compound heterozygous *noggin*[9] and *chordin* parents in a B6SJLF1 background (Jackson Laboratories).

Genotyping

DNA was extracted from extraembryonic membranes of E8.5 and older embryos. For E6.0 to E7.5 embryos the whole animal was digested overnight upon completion of *in situ* analysis and photography of each embryo. Digestion was at 55 °C in 50 mM Tris-HCl, pH 8, 100 mM EDTA, 100 mM NaCl, 1% SDS, 0.5 mg ml^{-1} Proteinase K. After adding NaCl to 1 M final concentration, the mix was centrifuged for 15 min, the supernatant recovered, and DNA precipitated with an equal volume of isopropanol. The genotype of the specimens was determined by polymerase chain reaction using the following primers: *chordin* primers were a mixture of three oligonucleotides, Chd-F (5′-GAGTTAGGAGGTGGAGCTCTACAC-3′), Chd-R (5′-GGTAGGAGACAGAGAAGCGTAAACT-3′) and Neo-2 (5′-GTTCCACATACACTTCATTCTCAG-3′), which yielded bands of 417 and 285 base pairs (bp) for the wild-type and mutant alleles respectively. *Noggin* primers were a mixture of three oligonucleotides, Nog-1 (5′-GCATGGAGCGCTGCCCCAGC-3′), Nog-2 (5′-GAGCAGCGAGCGCAGCAGCG-3′) and β-Gal (5′-AAGGGCGATCGGTGCGGGCC-3′), which yielded bands of 211 and 160 bp for the wild-type and mutant alleles respectively.

In situ hybridization

Whole-mount *in situ* hybridization was performed as described[18]. The *chordin* probe used was a subclone spanning nucleotides 145–2016 of the full-length mouse *chordin* complementary DNA (accession number AF096276), linearized with *Not*I and transcribed with T3 RNA polymerase. The *Krox20* probe was linearized with *Bam*H1 and transcribed with T3 RNA polymerase. *En-1*, *Hnf3β* and *Bmp-4* were linearized with *Cla*I, *Asp*700 and *Bam*H1 respectively, and transcribed with T7 RNA polymerase. Other antisense probes used were: *noggin* (ref. 9), *Shh* (ref. 9), *Six3* (ref. 11), *Hesx1* (ref. 12), *Lim1* (ref. 14) and *Cer-l* (ref. 18).

Received 21 September; accepted 25 November 1999.

1. Beddington, R. S. P. & Robertson, E. J. Axis development and early asymmetry in mammals. *Cell* **96,** 195–209 (1999).
2. Tam, P. L. & Behringer, R. R. Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev.* **68,** 3–25 (1997).
3. Bouwmeester, T., Kim, S. H., Sasai, Y., Lu, B. & De Robertis, E. M. Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. *Nature* **382,** 595–601 (1996).
4. Piccolo, S. *et al.* Cerberus induces head structures by binding to and inhibiting Nodal, BMP and Wnt signals in the extracellular space. *Nature* **397,** 707–710 (1999).
5. Zorn, A. M., Butler, K. & Gurdon, J. B. Anterior endomesoderm specification in *Xenopus* by Wnt/β-catenin and TGF-β signalling pathways. *Dev. Biol.* **209,** 282–297 (1999).
6. Zaraisky, A. G. *et al.* The homeobox-containing gene *XANF-1* may control development of the Spemann organizer. *Development* **121,** 3839–3847 (1995).
7. Piccolo, S., Sasai, Y., Lu, B. & De Robertis, E. M. Dorsoventral patterning in *Xenopus*: Inhibition of ventral signals by direct binding of Chordin to BMP-4. *Cell* **86,** 589–598 (1996).
8. Zimmerman, L. B., De Jesús-Escobar, J. M. & Harland, R. M. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* **86,** 599–606 (1996).
9. McMahon, J. A. *et al.* Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. *Genes Dev.* **12,** 1438–1452 (1998).
10. Chiang, C. *et al.* Cyclopia and defective axial patterning in mice lacking *Sonic hedgehog* gene function. *Nature* **383,** 407–413 (1996).
11. Oliver, G. *et al. Six-3*, a murine homologue of the *sine oculis* gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development. *Development* **121,** 4045–4055 (1995).
12. Thomas, P. & Beddington, R. Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Curr. Biol.* **6,** 1487–1496 (1996).
13. Dattani, M. T. *et al.* Mutations in the homeobox gene *HESX1/Hesx1* associated with septo-optic dysplasia in human and mouse. *Nature Genet.* **19,** 125–133 (1998).
14. Shawlot, W. & Behringer, R. R. Requirement for *Lim1* in head-organizer function. *Nature* **374,** 425–430 (1995).
15. Ang, S. L. & Rossant, J. *HNF-3*β is essential for node and notochord formation in mouse development. *Cell* **78,** 561–574 (1994).
16. Rhinn, M. *et al.* Sequential roles for *Otx2* in visceral endoderm and neuroectoderm for forebrain and midbrain induction and specification. *Development* **125,** 845–856 (1998).
17. Dufort, D., Schwartz, L., Harpal, K. & Rossant, J. The transcription factor HNF3β is required in visceral endoderm for normal primitive streak morphogenesis. *Development* **125,** 3015–3025 (1998).
18. Belo, J. A. *et al. Cerberus-like* is a secreted factor with neuralizing activity expressed in the anterior primitive endoderm of the mouse gastrula. *Mech. Dev.* **68,** 45–57 (1997).
19. Liu, P. *et al.* Requirement for *Wnt3* in vertebrate axis formation. *Nature Genet.* **22,** 361–365 (1999).
20. Litingtung, Y., Lei, L., Westphal, H. & Chiang, C. Sonic hedgehog is essential for the development of the foregut. *Nature Genet.* **20,** 58–61 (1998).
21. Motoyama, J. *et al.* Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus. *Nature Genet.* **20,** 54–57 (1998).
22. Meyers, E. N. & Martin, G. R. Differences in left–right axis pathways in mouse and chick: functions of FGF8 and SHH. *Science* **285,** 403–406 (1999).
23. Hammerschmidt, M., Serbedzija, G. N. & McMahon, A. P. Genetic analysis of dorsoventral pattern formation in the zebrafish: requirement of a BMP-like ventralizing activity and its dorsal repressor. *Genes Dev.* **10,** 2452–2461 (1996).
24. Schulte-Merker, S., Lee, K. J., McMahon, A. P. & Hammerschmidt, M. The zebrafish organizer requires *chordino*. *Nature* **387,** 862–863 (1997).
25. Lyons, K. M., Hogan, B. L. M. & Robertson, E. J. Colocalization of BMP 7 and BMP 2 RNAs suggests that these factors cooperatively mediate tissue interactions during murine development. *Mech. Dev.* **50,** 71–83 (1995).
26. Arkell, R. & Beddington, R. S. P. BMP-7 influences pattern and growth of the developing hindbrain of mouse embryos. *Development* **124,** 1–12 (1997).
27. Solloway, M. J. & Robertson, E. J. Early embryonic lethality in *Bmp5; Bmp7* double-mutant mice suggests functional redundancy within the 60A subgroup. *Development* **128,** 1753–1768 (1999).
28. Golden, J. A *et al.* Ectopic bone morphogenetic proteins 5 and 4 in the chicken forebrain lead to cyclopia and holoprosencephaly. *Proc. Natl Acad. Sci. USA* **54,** 623–634 (1999).
29. Furuta, Y., Piston, D. W. & Hogan, B. L. M. Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. *Development* **124,** 2203–2212 (1997).
30. Piccolo, S. *et al.* Cleavage of Chordin by the Xolloid metalloprotease suggests a role for proteolytic processing in the regulation of Spemann organizer activity. *Cell* **91,** 407–416 (1997).

Acknowledgements

We thank R. Beddington, G. Oliver, D. Wilkinson and A. Joyner for gifts of plasmids, K. Woo, C. De Robertis, B. Yoon and A. Burnett for help with genotyping, and K. Lyons, G. Weinmaster, E. Delot and members of our laboratories for comments on the manuscript. This work was supported by the HHMI, NIH, MRC Canada, Whitehead Foundation and the Norman Sprague Chair.

Correspondence and requests for materials should be addressed to E.M.D.R.

Maintenance of functional equivalence during paralogous Hox gene evolution

Joy M. Greer, John Puetz, Kirk R. Thomas & Mario R. Capecchi

Howard Hughes Medical Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA

Biological diversity is driven mainly by gene duplication followed by mutation and selection. This divergence in either regulatory or protein-coding sequences can result in quite different biological functions for even closely related genes. This concept is exemplified by the mammalian Hox gene complex, a group of 39 genes which are located on 4 linkage groups, dispersed on 4 chromosomes[1–4]. The evolution of this complex began with amplification in *cis* of a primordial Hox gene to produce 13 members, followed by duplications in *trans* of much of the entire unit. As a consequence, Hox genes that occupy the same relative position along the 5′ to 3′ chromosomal coordinate (*trans*-paralogous genes) share more similarity in sequence and expression pattern than do adjacent Hox genes on the same chromosome. Studies in mice indicate that although individual family members may have unique biological roles, they also share overlapping functions with their paralogues[5–12]. Here we show that the proteins encoded by the paralogous genes, *Hoxa3* and *Hoxd3*, can carry out identical biological functions, and that the different roles attributed to these genes are the result of quantitative modulations in gene expression.

The paralogous genes *Hoxa3*, *Hoxb3* and *Hoxd3* show virtually identical expression patterns of messenger RNA and share, in a pairwise comparison, ~50% identity in protein-coding sequences[13–17]. Nevertheless, mice lacking either a functional *Hoxa3* or *Hoxd3* gene show no obvious overlap in phenotype: the former have deficiencies in pharyngeal tissues derived from mesenchymal neural crest[18], the latter in somitic, mesodermally derived tissues of the axial skeleton[19]. Although the uniqueness of the single-mutant phenotypes implies a different qualitative role for each gene product, the analysis of animals carrying different combinations of mutant alleles of the *Hox3* paralogues suggests that there is also functional overlap between the three genes, and that the important parameters that mediate these interactions may be quantitative in nature[5,17,20–22]. For example, in animals that are

homozygous for null alleles of *Hoxd3*, the defect in the formation of the atlas is primarily restricted to the anterior arch[19]. The severity of this phenotype is increased quantitatively by a serial reduction in the number of functional copies of either the *Hoxa3* or the *Hoxb3* genes. In the $Hoxd3^{-/-}$ background, absence of one copy of either paralogue reduces the size of the entire atlas by ~50%; absence of any two copies, either $Hoxa3^{-/-}$ or $Hoxb3^{-/-}$, or $Hoxa3^{+/-}$;$Hoxb3^{+/-}$, eliminates the entire atlas[5,17].

To examine directly the nature of functional overlap[8,23,24] within the *Hox3* family, we have exchanged reciprocally, in the genome of mice, the protein-coding portions of the *Hoxa3* and *Hoxd3* genes (Fig. 1a). Thus, we have generated mice that lack any HoxA3 protein, but instead express the HoxD3 protein from both the *Hoxa3* and *Hoxd3* loci (Fig. 1b), as well as mice that lack HoxD3 protein but express HoxA3 from both loci. The alleles used in this study are shown in Fig. 1c and include the two wild-type alleles, $Hoxa3^{+}$ and $Hoxd3^{+}$; the null alleles, $Hoxa3^{null}$ and $Hoxd3^{null}$; and the paralogous replacement alleles, $Hoxa3^{D3}$ and $Hoxd3^{A3}$. The null allele of *Hoxa3* contains a translation stop codon in the first exon and a frameshift mutation in the second exon. The *Hoxd3* null allele contains a deletion/insertion encompassing the homeodomain-coding region of the second exon. The $Hoxa3^{D3}$ allele was generated by replacing the protein-coding domains of both *Hoxa3* exons with the corresponding sequences from the *Hoxd3* gene. All sequences flanking the protein-coding domains, 5′ of the ATG, the entire intron and 3′ of the TGA termination codon, are from the *Hoxa3* locus. Similarly, the $Hoxd3^{A3}$ locus, in which the HoxA3 protein is synthesized from the *Hoxd3* locus, retains all the surrounding *Hoxd3* nucleotides. All alleles were introduced into their appropriate genomic loci by homologous recombination in murine ES cells[25]. It should also be noted that the selectable markers (the neo^{r} genes), which are required for isolation of the targeted mutations, have all been replaced with a single, 32-base-pair (bp) *loxP* element[26], thus eliminating potential effects of marker gene expression on neighbouring genes.

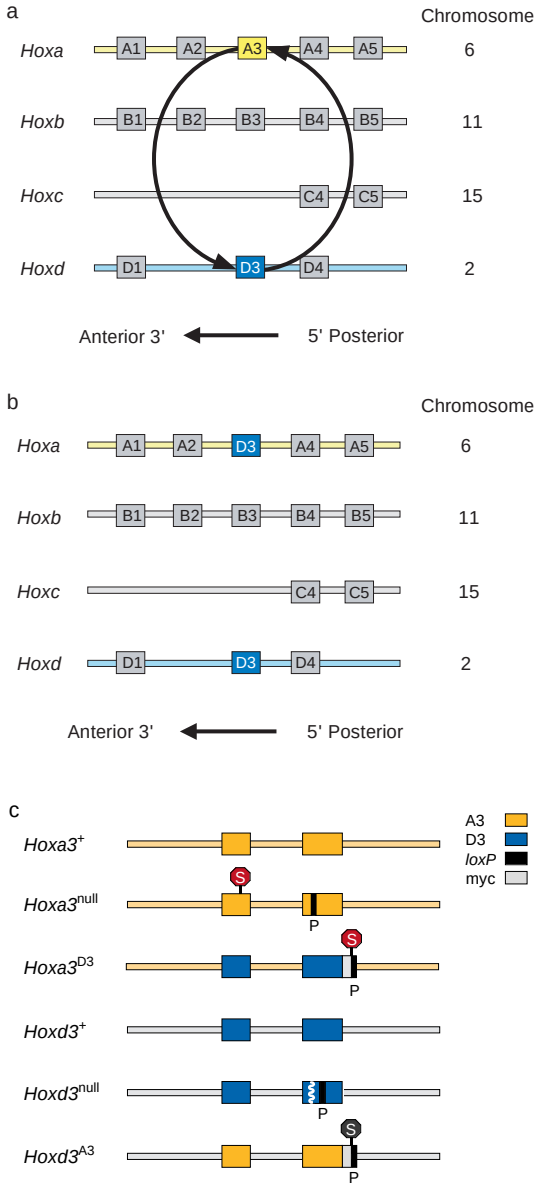

Figure 1 Representation of the wild-type (**a**) and one of the mutant Hox gene clusters (**b**). The individual alleles used are shown in (**c**). **a**, The 3′ end of the wild-type Hox cluster, showing the relationship of paralogous genes. The black arrows indicate the paralogous protein-coding regions that were transplaced. **b**, The 3′ end of the Hox cluster in animals that express the HoxD3 protein from the *Hoxa3* locus, referred to as the $Hoxa3^{D3}$ allele. **c**, The six alleles used in this study. From top to bottom the alleles are wild-type *Hoxa3*, $Hoxa3^{null}$, $Hoxa3^{D3}$, wild-type *Hoxd3*, $Hoxd3^{null}$ and $Hoxa3^{A3}$. Solid boxes represent protein-coding sequences, with those of HoxA3 and HoxD3 shown in dark yellow and dark blue, respectively. The *Hoxa* flanking non-coding sequences are in pale yellow, and *Hoxd* *cis*-acting regulatory sequences are in light blue. The homeodomain deletion in the *Hoxd3* null allele is white. Stop codons (S), *loxP* sites and the myc epitopes are indicated.

Mice homozygous for null alleles at *Hoxa3* are characterized by perinatal lethality, absence of the thymus, and malformation of the hyoid bone[17,18,20,21]. The glandular and skeletal abnormalities are exacerbated by mutations in the *Hoxd3* locus[5,17,21]. To confirm that our null allele for *Hoxa3*, without the neo^{r} gene, conferred the originally described mutant phenotype, we crossed animals heterozygous for wild-type and null alleles ($Hoxa3^{+/null} \times Hoxa3^{+/null}$). As shown in Table 1, no adult progeny of the $Hoxa3^{null/null}$ genotype were recovered. As a test of whether the HoxD3 protein is sufficient to complement this *Hoxa3* deficiency, animals heterozygous for the wild-type and the *Hoxd3* transplacement alleles ($Hoxa3^{D3/+}$) were crossed with animals heterozygous for the wild-type and null alleles ($Hoxa3^{+/null}$). Adult progeny recovered from this cross represented all four possible genotypes at roughly normal mendelian frequencies (Table 1). The viability of the compound heterozygotes, $Hoxa3^{D3/null}$, shows that the HoxD3 protein, when expressed at the *Hoxa3* locus, complements the absence of HoxA3 protein.

To determine the extent of this complementation, embryos representing all *Hoxa3* allelic combinations were examined histologically. At embryonic day 17.5 (E17.5), the most striking abnormality of homozygous null ($Hoxa3^{null/null}$) embryos is the complete

Table 1 $Hoxa3^{D3}$ rescues the lethality of $Hoxa3^{null/null}$

Parental genotypes	$Hoxa3^{+/null} \times Hoxa3^{+/null}$			
Possible genotypes of progeny	+/+	+/null	null/null	
Number of progeny observed	13	28	0	
Parental genotypes	$Hoxa3^{+/null} \times Hoxa3^{D3/+}$			
Possible genotypes of progeny	+/+	+/null	D3/+	D3/null
Number of progeny observed	12	13	18	10

Table 2 Fraction of mice exhibiting normal glandular and skeletal structures

Genotype at the *Hoxa3* locus	+/+	null/null	D3/null	D3/D3	
Thymus	4/4	0/5	5/5	3/3	
Hyoid	18/18	0/7	11/11	6/6	
Genotype at the *Hoxd3* locus	+/+	+/null	null/null	A3/A3	A3/null
Anterior arch of the atlas	8/8	10/10	0/7	10/10	8/9
Dens	8/8	9/10	1/7	10/10(9/10)*	9/9 (6/9)*

* Indicates number of mice in which complementation was complete. Incomplete complementation of the *Hoxd3* null mutation by a single copy of $Hoxd3^{A3}$ may result from a number of independent factors. It could represent the 15% haplo-insufficiency normally seen in *Hoxd3* heterozygotes[19] (10% in this study), it may reflect a depression in protein activity due to the myc-epitope tag at the carboxy terminus of the $Hoxd3^{A3}$ protein, and/or it may reflect slight differences in the biological activities of HoxA3 and HoxD3 in bone morphogenesis.

absence of the thymus (Fig. 2, Table 2). However, replacement of one or both copies of HoxA3 protein with the HoxD3 protein restores this organ. The effect of the various *Hoxa3* mutations on the throat cartilage was also assessed. Alterations of the hyoid, characteristic of embryos homozygous for the null allele of *Hoxa3*, are reversed by expression of the HoxD3 protein at the *Hoxa3* locus, as shown in Fig. 2 and summarized in Table 2. Moreover, adult animals lacking the HoxA3 protein but carrying one or two copies of the $Hoxa3^{D3}$ allele appear anatomically completely normal (data not shown).

A conclusion from this data is that the HoxD3 protein is functionally equivalent to the HoxA3 protein if it is expressed in the context of the *Hoxa3* gene. A corollary to this conclusion is that the HoxA3 protein, if expressed at the *Hoxd3* locus will be unable to complement null mutations at *Hoxa3*. To test this, we crossed mice heterozygous for the *Hoxa3* mutation, $Hoxa3^{+/null}$, with mice heterozygous for the same mutation but that also carried the HoxA3 protein-coding domains inserted at the *Hoxd3* locus ($Hoxa3^{+/null}$;$Hoxd3^{A3/null}$). No adult progeny homozygous for the *Hoxa3* null allele were ever recovered from this cross. However, five dead newborn animals of the genotype $Hoxa3^{null/null}$;$Hoxd3^{A3/+}$ were analysed for the status of the throat bones and cartilage. As shown in Fig. 2, these structures were indistinguishable from those found in the $Hoxa3^{null/null}$ embryos, showing that when the HoxA3 protein is expressed from the *Hoxd3* locus, it is incapable of rescuing the *Hoxa3* mutant phenotype. Embryos of the same genotype, $Hoxa3^{null/null}$;$Hoxd3^{A3/+}$, were also examined histologically, and no thymus was detected (Fig. 2).

Hoxd3 is required for the proper development of the cervical vertebrae[19]. Mice homozygous for a *Hoxd3* null mutation lack the anterior arch of the atlas and the dens of the axis. To assess the ability of the HoxA3 protein to function during vertebral morphogenesis, we replaced the protein-coding sequences of the *Hoxd3* gene with *Hoxa3* protein-coding sequences, leaving all adjacent, non-coding sequences intact. We were able to generate viable, fertile adult animals in which the three *Hoxd3* alleles, $Hoxd3^{+}$, $Hoxd3^{null}$ and $Hoxd3^{A3}$, were represented in all six possible combinations. Skeletal analyses were carried out on adult animals of each genotype with emphasis on the structure of the atlas and axis (Fig. 3, Table 2). Mice homozygous for the null allele lack the anterior arch of the atlas as well as the dens. Both of these structures were restored by the presence of two copies of the $Hoxd3^{A3}$ allele. A single copy of the $Hoxd3^{A3}$ allele also complements the mutant phenotype, but with incomplete penetrance. In three of nine animals of the $Hoxd3^{A3/null}$ genotype, the dens was fused (either completely or partially) to the anterior arch of the adjacent atlas; in one of these same individuals, the arch itself was incompletely formed. Neither the atlas nor axis phenotype was suppressed by the expression of the HoxD3 protein at the *Hoxa3* locus (data not shown), showing again the critical role of *cis*-acting sequences in determining the functional specificity of the gene products.

The hybrid alleles generated in this study are characterized by the precise engineering that distinguishes between coding and non-coding sequences, and by the targeting of single copies of the coding moieties to the chromosomal loci of their paralogues. Promoters, enhancers, intron-encoded splicing signals and 3′ mRNA processing signals are thus the ones that are normal for the host locus. We propose that this provides the most precise means of retaining the normal temporal, spatial and quantitative parameters of gene expression.

The successful bi-directional complementation of both the *Hoxa3* and *Hoxd3* genes has proved informative on several counts. First, it shows that these proteins, which share less than 50% identity in amino-acid sequence, are capable of carrying out equivalent biological functions in the processes recognized to require *Hox3* gene activity. Second, it provides direct evidence that the different roles played by these genes during embryogenesis are mainly the result of

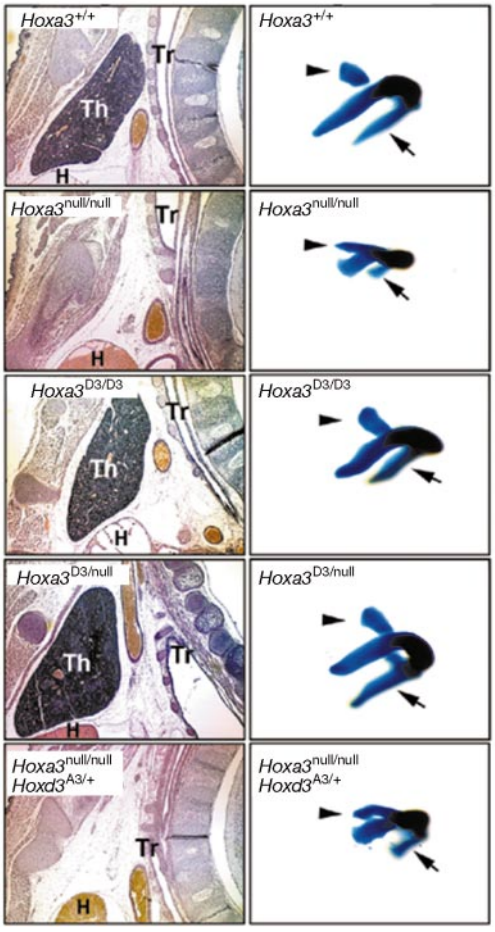

Figure 2 Thymus and hyoid phenotypes of *Hoxa3* null mutants and complementation by the $Hoxa3^{D3}$ allele. Left, parasagittal sections from the throat regions of E17.5 embryos. Right, dissected hyoid bones from either E18.5 embryos or newborn animals. Arrows and arrowheads indicate the lesser and greater horns of the hyoid bone, respectively. Th, thymus; Tr, trachea; H, heart.

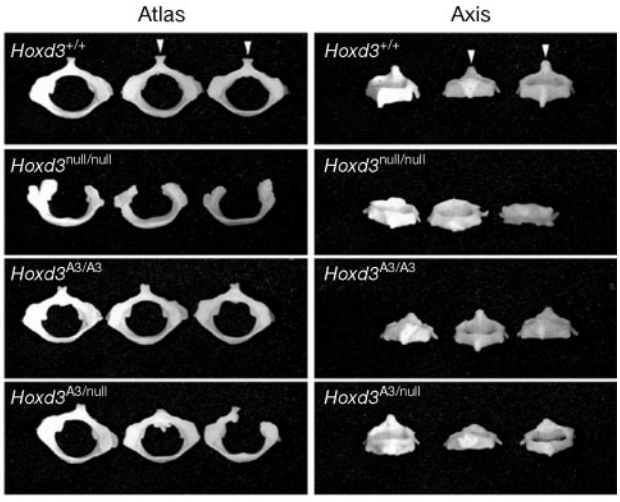

Figure 3 Atlas and axis phenotypes of *Hoxd3* null mutants and complementation by the $Hoxd3^{A3}$ allele. The atlas and axis from three animals of each genotype are shown. The atlas and axis occupying the same relative positions in each panel are from the same animal. Arrowheads denote the anterior arch of the atlas (left panel) and the dens of the axis (right panel).

cis-acting sequences that modulate expression of the individual loci. Whether equivalent scenarios hold for other paralogous Hox gene families remains to be tested. We want to stress that the non-equivalence of *Hoxa3* and *Hoxd3* in their normal, wild-type, chromosomal environments may not be due to qualitative differences in the spectrum of cell-specific expression. The expression patterns of *Hoxa3*, *Hoxb3* and *Hoxd3* are very similar, if not completely overlapping. Rather, we suggest that the important parameter controlling, for example, the differential function of *Hoxa3* and *Hoxd3* in mesenchymal neural crest and somitic mesoderm, respectively, may be that each gene is expressed in particular cell types at sufficient levels to activate its respective developmental programme; *Hoxa3* being expressed at a higher concentration in one context and *Hoxd3* in another. In this model, qualitative differences in the behaviour of genes are the consequence of quantitative differences in their expression levels. Further, the cross-exacerbations of the mutant phenotypes in the double mutants argue for co-expression of these paralogues in all the tissues of the neck region and that they function coordinately. At the molecular level, these observations suggest that these Hox genes mediate their functions by collectively controlling common target genes through conserved target *cis* elements. The outputs from the target genes would thus reflect signals generated from the combined binding activities of all of the paralogous Hox3 products. The sensitivity of these developmental programmes to the overall combined concentration of paralogous gene products is demonstrated by the observation that, although mice heterozygous for individual *Hox3* paralogous genes are outwardly normal, most mice heterozygous at all three *Hox3* loci die perinatally[17].

It is remarkable that in the estimated 530 million years since quadruplication of the Hox complex in early vertebrates, the near perfect functional equivalence described here has been maintained for the HoxA3 and HoxD3 proteins. We suggest that this conservation of functional equivalence may not be an accident of evolution, but rather an indication that equivalence of function is at the very core of normal Hox gene function during development, perhaps facilitating a necessary quantitative interaction among these proteins. □

Methods

Allele generation

The null allele of *Hoxa3*, *Hoxa3*^{null}, has been described[26]. It contains an 8-bp insertion at the unique *Eco47*III site in exon 1 and a *loxP*-containing insertion at the *Bgl*II site in the homeodomain of exon 2. The null allele of *Hoxd3*, *Hoxd3*^{null}, contains a *loxP* site instead of the 249 nucleotides between the two *Bgl*II sites in exon 2. This mutant allele was created by the self-excising *neo*^{r} cassette, ACN, also used to generate *Hoxa3*^{null}. The exchange of coding sequences required to generate *Hoxa3*^{D3} and *Hoxd3*^{A3} first necessitated the sequencing of all junctions between coding and non-coding regions of both genes. The sequences are as follows:

Hoxa3 exon 1, 5′-aaacatcgcg**ATG**CAAAAAG.....TCCAGCTCAGgtaaatgaat-3′;

Hoxa3 exon 2, 5′-gtcctggcagGGGAGAGCTG.....CACCCACCTGtgatgatggg-3′;

Hoxd3 exon 1, 5′-tgagtgcacc**ATG**CAGAAGG.....GCCACTTCAGgtagttctct-3′;

Hoxd3 exon 2, 5′-ccttgcccagGAGAGAACTG.....GACACATCTGtagctgttgc-3′;

where lower and upper case letters indicate non-coding and coding sequences, respectively, and start- and stop-translation codons are indicated by bold type and underline, respectively.

Restriction endonuclease cleavage sites were identified in all regions 5′ and 3′ to each coding region and used to remove the coding sequences and linked nucleotides from cloned copies of the two genes. Combinations of synthetic oligonucleotides, restriction-endonuclease-generated fragments and PCR-generated fragments were used to reconstruct the transplaced coding sequences and any missing host, flanking nucleotides. These were then re-inserted into the appropriate sites. Exons and flanking sequences were sequenced before construction of the final targeting vectors. Both the Hoxa3^{D3} and Hoxd3^{A3} coding regions were fused at the 3′ end to an 11-amino-acid c-myc epitope, and a *loxP*-flanked *neo*^{r} gene was inserted immediately 3′ of the termination codon in exon 2 of each gene. The alleles were introduced into mouse ES cell line RI[27] and subjected to positive–negative selection[28] before identification of recombinant cell lines. Two cell lines representing each allele were used to generate mice by blastocyst injection. The *neo*^{r} marker gene was removed by mating the progeny of the chimaeric animals with a *Cre*-expressing 'deleter mouse'[29] before colony expansion and phenotypic analysis.

Phenotypic analysis of embryos

The characterization of the pharyngeal organs in E17.5 embryos was carried out after serial sectioning as previously described[30]. Embryos from timed pregnancies were fixed in Bouin's reagent for 4 days, rinsed in a sucrose/phosphate solution and hemisected with a razor blade. The samples were then embedded in paraffin, sectioned sagittally (10μm) and stained on glass slides with haematoxylin and eosin. E18.5 day embryos used for skeletal analysis were fixed in ethanol, stained with alazarin red and alcian blue, and cleared in glycerol as described[6].

Preparation of adult skeletons

Six- to ten-week-old adult animals were killed by CO_2 asphyxiation and the head and cervical region removed. After removal of the skin, the samples were boiled for 5 min in water and incubated overnight in a solution of 0.5% trypsin and 30% borax. The sample was boiled again in water for 10 min, and the cleaned bones stored in 95% ethanol.

Received 21 October; accepted 25 November 1999.

1. Scott, M. P. Vertebrate homeobox gene nomenclature. *Cell* **71,** 551–553 (1992).
2. Kappen, C., Schughart, K. & Ruddle, F. H. Two steps in the evolution of *Antennapedia*-class vertebrate homeobox genes. *Proc. Natl Acad. Sci. USA* **86,** 5459–5463 (1989).
3. Duboule, D. & Dollé, P. The structural and functional organization of the murine *Hox* gene family resembles that of *Drosophila* homeotic genes. *EMBO J.* **8,** 1497–1505 (1989).
4. Holland, P. W. H. & Garcia–Fernandez, J. *Hox* genes and chordate evolution. *Dev. Biol.* **173,** 382–395 (1996).
5. Condie, B. G. & Capecchi, M. R. Mice with targeted disruptions in the paralogous genes *hoxa-3* and *hoxd-3* reveal synergistic interactions. *Nature* **370,** 304–307 (1994).
6. Davis, A. P., Witte, D. P., Hsieh–Li, H. M., Potter, S. S. & Capecchi, M. R. Absence of radius and ulna in mice lacking *hoxa-11* and *hoxd-11*. *Nature* **375,** 791–795 (1995).
7. Horan, G. S. B. *et al.* Compound mutants for the paralogous *hoxa-4, hoxb-4,* and *hoxd-4* genes show more complete homeotic transformations and a dose-dependent increase in the number of vertebrae transformed. *Genes Dev.* **9,** 1667–1677 (1995).
8. Zákány, J., Gérard, M., Favier, B., Potter, S. S. & Duboule, D. Functional equivalence and rescue among group 11 *Hox* gene products in vertebral patterning. *Dev. Biol* **176,** 325–328 (1996).
9. Gavalas, A. M. *et al. Hoxa1* and *Hoxb1* synergize in patterning the hindbrain, cranial nerves and second pharyngeal arch. *Development* **125,** 1123–1136 (1998).
10. Studer, M. *et al.* Genetic interactions between *Hoxa1* and *Hoxb1* reveal new roles in regulation of early hindbrain patterning. *Development* **125,** 1025–1036 (1998).
11. Fromental–Ramain, C. *et al. Hoxa-13* and *Hoxd-13* play a crucial role in the patterning of the limb autopod. *Development* **122,** 2997–3011 (1996).
12. Rossel, M. & Capecchi, M. R. Mice mutant for both *Hoxa1* and *Hoxb1* show extensive remodeling of the hindbrain and defects in craniofacial development. *Development* **126** (in the press).
13. Lonai, P., Arman, E., Czosnek, H., Ruddle, F. H. & Blatt, C. New murine homeoboxes: structure, chromosomal assignment and differential expression in adult erythropoiesis. *DNA* **6,** 409–418 (1987).
14. Hunt, P. *et al.* A distinct *Hox* code for the branchial region of the vertebrate head. *Nature* **353,** 861–864 (1991).
15. Hunt, P. *et al.* The branchial *Hox* code and its implications for gene regulation, patterning of the nervous system and head evolution. *Dev. Suppl.* **2,** 63–77 (1991).
16. Sham, M. H. *et al.* Analysis of the murine *Hox-2.7* gene: conserved alternative transcripts with differential distributions in the nervous system and the potential for shared regulatory regions. *EMBO J.* **11,** 1825–1836 (1992).
17. Manley, N. R. & Capecchi, M. R. *Hox* group 3 paralogous genes act synergistically in the formation of somitic and neural crest–derived structures. *Dev. Biol.* **192,** 274–288 (1997).
18. Chisaka, O. & Capecchi, M. R. Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *hox-1.5*. *Nature* **350,** 473–479 (1991).
19. Condie, B. G. & Capecchi, M. R. Mice homozygous for a targeted disruption of *Hoxd-3* (*Hox-4.1*) exhibit anterior transformations of the first and second cervical vertebrae, the atlas and the axis. *Development* **119,** 579–595 (1993).
20. Manley, N. R. & Capecchi, M. R. The role of *hoxa-3* in mouse thymus and thyroid development. *Development* **121,** 1989–2003 (1995).
21. Manley, N. R. & Capecchi, M. R. *Hox* group 3 paralogs regulate the development and migration of the thymus, thyroid and parathyroid glands. *Dev. Biol.* **195,** 1–15 (1998).
22. Capecchi, M. R. *Hox* genes and mammalian development. *Cold Spring Harb. Symp. Quant. Biol.* **62,** 273–281 (1997).
23. Fitzgerald, K., Wilkinson, H. A. & Greenwald, I. *Glp-1* can substitute for *lin-12* in specifying cell fate decisions in *Caenorhabditis elegans. Development* **119,** 1019–1027 (1993).
24. Hanks, M., Wurst, M., Anson–Cartwright, L., Auerbach, A. B. & Joyner, A. L. Rescue of the *En–1* mutant phenotype by replacement of *En-1* with *En-2*. *Science* **269,** 679–682 (1995).
25. Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51,** 503–512 (1987).
26. Bunting, M., Bernstein, K. E., Greer, J. M., Capecchi, M. R. & Thomas, K. R. Targeting genes for self-excision in the germline. *Genes Dev.* **13,** 1524–1528 (1999).
27. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90,** 8424–8428 (1993).
28. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336,** 348–352 (1988).
29. Schwenk, F., Baron, U. & Rajewsky, K. A *cre*-transgenic mouse strain for the ubiquitous deletion of *loxP*-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23,** 5080–5081 (1995).
30. Thomas, K. R., Musci, T. S., Neumann, P. E. & Capecchi, M. R. *Swaying* is a mutant allele of the proto-oncogene *Wnt-1*. *Cell* **67,** 969–976 (1991).

Acknowledgements

We thank all members of the Capecchi laboratory's tissue culture support group and animal care facility for their expertise. Assistance from L. Oswald, P. Reid and D. Lim for manuscript preparation, and R. Beglarian for histology is appreciated. J.M.G. was supported by the Dee Fellowship and a NIH Genetics Training Grant.

Correspondence and requests for materials should be addressed to M.R.C. (e-mail: mario.capecchi@genetics.utah.edu).

The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences

J. Parkhill*, B. W. Wren†, K. Mungall*, J. M. Ketley‡, C. Churcher*, D. Basham*, T. Chillingworth*, R. M. Davies*, T. Feltwell*, S. Holroyd*, K. Jagels*, A. V. Karlyshev†, S. Moule*, M. J. Pallen§, C. W. Penn‖, M. A. Quail*, M-A. Rajandream*, K. M. Rutherford*, A. H. M. van Vliet¶, S. Whitehead* & B. G. Barrell*

* *The Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*
† *Department of Infectious and Tropical Diseases, The London School of Hygiene and Tropical Medicine, Keppel St, London WCIE 7HT, UK*
‡ *Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK*
§ *Department of Microbiology and Immunobiology, Queen's University Belfast, Grosvenor Road, Belfast BT12 6BN, UK*
‖ *School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK*
¶ *Departments of Medical Microbiology and Gastroenterology, Faculty of Medicine, Vrije Universiteit, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands*

***Campylobacter jejuni*, from the delta-epsilon group of proteobacteria, is a microaerophilic, Gram-negative, flagellate, spiral bacterium—properties it shares with the related gastric pathogen *Helicobacter pylori*. It is the leading cause of bacterial food-borne diarrhoeal disease throughout the world[1]. In addition, infection with *C. jejuni* is the most frequent antecedent to a form of neuromuscular paralysis known as Guillain–Barré syndrome[2]. Here we report the genome sequence of *C. jejuni* NCTC11168. *C. jejuni* has a circular chromosome of 1,641,481 base pairs (30.6% G+C) which is predicted to encode 1,654 proteins and 54 stable RNA species. The genome is unusual in that there are virtually no insertion sequences or phage-associated sequences and very few repeat sequences. One of the most striking findings in the genome was the presence of hypervariable sequences. These short homopolymeric runs of nucleotides were commonly found in genes encoding the biosynthesis or modification of surface structures, or in closely linked genes of unknown function. The apparently high rate of variation of these homopolymeric tracts may be important in the survival strategy of *C. jejuni*.**

Human infection is usually acquired by the consumption of contaminated food (especially poultry) or water[1]. Motile campylobacters colonize the intestines of a wide range of animals, but in immunologically naive humans infection frequently results in an inflammatory enterocolitis. The number of cases of *Campylobacter* infection reported in England and Wales in 1998 increased by 17% from the previous year, with the number of reported cases now more than double that due to *Salmonella*[3]. Despite its importance, effective control of *Campylobacter* in the food chain and the design of disease prevention strategies are hindered by a poor understanding of the genetics, physiology and virulence of this organism.

The genome of *C. jejuni* NCTC11168 is 1,641,481 base pairs (bp) in length. Of the 1,654 predicted coding sequences (CDS), at least 20 probably represent pseudogenes; the average gene length is 948 bp, and 94.3% of the genome codes for proteins, making it the densest bacterial genome sequenced to date. The bias towards G on the leading strand of the chromosome[4] indicates that the origin of replication is near to the start of the *dnaA* gene. Strand bias is also evident in the CDSs; overall, 61.1% are transcribed in the same direction as replication (Fig. 1). We discovered two large regions of lower G+C content that encompass CDSs Cj1135–Cj1148 (25.4%) and Cj1421–Cj1442 (26.5%); these correspond to genes within the lipooligosaccharide (LOS) and extracellular polysaccharide (EP) biosynthesis clusters, respectively. Functional information (matches to genes of known function, or informative hydrophobicity profiles) could be deduced for 77.8% of the 1,654 CDSs, wheras 13.5% matched genes of unknown function in the database and 8.7% had no database match, or other functional information. The unusually low number of unknowns reflects the preponderance of predicted membrane, periplasmic and lipoproteins; these make up 10.3%, 7.8% and 2.3 % of the CDSs, respectively, and many of these have no database matches.

One surprising feature of the *C. jejuni* genome is the almost complete lack of repetitive DNA sequences. In fact, there are only four repeated sequences within the entire genome; three copies of the ribosomal RNA operon (6 kilobases (kb)) and three duplicated or triplicated CDSs. Apart from Cj0752, which is similar to part of IS605 *tnpB* from *H. pylori*, there is no evidence of any functional inserted sequence (IS) elements, transposons, retrons or prophages

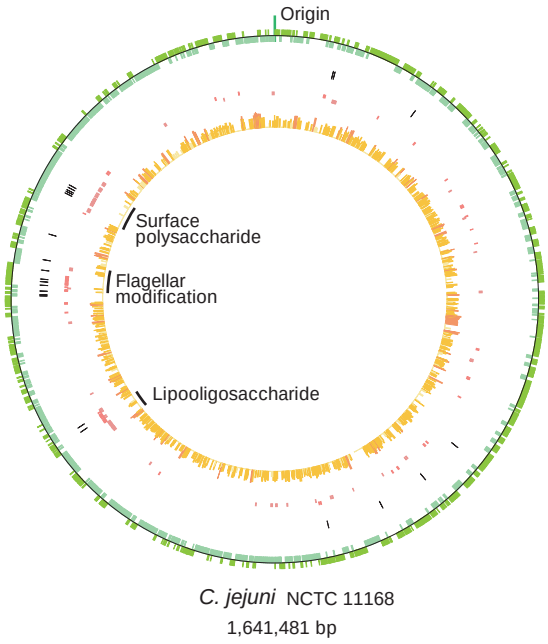

Figure 1 Circular representation of the *C. jejuni* genome. From the outside to the inside, the first circle shows coding sequences transcribed in the clockwise direction in dark green; the second shows coding sequences in the anticlockwise direction in pale green. The putative origin of replication is marked. The third shows the positions of hypervariable sequences in black, and the fourth and fifth show genes involved in the production of surface structures: clockwise in dark red and anticlockwise in pale red. The innermost histogram shows the similarity of each gene to its *H. pylori* orthologue, where present; the height of the bar, and the intensity of the colour, are proportional to the degree of similarity. The clusters of genes responsible for LOS biosynthesis, EP biosynthesis and flagellar modification are marked.