Although these experimental studies clearly demonstrate that the function of transcription factors has changed during evolution, it is not clear if these differences are the result of adaptive molecular evolution. There are, by contrast, several studies that provide evidence of the action of diversifying selection acting on transcription factor genes (Barrier *et al.* 2001), but it remains unclear whether these adaptive changes are related to novel developmental functions or other physiological adaptations unrelated to their role in development. We think that the origin of mammalian female reproductive organs and recent advances in methods for detecting molecular adaptation will provide an excellent opportunity to investigate adaptive changes associated with the recruitment of transcription factor genes into derived developmental functions.

The authors thank C. Amemiya and T. Powers for providing the unpublished coelacanth *HoxA-11* sequence, W. J. Murphy for providing hyrax genomic DNA, A. Caccone and J. R. Powell for providing the giant Galapagos tortoise DNA, as well as R. Jirtle and K. Killian for providing opossum, platypus and echidna genomic DNA. The authors also thank Z. Yang for answering questions about PAML. This work was supported by a grant from the National Science Foundation to G.P.W. (INB-0321470), a fellowship from the Doctoral Scholarship Programme of the Austrian Academy of Sciences to J.J.R., and a National Science Foundation Graduate Research Fellowship to C.W.D.

REFERENCES

Barrier, M., Robichaux, R. H. & Purugganan, M. D. 2001 Accelerated regulatory gene evolution in an adaptive radiation. *Proc. Natl Acad. Sci. USA* **98**, 10208–10213.

Bielawski, J. P. & Yang, Z. 2003 Maximum likelihood methods for detecting adaptive evolution after gene duplication. *J. Struct. Func. Genomics* **3**, 201–212.

Fares, M. A., Bezemer, D,. Moya, A. & Marin, I. 2003 Selection on coding regions determined *Hox7* genes evolution. *Mol. Biol. Evol.* **30**, 2104-2112.

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

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

Grenier, J. K. & Carroll, S. B. 2000 Functional evolution of the Ultrabithorax protein. *Proc. Natl Acad. Sci. USA* **97**, 704–709.

Haack, H. & Gruss, P. 1993 The establishment of murine Hox-1 expression domains during patterning of the limb. *Devl Biol.* **157**, 410–422.

Higgins, D., Thompson, J., Gibson, T., Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994 Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Holland, P. 1992 Homeobox genes in vertebrate evolution. *BioEssays* **14**, 267–273.

Hsieh-Li, H. M., Witte, D. P., Weinstein, M., Branford, W., Li, H., Small, K. & Potter, S. S. 1995 *Hoxa-11* structure, extensive antisense transcription, and function in male and female fertility. *Development* **121**, 1373–1385.

Kappen, C., Schughart, K. & Ruddle, F. H. 1989 Two steps in the evolution of Antennapedia-class vertebrate homeobox genes. *Proc. Natl Acad. USA* **86**, 5459–5463.

Kessel, M. & Gruss, P. 1990 Murine development control genes. *Science* **249**, 374–379.

Killian, J. K., Buckley, T. R., Stewart, N., Munday, B. L. & Jirtle, R. L. 2001 Marsupials and eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution. *Mamm. Genome* **12**, 513–517.

Kobayashi, A. & Behringer, R. R. 2003 Developmental genetics of the female reproductive tract in mammals. *Nat. Rev. Genet.* **4**, 969–980.

Kumar, S. & Hedges, S. B. 1998 A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920.

Li, W.-H. 1993 Unbiased estimation of the rates of synonymous and non-synonymous substitution. *J. Mol. Evol.* **36**, 96–99.

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

Mann, R. S. & Morata, G. 2000 The developmental and molecular biology of genes that subdivide the body of *Drosophila*. *A. Rev. Cell Dev. Biol.* **16**, 3861–3871.

Martinez-Castilla, L. P. & Alvarez-Buylla, E. R. 2003 Adaptive evolution in the *Aribidopsis* MADS-box gene family inferred from its complete resolved phylogeny. *Proc. Natl Acad. Sci. USA* **100**, 13407–13412.

Mathews, S., Burleigh, J. G. & Donoghue, M. J. 2003 Adaptive evolution of the photosensory domain of phytochrome *A* in early angiosperms. *Mol. Biol. Evol.* **20**, 1087–1097.

Messier, W. & Stewart, C.-B. 1997 Episodic adaptive evolution of primate lysozymes. *Nature* **385**, 151–154.

Mortlock, D. P., Sateesh, P. & Innis, J. W. 2000 Evolution of N-terminal sequences of the vertebrate HOXA13 protein. *Mamm. Genome* **11**, 151–158.

Nei, M. & Gojobori, T. 1986 Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitution. *Mol. Biol. Evol.* **3**, 418–426.

Phillips, M. J. & Penny, D. 2003 The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylogenet. Evol.* **23**, 171–185.

Rambaut, A. 1996 Se-Al: sequence alignment editor. Available at http://evolve.zoo.ox.ac.uk/

Ranganayakulu, G., Elliott, D. A., Harvey, R. P. & Olson, E. N. 1998 Divergent roles for NK-2 class homeobox genes in cardionenesis in flies and mice. *Development* **125**, 3037–3048.

Rodriguez-Trelles, F., Tarrio, R. & Ayala, F. R. 2003 Convergent neofunctionalization by positive Darwinian selection after ancient recurrent duplications of the xanthine dehydrogenase gene. *Proc. Natl Acad. Sci. USA* **100**, 13413–13417.

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

Romer, A. S. & Parsons, T. S. 1997 *The vertebrate body*. Fort Worth, TX: International Thomson Publishing.

Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murtha, M. T. & Pendleton, J. W. 1994 Evolution of Hox genes *A. Rev. Genet.* **28**, 423–442.

Stadler, H. S., Higgins, K. M. & Capecchi, M. R. 2001 Loss of Eph-receptor expression correlates with loss of cell adhesion and chondrogenic capacity in Hoxa13 mutant limbs. *Development* **128**, 4177–4188.

Starck, V. D. 1975 *Embryologie: Ein Lehrbush auf allgemein biologischer Grundlage*. Stuttgart: Georg Thieme Verlag.

Satokata, I., Benson, G. & Maas, R. 1995 Sexually dimorphic sterility phenotypes in Hoxa-13 deficient mice. *Nature* **347**, 460–463.

Shashikant, C. S., Utset, M. F., Violette, S. M., Wise, T. L., Einat, P., Einat, M., Pendleton, J. W., Schughart, K. & Ruddle, F. H. 1991 Homeobox genes in mouse development. *Crit. Rev. Eukar. Gene Express.* **1**, 207–245.

Suzuki, Y. & Nei, M. 2004 False-positive selection identified by ML-based methods: examples from the Sig1gene of the diaton *Thalassiosira weissflogii* and the tax gene of a human T-cell lymphotropic virus. *Mol. Biol. Evol.* **21**, 914–921.

Taylor, H. S., Vanden, H. G. & Igarashi, P. A. 1997 A conserved Hox axis in the mouse and human reproductive system: late establishment and persistent expression of the Hoxa cluster genes. *Biol. Reprod.* **57**, 1338–1345.

van Oostveen, J., Biji, I., Raaphorst, F., Walboomers, J. & Meijer, C. 1999 The role of homeobox genes in normal hematopoiesis and hematological malignancies. *Leukemia* **13**, 1675–1690.

Warot, X., Fromental-Ramin, C., Fraulob, V., Chambon, P. & Dollé, P. 1997 Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts. *Development* **124**, 4781–4791.

Woodburne, M. O., Rich, T. S. & Springer, M. S. 2003 The evolution of tribospheny and the antiquity of mammalian clades. *Mol. Phylogenet. Evol.* **28**, 360–385.

Yang, Z. 1997 PAML, a program package for phylogenetic analysis by maximum likelihood. *Comp. App. Bio. Sci.* **13**, 555–556.

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

Yang, Z. & Bielawski, J. P. 2000 Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–502.

Yang, Z. & Nielsen, R. 2002 Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**, 908–917.

Zhang, J. 2004 Frequent detection of positive selection by the likelihood method with branch-site models. *Mol. Biol. Evol.* **21**, 1332-1339.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.

Visit www.journals.royalsoc.ac.uk and navigate through to this article in *Proceedings*: Biological Sciences to see the accompanying electronic appendices.