Appendix 1

Selection of model of amino acid substitution.

In the SOWHL tests performed on amino acid alignments either the WAG (Whelan and Goldman 2001) or the JTT (Jones et al. 1992) replacement matrix was used. The WAG matrix was selected over others on the basis that it was estimated from a large database of different globular protein families, as opposed being tailored to a specific family (as in the mtREV24 and mtmamm matrices), and it avoids some of the pitfalls of matrices derived from counting-based methods (such as the Dayhoff and JTT matrices, see Whelan and Goldman (Whelan and Goldman 2001) and references therein). The JTT matrix was used to test the assumptions of the SOWHL test (see main text). We used the second order Akaike Information Criterion (AIC_c) to select other aspects of the model, namely whether to estimate amino acid frequencies from the data and whether to include a gamma model of rate variation among sites (Hurvich and Tsai 1989). The Akaike Information Criterion (AIC) is a measure of the amount of information lost when a particular model is used to approximate a given situation. The second order AIC is used when the number of free parameters in the model is close to the number of sites in the alignment (Posada and Buckley 2004), as is the case here. The AIC_c is calculated as:

$$AIC_c = -2\ln L + 2K + \frac{2K(K+1)}{n-k-1}$$

where L is the likelihood of a given phylogenetic tree, K is the number of free parameters in the model being assumed and n is the number of sites in the alignment. An AIC_c score is calculated for each model under consideration. Models are then compared using the difference in AIC_c (Δ) calculated as follows:

$$\Delta_i = AIC_i - AIC_{min}$$

where AIC_{min} is the lowest obtained AIC_c value among candidate models (i.e. the AIC_c of the best model). Models with a $\Delta_i > 10$ can be confidently rejected as a worse fit to the data than the best model (Posada and Crandall 1998).

Four different models of protein evolution were considered, using the alignment of homeodomains shown in figure 1 (main text). These were, WAG, WAG+F (which allows amino acid frequencies to be estimated from the data), WAG+Γ (which allows rates of evolution to vary across sites according to a gamma distribution with 8 categories), and the WAG+F+Γ (which allows amino acid frequencies to be estimated, as well as allowing gamma distributed rates to be estimated).

The AIC_c calculation indicated that all models could be rejected in favour of the WAG+Γ model, in which rates of evolution are allowed to vary across sites according to a gamma distribution, but amino acid frequencies are not estimated from the data. For SOWHL tests using the JTT replacement matrix, the JTT+Γ model was used.

References

Hurvich CM, Tsai CL (1989) Regression and Time Series Model Selection in Small Samples. Biometrika 76(2): 297-307.

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8(3): 275-282.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14(9): 817-818.

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol 53(5): 793-808.

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18(5): 691-699.

Appendix 2

The following table shows details of the sequences used in all analyses. All alignments consisted of just the 60 amino acid homeodomain.

Supplementary Table 1 Sequences used in alignments.			
Species	**Abbreviation**	**Hox Orthology**	**GenBank Accession**
Mus musculus	Mmu_HoxA1	*Hox1*	NM_010449
Mus musculus	Mmu_HoxB1	*Hox1*	NM_008266
Branchiostoma floridae	Bfl_Hox1	*Hox1*	AB028206
Chaetopterus variopedatus	Cva_Hox1	*Hox1*	AF163856
Nereis virens	Nvi_Hox1	*Hox1*	AF151663
Mus musculus	Mmu_HoxA2	*Hox2*	NM_010451
Branchiostoma floridae	Bfl_Hox2	*Hox2*	AB028207
Mus musculus	Mmu_HoxB2	*Hox2*	NM_134032
Chaetopterus variopedatus	Cva_Hox2	*Hox2*	AF163857
Archegozetes longisetosus	Alo_pb	*Hox2*	AF071406
Mus musculus	Mmu_HoxA3	*Hox3*	NM_010452
Mus musculus	Mmu_HoxB3	*Hox3*	NM_010458
Branchiostoma floridae	Bfl_Hox3	*Hox3*	X68045
Chaetopterus variopedatus	Cva_Hox3	*Hox3*	AF163858
Nereis virens	Nvi_Hox3	*Hox3*	AF151665
Mus musculus	Mmu_HoxA4	*Hox4*	NM_008265
Mus musculus	Mmu_HoxB4	*Hox4*	NM_010459
Branchiostoma floridae	Bfl_Hox4	*Hox4*	AB028208
Nereis virens	Nvi_Dfd	*Hox4*	AF151666
Helobdella triserialis	Htr_Lox18	*Hox4*	AF137385
Mus musculus	Mmu_HoxA5	*Hox5*	NM_010453
Branchiostoma floridae	Bfl_Hox5	*Hox5*	Z35145

Mus musculus	Mmu_HoxB5	*Hox5*	NM_008268
Lingula anatina	Lan_Scr	*Hox5*	AF144674
Nereis virens	Nvi_Scr	*Hox5*	AF151667
Helobdella robusta	Hro_Lox2	*Hox6/7/8*	Z14313
Nereis virens	Nvi_Lox2	*Hox6/7/8*	AF151668
Nereis virens	Nvi_Lox4	*Hox6/7/8*	AF151669
Hirudo medicinalis	Hme_Lox4	*Hox6/7/8*	S79240
Nereis virens	Nvi_Lox5	*Hox6/7/8*	AF151671
Helobdella robusta	Hro_Lox5	*Hox6/7/8*	AF004387
Mus musculus	Mmu_HoxA6	*Hox6/7/8*	NM_010454
Mus musculus	Mmu_HoxB6	*Hox6/7/8*	NM_008269
Branchiostoma floridae	Bfl_Hox6	*Hox6/7/8*	Z35146
Mus musculus	Mmu_HoxB7	*Hox6/7/8*	NM_010460
Mus musculus	Mmu_HoxA7	*Hox6/7/8*	NM_010455
Branchiostoma floridae	Bfl_Hox7	*Hox6/7/8*	Z35147
Mus musculus	Mmu_HoxB8	*Hox6/7/8*	NM_010461
Mus musculus	Mmu_HoxC8	*Hox6/7/8*	NM_010466
Branchiostoma floridae	Bfl_Hox8	*Hox6/7/8*	Z35148
Mus musculus	Mmu_HoxA9	*Hox9+*	NM_010456
Mus musculus	Mmu_HoxB9	*Hox9+*	NM_008270
Branchiostoma floridae	Bfl_Hox9	*Hox9+*	Z35149
Lingula anatina	Lan_Post1	*Hox9+*	AF144679
Nereis virens	Nvi_Post1	*Hox9+*	AF151672
Mus musculus	Mmu_HoxA10	*Hox9+*	NM_008263
Mus musculus	Mmu_HoxC10	*Hox9+*	NM_010462
Branchiostoma floridae	Bfl_Hox10	*Hox9+*	Z35150
Lingula anatina	Lan_Post2	*Hox9+*	AF144680
Nereis virens	Nvi_Post2	*Hox9+*	AF151673
Mus musculus	Mmu_Gsh1	*Gsx*	NM_008178

Mus musculus	Mmu_Gsh2	*Gsx*	NM_133256
Branchiostoma floridae	Bfl_Gsx	*Gsx*	AF052463
Capitella species	Csp_Gsx	*Gsx*	DQ132894
Mus musculus	Mmu_Xlox	*Xlox*	NM_008814
Branchiostoma floridae	Bfl_Xlox	*Xlox*	AF052464
Capitella species	Csp_Xlox	*Xlox*	DQ102390
Phascolion strombi	Pst_Xlox	*Xlox*	AF363233
Mus musculus	Mmu_Cdx1	*Cdx*	NM_009880
Mus musculus	Mmu_Cdx2	*Cdx*	NM_007673
Capitella species	Csp_Cdx	*Cdx*	DQ102389
Branchiostoma floridae	Bfl_Cdx	*Cdx*	AF052465
Nereis virens	Nvi_Cdx	*Cdx*	AY117546
Mus musculus	Mmu_Evx1	*Evx*	NM_007966
Mus musculus	Mmu_Evx2	*Evx*	BC061467
Branchiostoma floridae	Bfl_Evx	*Evx*	AF374191
Helobdella robusta	Hro_Eve	*Evx*	AF409098
Theromyzon trizonare	Ttr_Eve	*Evx*	AY050275