Genomes & Developmental Control

Transgenic analysis of Hoxd gene regulation during digit development

Federico Gonzalez, Denis Duboule⁎, François Spitz ⁎

National Center for Competence in Research “Frontiers in Genetics”, Department of Zoology and Animal Biology, University of Geneva, Sciences III, Quai Ernest Ansermet 30, 1211 Geneva 4, Switzerland

Received for publication 11 January 2007; revised 27 February 2007; accepted 19 March 2007
Available online 23 March 2007

Abstract

In tetrapods, posterior Hoxd genes (from groups 10 to 13) are necessary to properly pattern the developing autopods, including the number and identities of digits. Their coordinated expression is achieved by sharing a global control region (GCR), which was isolated and localized 200 kb 5′ (centromeric) of the gene cluster. However, in transgenic assays, the GCR was unable to fully recapitulate all aspects of the endogenous Hoxd expression patterns during distal limb development. In this paper, we further analyze the regulatory potential of this locus and report the characterization of Prox, a second enhancer element that contributes to the transcriptional activity of posterior Hoxd genes in developing distal limb buds. We show that the GCR and Prox elements complement each other and work in combination to correctly establish the late phase of Hoxd genes expression. Based on DNA sequence conservation and transgenic assays, we discuss the functions of these regulatory regions as well as a potential evolutionary scheme accounting for their emergence along with the evolution of tetrapod limbs.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Hox genes; Global and shared enhancers; Regulatory landscapes; Limb morphogenesis; Evolution of gene regulation

Introduction

Genes of the Hox family encode transcription factors essential for proper morphogenesis along the major body axis, where their precise transcription in time and space determines the fate of a variety of structures derived from all three germ layers. The property of these gene products to organize our body plan is ancient, and most animals showing a bilateral symmetry implement this genetic process, to some extent (see Deschamps and van Nes, 2005). In vertebrates, this gene family was amplified following genomic duplications and four clusters of genes are found in tetrapods; HoxA, B, C and D. This increase in number of Hox gene clusters likely allowed for subsequent functional co-options of some clusters in parallel with the emergence of tetrapod-specific traits. The paradigm of such an evolutionary process is the functional recruitment of both HoxA and HoxD cluster genes to organize the development of our limbs and external genital organs (Dollé et al., 1989, 1991; Haack and Gruss, 1993; Zakany et al., 1997).

The function of Hoxa and Hoxd genes during limb development has been largely documented, in particular using mouse molecular genetics. Within these two clusters, genes belonging to paralogy groups 9 to 13, i.e. related in sequence to the Drosophila Abdominal B gene (Abd-B), are activated along with limb bud outgrowth, and are expressed in progressively more distal domains that mirror the physical sequence of genes along the chromosome (Dollé et al., 1989). Removing the function of particular paralogy groups severely impairs the development of the corresponding parts of the limbs, such as the forearms upon abrogation of group 11 genes (Davis et al., 1995). Accordingly, the complete absence of Hoxa and Hoxd gene function leads to heavily truncated limbs, containing only a proximal remnant of the humerus (Kmita et al., 2005).
The regulatory controls responsible for these expression patterns during limb development have been investigated in some details, at least for the HoxD cluster. During limb bud outgrowth and development, Hoxd genes are expressed in two waves. The first wave (early phase) concerns all genes of the cluster, but Hoxd12 and Hoxd13, and organizes an expression pattern that will subsequently correspond to a large proximal domain of the limb including part of the stylopodium and the zeugopodium. In forelimbs, this domain will generate the distal arm and forearm. A second wave of activation (late phase), starting about one day later, involves the expression of the four 5'-located genes Hoxd10 to Hoxd13 in the presumptive autopodium, the most distal part of the limbs. This late phase is required for the development of hands and feet (see Tarchini and Duboule, 2006).

Previous studies involving various chromosomal engineering strategies in vivo have led to two major conclusions regarding the modes of transcriptional regulation associated with these phases of expression. Firstly, regulatory sequences (enhancers) underlying both phases are located outside the HoxD cluster itself (Spitz et al., 2001). Secondly, the early and late phases are controlled by regulations that can be separated from one another, reflecting their different evolutionary origins.

While the early phase is controlled by an activation-inhibition mechanism relying upon sequences located at either sides of the cluster (Spitz et al., 2005; Tarchini and Duboule, 2006; Zakany et al., 2004), the expression in presumptive digits was shown to depend on enhancer sequences located centromeric (5'-) to the cluster (Spitz et al., 2003, 2005).

To identify this centromeric regulatory element, we developed a locus-targeted enhancer-trap approach, which allowed us to test the regulatory potential of several bacterial artificial chromosomes (BACs) mapping at the HoxD locus (Spitz et al., 2003). A transgene, in which a LacZ reporter gene driven by a minimal β-globin promoter had been inserted in an intron of Lunapark (Lnp) on BAC RP11-504O20 (Fig. 1B; 504βLac), showed reproducible expression in distal limbs, external genitalia, dorsal and ventral neural tube, hindbrain, midbrain, forebrain and in both the heart and the eyes (Spitz et al., 2003; Fig. 1C). The expression pattern in the limbs resembled that of endogenous posterior Hoxd genes at comparable stages, though with some differences. The 504O20 BAC also contained regulatory activities in several sites of the developing central nervous system, including some cell types or tissues that normally express both Lunapark and Evx2, the two transcription units localized between the HoxD cluster and this region.

![Fig. 1. The GCR and Prox sequences display distal limb expression potential in transgenic mice.](image-url)

- **(A)** Schematic representation of the human HOXD cluster and 5' surrounding sequences with the two major conserved regions within the GCR, represented in red (CsA) and yellow (CsB) boxes. LNP (purple), EVX2 (orange) and HOXD (dark yellow) exons are highlighted. The arrows above the genes reflect their respective transcriptional orientations.
- **(B)** Transgenes injected in this study: human BAC RP11-504O20 and RP11-514D19 were engineered to produce 504βLac, 514βLac and 514D11Lac. The GCR and the Prox regions were cloned and linked either to the βLacZ or to the mouse Hoxd11LacZ reporter genes, generating GCR-βLac, GCR-d11Lac and d11Lac-Prox. The grey boxes outline the GCR and the Prox regions. Dark blue boxes represent LacZ coding sequences; the light blue box is the human βglobin minimal promoter; blue and yellow boxes are the Hoxd11LacZ reporter gene.
- **(C)** E11.5 X-gal staining of transgenic embryos corresponding to the constructs depicted in panel B; fb, forebrain; mb, midbrain; abh, anterior hindbrain; dnt, dorsal neural tube; vnt, ventral neural tube; dl, distal limbs; dfb, dorsal forebrain; gb, genital bud. **(D)** Activity of the d11Lac-Prox transgene at different stages of development.
Furthermore, these two genes are expressed in the developing autopodium, much like the posterior Hoxd genes, even though they do not have an overtly important function during limb development. This led to the conclusion that this DNA fragment contains a Global Control Region (GCR), which is able to regulate the expression of several genes in a variety of developing tissues and cell types. These genes may not necessarily be related in their functions but, by sharing a common genomic location, they set out a “regulatory landscape”, as defined by the transcriptional enhancing activity(ies) of the GCR (Spitz et al., 2003).

Within the region covered by the BAC 504O20, human versus mouse sequence comparisons revealed a 40 kb large conserved region located 3′ of LNP and delimited by two blocks, ca. 5 kb long each, referred to as conserved sequences A and B (Fig. 1A; CsA and CsB) and showing more than 80 percent interspecies sequence identity, with large stretches above 95 percent. Deletion of this region in the BAC 504fLac (504fLacAGCR) resulted in the loss of LacZ expression in limbs, dorsal neural tube, midbrain and forebrain (Spitz et al., 2003), suggesting that the GCR was localized within this deleted fragment.

In this report, we use functional transgenic approaches to further characterize the control Hoxd gene expression in developing limbs. First, we confirm that the 40 kb evolutionary conserved region (thereafter referred to as the GCR) contains the limb enhancer activity ascribed to BAC 504O20 and to the GCR. However, this activity showed significant differences with the endogenous expression of Hoxd13. First, both the anterior extension and overall expression level in the distal autopod were variable from one line to another: some animals showed a broad expression covering the whole autopod, whereas others displayed a more restricted and weaker posterior dot. Furthermore, expression of the reporter gene was much weaker in the medial part of stage E11 limbs, in contrast to the rather homogenous expression of Hoxd13 at the same stage. We identified an additional limb enhancer in the GCR-514 region, 40 kb evolutionary conserved region (thereafter referred to as BAC 514d19). The other BAC transgenes were constructed by ET-recombineering (Lee et al., 2001), starting with BAC RPCI11-1006L9. Four homology arms (nestI, nestIII, T7end and SP6end), from 150 to 400 bp were amplified by PCR using the BAC DNA as a template and the following pairs of primers:

- nestI 5′-AGTGGTACCCGGCCGCCGCCCATATAAAGTTGGCCAAAGG-3′
- nestII 5′-AGTGGTACCCGGCCGCCCATATAAAGTTGGCCAAAGG-3′
- nestIII 5′-AGTGGTACCCGGCCGCCCATATAAAGTTGGCCAAAGG-3′
- nestIV 5′-AGTGGTACCCGGCCGCCCATATAAAGTTGGCCAAAGG-3′

PCR products were subsequently digested with PstI and either NorI, or Acc65I, ligated to each other (nestI-T7, nestII-NorIII-NorIII-Sp6) and cloned into a pBluescript II SK (+) backbone digested with Acc65I and NorI. A reporter gene construct, either βLacZ (Spitz et al., 2003) or HoxdIIlacE/E (Gérard et al., 1993) was then inserted, linked to a kanamycin resistance marker, into the PstI site of the resulting plasmids. Targeting constructs were released by NorI digestion, gel-purified and electroporated into EL250 bacteria containing the RPCI11-1006L9 BAC after induction of the recombinases, as described previously (Spitz et al., 2003). This strategy allowed for the concurrent insertion of the transgene and deletion of the sequences located in-between the homologous arms. Recombinant clones were selected on kanamycin (20 μg/ml) LB agar plates at 30 °C, amplified and verified by restriction enzyme fingerprinting. The GCR-fLac and GCR-d11Lac transgenes were obtained by using the nestI-T7 homology arms flanking the corresponding reporter genes, while a targeting cassette using HoxdIIlacKanE/E flanked by nestIII-Sp6 (respectively nestI-norIII) led to the recombined BAC d11Lac-Prox (GCR-d11Lac-Prox).

To engineer BAC GCR-514, we first PCR-amplified a 1.4 kb large fragment corresponding to the T7 end of BAC RPCI11-387A1, with primers nestT7a and 387cen (5′-ATGCCATGGGAGAGGTAGGTGTGGC-3′) and either 514cen (5′-AGTGGTACCGCGGCCGCTTAGTAATTAAGGGATC-3′) or 514cen (5′-AGTGGTACCGCGGCCGCTTAATTAAGGGATC-3′) and 1 kb of sequences from the 387cen end. We cloned a FRT-flanked kanamycin marker in the corresponding reporter genes, checked and digested with NotI and either BAC RPCI11-1006L9 BAC after induction of the recombinases. We isolated kanamycin-resistant colonies, amplified them and identified those that corresponded to the expected GCR-514 fusion product by restriction enzyme fingerprinting. We further verified the presence of intact CsC, CsB and CsA by PCR analysis.

To engineer the 504fLacAA and 504fLacAB clones, we electroporated a PCR amplified Zeocin resistance cassette (Invitrogen) flanked by 50 bp large DNA fragments homologous to both ends of the regions to be deleted, into EL350 competent cells containing BAC 504O20[504O20lac6, transiently induced to express the ϕC13-galactosidase recombination system (Lee et al., 2001). Briefly, for deletion of CsA, PCR was used to flank the Zeocin resistance gene with the sequences from BAC 504O20. 5′-ATA TTT GCA AGG TGA AGG GTG ATA CAA TTA TTC AGC TAG AGC ATT CTC AG-3′, and 5′-CCA GGT ACA CAC ATA TAA GAA CAT TAC TTA CCA AGC CTC AAC TCT TAT AAG GT-3′, respectively. For the deletion of CsB, 5′-TTT GTG GTG TGT TCA TCT TCT AAA CCT ACT GTT TTC CTC TCC CCC AGT AGT G-3′ and 5′-GTT GTG GTG TAT CCG GTG GTG TAG GAG GGA AGA AATG TGG GTG AAG GT-3′, respectively. BACs carrying the deletions were recovered by plating cells on Zeocin LB agar plates (25 μg/ml) at 30 °C. Colonies were

Materials and methods

BAC transgenes

BACs 504fLac and 514D11Lac were previously described (Spitz et al., 2003) and were referred to as 504o20fLac6 and 514d19-D11LacZ. 514fLac was obtained by in vitro random transposition of Tn5fLac-Z (Spitz et al., 2003). Mapping of the insertion sites was made by Southern blot analysis using a KanR probe to determine the number of Tn5fLac-Z integrations. Further restriction enzyme fingerprinting verified the integrity of the genomic insert. Using XhoI digestion and field inversion electrophoresis (FIEGE), we mapped the insertion in a 57 kb fragment located at the 5′ end of BAC 514d19. The other BAC transgenes were constructed by ET-recombineering (Lee et al., 2001), starting with BAC RPCI11-1006L9. Four homology arms (nestI, nestIII, T7end and SP6end), from 150 to 400 bp were amplified by PCR using the BAC DNA as a template and the following pairs of primers:
amplified and analyzed both by PCR and restriction fingerprinting to check for both the targeted deletion and the integrity of the BAC. BACs showing the expected deletions were linearized with PstI-ScaI and injected into C57BL/6×DBA F1 fertilized mouse oocytes as previously described (Spitz et al., 2001).

Plasmid transgenes

To produce CsA-{\beta}Lac, we digested BAC 504O20 with AvrII, and cloned the 11,545 bp large DNA fragment containing CsA in pBlueScript II SK(−), after digestion with XbaI to produce pSK11545. We cloned the βlacZeo cassette from pSK11545B, after digestion with Bsp120I-Host, in the NorI site of pSK11545 and obtained pSK11545Blac. Finally, pSK11545Blac was digested with Acc65I and NorI and the 5529 bp large Acc65I-NorI fragment containing CsA, and the 6992 bp large Acc65I-NorI fragment containing pBlueScript II SK(−) backbone and the βlacZeo cassette, were recovered and cloned together to obtain CsA-B{\beta}Lac. This construct was cut by Acc65I-NorI to eliminate backbone sequences, purified and injected.

To produce CsB-{\beta}Lac and CsB-{β}lac−(−), we digested BAC 504O20 with AvrII, and cloned the 15936 bp large DNA fragment containing CsB in pBlueScript II SK(−), digested by XbaI to produce pSK15936. pSK15936 was digested with Smal and Swal, and the 4661 bp large fragment containing CsB was cloned into pSK15936B opening with Bsp120I and treated with Klenow. Two orientations were recovered, corresponding to CsB-B{\beta}Lac and CsB-{\beta}lact−(−). For injection, the constructs were digested with Acc65I-NorI to eliminate backbone sequences and purified.

To produce CsC-{\beta}Lac, we digested BAC 514d19 with EcoRI, and cloned the 4739 bp large DNA fragment containing CsC in pBlueScript II SK(−), digested by EcoRI, to produce pSK4739. We cloned the βlacZeo cassette from pSK4739B with Bsp120I-Host, in pSK4739, after linearization with NorI. Both orientations were recovered. The one where the βglucat promoter was in the opposite orientation with respect to CsC was selected (CsC-{\beta}lac). For injection, CsC-{\beta}Lac was digested with Acc65I and Swal to eliminate backbone sequences and purified. The h{\beta}lac transgene was generated by inserting a 4.2 kb large KpnI fragment, isolated from pSK4739, into the unique KpnI site of pSkiII-B{\beta}lac. We selected a clone where the CsC fragment was oriented such like the βlac reporter gene had the same orientation, with respect to CsC, than that of the neighboring gene {\upsilon}up in the normal situation. The h{\psi}C{\beta}lac-(AC(NH)2)-{\beta}lac, respectively) transgene was released by digestion with SfiI and NorI (Nhel and NorI, respectively), purified and injected into mouse fertilized oocytes, as previously described (Spitz et al., 2001).

For the chicken CsB region, we screened a chick BAC library (Crozinnjans et al., 2000) spotted on filters (MRC-HMG1). We identified BACs containing the chicken GCR with probes for the chick CsA and CsB regions obtained after PCR amplification of chicken genomic DNA using degenerate primers. An EcoRINhel fragment containing the chicken CsB orthologous region was isolated from BAC 95P20 and cloned downstream of the βlac reporter gene, leading to the cB{\beta}lac transgene. To delete the tetrapod-specific regions B1 and B2, we PCR amplified a kanamycin resistance gene flanked by FRT sites using pairs of primer providing 50 bp large sequence homologies flanking each region (for B1: 5′-GGT CAT CTG CAT GAG GGG GTG GTC ATT GAT CAT CCT CTG CCT GGT GAC AGG TGA G-3′ and 5′-ACA TAT CGC AGA GTG TGG GGG GAA AAT CAA ATA TTG CTT CTT CCA TTA TTT-3′; for B2: 5′-GGT TTG TTG TTA AAC AGT TCT TGT TAG TCA TTT AAA CAT AGT GCA TCT TGT GTG-3′ and 5′-GGG CAT TCA TCT GTG TGG CAA TAT CAT CTG ATA CTT GTG TAA CCC TTT TTT-3′). The PCR products were co-electroporated with the cB{\beta}lac plasmid in EL250 cells used for recombination. Colonies that had integrated the KAN marker in B1 (B2, respectively) were selected on Kanamycin plates at 30 °C. Fresh EL250 cells were electroporated with plasmid DNA prepared from the resistant colonies, then plated on LB agar plates with kanamycin to avoid colonies containing a mixture of targeted and non-targeted plasmids. One colony containing the correctly targeted plasmid was then expanded and cultured with 0.1% L-arabinose for 1 h to express transiently the FLP recombinase, such as to delete the Kanamycin marker. Colonies were replicated on Ampicillin and Ampicillin + Kanamycin LB agar plates and Amp6 and Kan6 colonies were further analyzed.

The resulting plasmids contained the cB{\beta}lac transgene where B1 (B2, respectively) had been replaced by a 100 bp large linker (from the pGEMI polylinker) with FRT site, leading to the cBΔB1-{\beta}lac (cBΔB2-{\beta}lac, respectively) transgene.

Genotyping

504{\beta}lac, 514{\beta}lac, 504{\beta}lacΔA, 504{\beta}lacΔB and GCR-β{lac} transgenic animals were genotyped by PCR, using primers specific for the Trn7lacZ transposon (LacZ5′rev: 5′-TGG GTA ACG AGG GGG TCT TCC-3′ and Kan903-pre: 5′-GGT ATG AGT CAG CAA CAC CTT CTT C-3′), or by Southern blot using a HindIII digestion and a KanR probe, revealing transgene-specific bands at different sizes depending on the BAC and the insertion site of the Trn7lacZ transposon. 514D11lac transgenic animals were genotyped by Southern blot using a BamHI restriction digestion and a KanR probe, generating a 1264 bp large transgene-specific band. GCR-d11lac, d11lac-Prox and GCR-d11lac-Prox animals were routinely genotyped by PCR using primers in the LacZ and Kan markers (LacZ3′up: 5′-AAG CAC ATG GCT GAA TAT CGA CCG-3′ and Kansil3: 5′-CAG AAG CCC GTG CTT TTT CCA CGG TTG-3′) as well as checked with primer pairs spanning the BAC, as described previously (Spitz et al., 2003).

CsA-β{lac}, CsB-β{lac}, CsB-βlact−(−) and CsC-β{lac} transgenic animals were genotyped by PCR, using primers specific for the βlacZeo cassette (LacZeo5′fw: 5′-TCA ACA GCA ACT GAT GGA AAC-3′ or LacZeo3′fw: 5′-ACT AAA CCA TGG CCA AGT GGA-3′) or by Southern blot using a PvuII restriction digestion and a LacZ probe generating a 2557 bp large transgene-specific DNA fragment. Chicken transgenes were genotyped with chicken-specific primers for CsB (db499: 5′-GCAAGCCCTGCAATGCCCCGGGG-3′ and cB832r: 5′-CATTGAGGGCTTCTTCTTAGAT-3′).

Interaspers sequence comparison

For interspecies DNA sequence comparison, we retrieved genomic sequences 5′ of the human, mouse, chicken, Xenopus, fugu and tetraodon Hoxd loci from the Ensembl database (last update 03/06). Alignments were performed using the mVISTA program AVID (Bray et al., 2003; Mayor et al., 2000); window size 50 bp; homology threshold 65%.

LacZ staining and whole-mount in situ hybridization

Mouse embryos were collected and assayed for LacZ or Hoxd gene expression following established protocols and probes, as described previously (Spitz et al., 2003).

Pictures

For the transgenic fetuses shown in Figs. 1–4, the backgrounds of the pictures (outside the biological material) were artificially homogenized for the sake of comparison.

Results

GCR-dependent expression in the CNS and distal limb buds

The GCR was initially defined as a region of 40 kb of DNA sequences highly conserved between mouse and human, whose deletion abrogated expression of a reporter transgene in developing limbs and most of the neural tube and CNS, in the context of the 504{\beta}lac transgene (Spitz et al., 2003). In order to further characterize the regulatory function(s) of the GCR, we engineered the human BAC RPC111-1006L9 by using ET recombination, to produce a transgene consisting of the 40 kb large GCR fragment with a β{lac}Z reporter vector (Fig. 1B; GCR-β{lac}). Two F0 transgenic embryos were obtained, both with staining in distal limbs, dorsal neural tube, hindbrain,
midbrain and forebrain domains (Fig. 1C), i.e. in patterns complementary to those obtained upon deletion of the GCR from BAC 504 (504βLacΔGCR), thus validating the approach. However, staining in distal limbs was mainly observed in the ventral and dorsal mesenchyme, whereas expression was weaker in deeper mesenchyme (not shown). A similar trend had already been reported when using the 504βLac transgenes (Spitz et al., 2003), in contrast to the homogeneous expression of posterior Hoxd genes, normally detected throughout the distal mesenchyme, whenever assayed either by in situ hybridization or by LacZ staining of knock-in alleles (Kmita et al., 2000; Kondo et al., 1996; Van der Hoeven et al., 1996).

To test whether this discrepancy was due to a weak response of the βLacZ reporter gene to the GCR enhancers, we generated a transgene where the GCR was associated with a Hoxd11LacZ reporter gene (Fig. 1B; GCR-d11Lac). Hoxd11LacZ is not expressed in distal limbs, when inserted randomly in the genome, but becomes active in this region whenever inserted 5′ of the HoxD cluster, i.e. under the control of the GCR (Gérard et al., 1993; Van der Hoeven et al., 1996). We obtained five GCR-d11Lac transgenic embryos, three of them showing an X-gal staining pattern similar to GCR-βLac in limbs, dorsal neural tube, hindbrain, midbrain and forebrain (Fig. 1C). Expression was also detected in caudal trunk corresponding to the

Fig. 2. Combination of GCR and Prox. (A) Schematic representation of the human HOXD cluster and 5′ surrounding sequences with CsA, CsB (red and yellow boxes, respectively), LNP (purple), EVX2 (orange) and HOXD (dark yellow) exons are shown. The arrows above the genes reflect the respective transcriptional orientations. The GCR and Prox are underlined in dark grey. (B) The GCR-d11Lac-Prox transgene was obtained by a deletion of the LNP region in BAC RP11-1006L9 (1006) and replacement by the Hoxd11LacZ reporter gene. (C) The GCR-514 transgene was obtained by fusion of the 5′ part of 1006, containing the GCR, with the 3′ part of 514, containing Prox, EVX2, HOXD13 to HOXD8. (D) Left: a GCR-d11Lac-Prox transgenic embryo (line 8) stained with X-gal at E11.5. From left to right, lateral and transversal views of transgenic forelimb buds stained with X-gal at E11.5: GCR-d11Lac, d11Lac-Prox and GCR-d11Lac-Prox. Right: endogenous Hoxd13 transcript distribution (Mmd13) at the same stage. Red double arrows indicate the region of the limb bud internal mesenchyme gained in GCR-d11Lac-Prox transgenic animals (line13). (E) Activity of the GCR-d11Lac-Prox transgene (line 8) at different stages of development. (F) Left: Whole-mount in situ hybridization of wt (control) and GCR-514 transgenic embryos with a human-specific HOXD13 RNA probe (HsD13) with higher magnifications of their forelimb buds. Right panel: CT scans of wt, del(11–13) and del(11–13); GCR-514 adult hindlimbs. In del(11–13) hindlimbs, carpal bones are reduced, malformed and partially fused; phalanges are reduced in number and size (white arrow) except digit I, which carries an additional bony mass at its base (white arrowhead). In most cases, synpolydactily is also observed on posterior digits (red arrowhead). In del(11–13); GCR-514 hindlimbs, the number of phalanges is still abnormal but digits are less reduced in size, the additional bony mass on digit one is reduced or absent (white arrowhead) and no synpolydactily is observed.
autonomous activity of the Hoxd11LacZ transgene (Gérard et al., 1993). In the genital bud, the GCR-d11LacZ transgenic embryos showed a broader domain of expression, when compared to both the Hoxd11LacZ transgene alone, and the GCR-βLacZ transgenic embryos, suggesting that, in the case of the genital bud, the GCR was acting more efficiently on a Hox promoter than on a minimal heterologous promoter. However, in distal limbs of GCR-d11LacZ transgenic embryos, deep mesenchymal cells were mostly negative, as with the βLacZ reporter constructs, suggesting that the GCR by itself could not fully account for the expression throughout the dorso-ventral axis of the limb, as normally observed for Lnp, Evx2 and posterior Hoxd genes (Fig. 2D). Therefore, while many regulatory traits were recapitulated, in this assay, by the 40 kb large GCR region, this DNA fragment failed to faithfully reproduce all aspects of the distal limb pattern.

Prox, a novel region with distal-limb enhancer potential

We hypothesized that additional limb-specific regulatory sequences are present at the vicinity of the HoxD cluster, in addition to the GCR, to properly regulate expression in developing distal limbs. We had previously tested the regulatory potential of BAC RPCI11-514D19 by inserting a LacZ reporter gene within the HOXD11 transcription unit (Fig. 1B; 514D11Lac). Three transgenic embryos were obtained, none of them showing LacZ staining in distal limbs (Spitz et al., 2003; Fig. 1C). Also, no limb-specific expression was detected by in situ hybridization with a human HOXD13 probe, using five other 514D11Lac transgenic embryos (Spitz et al., 2003). Similarly, a murine BAC including the Lnp to Evx2 intergenic region did not elicit this regulation in transgenic mice (Spitz et al., 2001). However, in a separate set of experiments, we injected a modified RPCI11-514D19 BAC, carrying the βLacZ reporter gene inserted after the second exon of LNP (514βLac; Fig. 1B) and recovered one embryo with LacZ staining in both the distal limbs and the CNS (Fig. 1C), even though this BAC did not overlap with the GCR. The location of this potential regulatory activity was excluded from the region overlapping with BAC RPCI11-504O20, as this expression pattern was never observed in any 504βLacΔGCR transgenic limbs. In addition, previous analysis of the 20 kb large interval located immediately 3′ of Evx2 did not reveal any obvious limb enhancer element (relIII in Kondo and Duboule, 1999; unpublished results). Therefore, we tentatively mapped this potential distal limb element referred to as Prox (Fig. 1B) within a DNA segment upstream LNP.
To confirm this observation and refined the localization of this element, we linked a 35 kb large DNA fragment, corresponding to the Prox interval, to the Hoxd11LacZ reporter gene (Fig. 1B; d11Lac-Prox). This fragment was slightly smaller than the interval defined above, but the only blocks of sequences conserved between human and mouse in the missing part were exons of Lnp. In addition to the autonomous Hoxd11 expression in caudal trunk (Gérard et al., 1993), seven out of eight d11Lac-Prox F0 transgenic embryos showed LacZ expression in distal limbs and in the genital bud (Fig. 1C). Several of them also displayed expression in the anterior hindbrain and dorsal-most forebrain. In the limbs, the d11Lac-Prox embryos displayed an early inter-digital pattern, localized essentially to the ventral regions. At later stages of development, the staining concentrated around the condensing regions of the digits, whereas being somewhat excluded from both the digits themselves and the regressing inter-digital webbing (Fig. 1D). While the early inter-digital pattern was clearly distinct from the endogenous Hoxd13, Evx2 and Lnp expression patterns, the subsequent late staining was reminiscent of Hoxd13 expression around the perichondrium of condensing digits. These results suggest that Prox may contribute, in
addition to the previously identified GCR, to set up the expression pattern of Lnp, Evi2 and posterior HoxD genes in developing distal limbs.

Combination of GCR and Prox together re-enforces limb expression

Since neither GCR nor Prox was sufficient, in isolation, to recapitulate all aspects of the endogenous distal limb expression in our transgenic assay, we combined both sequences on the same transgene. We engineered a GCR-d11Lac-Prox construct by deleting the LNP region in BAC RP11-1006L9 (1006) and replacing it by the Hoxd11LacZ reporter gene (Fig. 2B). We obtained two transgenic stocks, which displayed similar expression patterns. Embryos of these two lines showed LacZ expression in distal limbs that was clearly improved in this transgenic configuration, when compared to either the GCR-d11Lac, or the d11Lac-Prox (Fig. 2D). Expression started in the posterior-proximal part of the limb, as early as day 10.5 (i.e. slightly earlier than with either the GCR, or the Prox element alone), to further extend distally and anteriorly (Fig. 2E), as described for the endogenous Hoxd13 transcript domain (Fig. 2D; Mmd13), including a uniformly stained mesenchyme. In the first line (Fig. 2E), the expression comprised the whole autopod (from presumptive digit V to I) at stage E11.5, much like the endogenous Hoxd13 gene. In the other line (Fig. 2D), at the same stage, the signal was more restricted posteriorly (excluding the presumptive digit I), whereas at later stages, LacZ expression was also scored in digit I, as was the case for the 504flLacZ transgene.

In order to test if the combination of these two DNA fragments indeed generated a genuine posterior Hoxd domain in the developing digits, we prepared an expression construct to functionally rescue loss of function alleles. We constructed a transgenic locus by fusing the 5′ end of BAC RPCI11-1006L19, containing the GCR, with the 3′ part of BAC 514D19, containing Prox, EVX2 and from Hoxd13 to Hoxd8 (Fig. 2C; GCR-514). This 178 kb large construct thus contained both the GCR and the Prox element directly upstream the EVX2-5′ HoxD interval. We produced five distinct transgenic lines (GCR-514) and tested the expression of the human Hoxd13 gene by using a riboprobe (HsD13) that did not cross-hybridize with the murine orthologous transcript. Three out of five lines expressed Hoxd13 in distal limbs (Fig. 2F), whereas signal was detected neither in non-transgenic animals, nor in the limbs of transgenic embryos for a 514D19 BAC including the Prox region but lacking the GCR (Spitz et al., 2003). However, the expression pattern did not extend as distal as for endogenous Hoxd13, suggesting that either another element was still required, or that the respective configuration of these elements was not optimal. It should be noted that the expression of reporter genes introduced within the endogenous Lnp-HoxD locus is not always as strong or as distal as the broad Hoxd13 pattern (e.g. Hoxd9LacZ in Kondo and Duboule, 1999; Hoxd12 in Kondo et al., 1996). This relative variability implies that the architecture of the locus might also play an important role in finely tuning the expression of the genes lying within.

Partial rescue of a Hoxd gene deletion by a GCR and Prox-driven mini cluster

We assessed the ability of the GCR and Prox sequences to fully control Hoxd gene expression in developing digits by means of a functional rescue. We crossed transgenic animals from one of the GCR-514 lines with Del(11–13) mice, i.e. mice carrying a targeted deletion of both Hoxd13 and Hoxd12, plus an insertion of LacZ reporter sequences within Hoxd11 (Zakany and Duboule, 1996). These latter mice have a triple loss-of-function in cis of Hoxd13, Hoxd12 and Hoxd11. Mice homozygous for Del(11–13) show small digit primordia, a disorganized cartilage pattern and an impaired skeletal mass (Zakany and Duboule, 1996). More specifically, Del(11–13) homozygous mutant hindlimbs display reduction, malformation and partial fusion of carpal bones, a drastic reduction in both the number and size of phalanges in digits II to V as well as synpolydactyly, i.e. the proximal fusion of digits. Moreover, a supernumerary bony mass was systematically observed on digit one, clearly associated with the loss-of-function of Hoxd13 (Fig. 2F; Del(11–13); Dölle et al., 1993).

The presence of the GCR-514 transgene in Del(11–13) mice led to a substantial reduction in the severity of the mutant phenotype, yet the affected structures were only partially corrected. The number of phalanges was still abnormal, but the size of the digits was much less reduced than in Del(11–13) mutant mice. Also the ectopic bony mass usually found on mutant digit one was drastically reduced, if not absent, and the synpolydactyly disappeared (Fig. 2F). In marked contrast, such a phenotypic correction was not observed with a human HOXD PAC transgene lacking the Prox and the GCR sequences (Spitz et al., 2001). While the rescuing effect of the GCR-514 transgene demonstrated that the combined presence of both the GCR and Prox element was able to drive expression in the correct domains, the partial aspect of this rescue may be explained as discussed above either by the intrinsic nature of the transgene (i.e. the absence of an additional limb enhancer or the artificial organization of the various elements in the transgenic DNA), or by a low expression level due to position-effect. It is also possible that the phenotype induced by the Del(11–13) allele partly derives from the earlier expression of e.g. Hoxd11, i.e. from a regulation that does not depend upon the GCR-Prox influence but instead, from the 3′ located early limb control region (ELCR; Tarchini and Duboule, 2006; Zakany et al., 2004).

Altogether, these results indicated that the GCR and the Prox elements are able to direct posterior Hoxd gene expression in distal limbs in vivo. Interestingly, the Prox element seems unable to exert its regulatory potential in the absence of the GCR, as shown for example with the 514D11Lac transgene, or in the Ulnaless mutation (Spitz et al., 2003; Herault et al., 1997). The addition of the GCR helps to overcome this inability, leading not only to an additive effect in terms of patterns, as shown with the GCR-d11-Prox transgene (Fig. 2D), but also to a broader range of action (the GCR-514 transgene; Fig. 2F).
Transgenic analyses of the GCR and Prox regions

In order to identify DNA fragments responsible for distal limb expression within either the GCR, or the Prox region, we looked at interspecies sequence comparisons spanning the 300 kb large DNA interval 5’ of the HOXD cluster and covered by the two BAC clones 504 and 514 (Fig. 3A). Within the GCR, the regions conserved among tetrapods (mammals, chicken and *Xenopus*) are the previously reported regions A and B (CsA, CsB; Spitz et al., 2003; Fig. 3). The specific regulatory potentials of CsA and CsB were analyzed using two different approaches. First, we engineered versions of the 504βLac construct deleted for either one of these sequences, independently (Fig. 3C; 504βLacΔA and 504βLacΔB), such as to test the effect of these deletions in a genomic environment partially reconstructed by the surrounding BAC sequences. Secondly, these regions were separately associated to the βLacZ reporter gene (Fig. 3C; CsA-βLac and CsB-βLac) such as to test their autonomous activity.

Out of the five transgenic lines produced with the 504βLacΔB construct, three showed lacZ staining. Positive cells were detected in the ventral neural tube, but all three lines were negative either in distal limbs, dorsal neural tube, midbrain or in the forebrain (Fig. 3D). This pattern was quite similar to that of the 504βLacGCR construct, further suggesting that the GCR-dependent enhancer activities require the 5 kb large conserved region B. When CsB was tested alone, upstream of the reporter gene, five out of the eight independent insertions expressed lacZ in a broad limb domain extending from distal to prospective zeugopod, as well as in the forebrain, midbrain, hindbrain and the dorsal neural tube. As for GCR-βLac transgenic embryos, X-gal staining in limbs remained quite superficial, with deep internal mesenchyme mostly negative. It did not extend as distally as with the GCR-βLac transgene, hence it appeared similar to the case of the GCR-514 transgene. However, these activities correlated well with those domains lost upon deletion of either the GCR or CsB. A similar expression pattern was observed when the CsB element was cloned in the reverse orientation (Fig. 4B; CsB-βLac(−)), showing that CsB acts in an orientation-independent manner.

As previously reported (Spitz et al., 2003; Figs. 1C, 3C), we could not assign any enhancer function to CsA. However, when we directly assayed CsA for its enhancer activity, three out of five CsA-βLac transgenic lines showed expression of the βLac reporter gene in the ventral neural tube and in a restricted domain of the anterior hindbrain. This set of results showed that CsB endorses most the enhancer function assigned previously to the GCR, including the limb expression, whereas CsA displays a ventral neural tube specificity, which was not revealed by previous BAC deletions. The differences in the expression patterns of CsB-βLac and 504βLac may point to the presence of non-autonomously acting limb elements located somewhere within the GCR, required to refine CsB activity. Alternatively, yet not exclusively, this may also reflect technological contingencies. In contrast to BAC clones, smaller transgenes often integrate in tandem arrays, leading to closely juxtaposed copies of the enhancer.

Sequence comparison highlighted stretches of evolutionary conserved sequences within the Prox region (conserved sequence C; CsC). The CsC element is located between *Lnp* and the *HoxD* cluster, in all tetrapod species analyzed in this work (mammals, birds and amphibians). In contrast, no significant homology to CsC was found within the *Lnp-HoxD* interval, either in the pufferfishes (Tetraodon and Fugu) or in the zebrafish genomes. We isolated a 4 kb large DNA fragment spanning CsC and cloned it downstream of the βlac reporter gene. Six transgenic lines were produced for this construct, among which three showed similar expression patterns in distal limbs and in the genital bud (Fig. 3D; CsC-βLac), mimicking the expression profiles observed when the d11lac-Prox transgene was used. In two lines, LacZ expression in limbs was restricted posteriorly, expanding distally and anteriorly in the prospective digit domain, but mostly in the inter-digital and ventral mesenchyme. In the other line, the expression domain in limbs was broader and comprised the whole distal mesenchyme.

Besides these elements, we assayed all other evolutionary conserved regions in this DNA interval for their potential to activate the βLac reporter gene in E11–E12 mouse transgenic embryos. While some of them showed enhancer activities, predominantly in distinct regions of the neural tube and of the brain, none of them was active in developing limbs (not shown). Therefore, we conclude that the GCR and Prox regions are likely the main sequences, if not the only ones, required to drive *Hoxd* gene expression in distal limbs.

Evolutionary dissection of CsB and CsC

A comparison between the CsB and CsC regions of various species revealed blocks of conservation. For CsB, three regions were distinctly conserved throughout vertebrates, whereas two regions (B1 and B2) were selectively found in tetrapods (Fig. 4A). The high conservation of B1 and B2 among tetrapods, their absence from fish genomes and the observation that the pufferfish GCR can control reporter gene expression in the same neural domains than the human GCR, though not in the limb (Spitz et al., 2003), made us focussing subsequent functional analysis on these tetrapod-specific regions.

We first cloned a 4.7 kb large DNA fragment containing the chicken CsB region and linked it to the βLac reporter gene. Four out of the six transgenic embryos produced showed expression in neural tissues and in the early limb bud, similar to what was observed with the human orthologous element. However, in these embryos, expression in the more proximal limb domain detected with the human CsB fragment (but not with the whole human GCR) was reduced, if not absent (Fig. 4B; cB-βLac). The avian CsB element could thus properly function in the mouse embryo and mostly recapitulated the limb expression pattern associated with the GCR. To test the role of the tetrapod-specific conserved regions B1 and B2, these regions were deleted from the chicken cB-βLac transgene. In both cases, we obtained embryos displaying limb patterns indistinguishable from that scored with the entire CsB element (3 out of 4 for cBΔB1-βLac, 2 out of 5 for cBΔB2-βLac). Therefore, these tetrapod-specific elements are likely dispensable for the limb expression pattern.
domain mediated by CsB. Interestingly however, both deletions led to the disappearance of the CsB-specific forebrain domain (Fig. 4B). In this context, it is noteworthy that the enhancer activity of the pufferfish GCR, when compared to tetrapod GCRs, was not only different in the limb bud, but also in the forebrain where it triggered expression in a broader domain, as compared to the dorsally restricted pattern observed with either the human or the chicken GCR/CsB (Spitz et al., 2003).

The CsC element, within Prox, is composed of two distinct evolutionary conserved regions. Region C1 is found in mammals, birds and amphibians. Region C2 is present in amniotes but not on the current Xenopus assembly (JGI4.1, August 2005), which contains a gap in this region, albeit 10 kb away from C1 (Fig. 4C). We assessed the contributions of these two regions by generating new $\beta$Lac transgenes, containing either C1 and C2 ($hC (Sp)$-$\beta$Lac), or C1 only ($hC (Nh)$-$\beta$Lac). With slight variations, probably due to transgene insertion sites, expression of $hC (Sp)$-$\beta$Lac was similar to that of Csc-$\beta$Lac transgenes (Figs. 3D, 4D), showing that most of the CsC regulatory potential was carried by this shorter fragment, which functions in an orientation-independent manner. Interestingly, this $hC (Sp)$-$\beta$LacZ transgene, containing both the C1 and C2 sequences, was highly penetrant in the limbs (five embryos out of five), like the larger Prox element. Few differences were scored between the patterns derived from either the 35 kb large Prox region, or the smaller CsC construct, possibly due to the use of different reporter systems ($\beta$LacZ versus Hoxd11LacZ).

For example, the dorsal forebrain domain observed in approximately half of the d11-Prox embryos was never scored in $hC (Sp)$-$\beta$LacZ or Csc-$\beta$LacZ embryos. In contrast, three out of five $hC (Sp)$-$\beta$LacZ embryos and two out of three CsC-$\beta$LacZ embryos showed LacZ expression in the developing face (Figs. 3D, 4D), unlike the large Prox transgene.

Five transgenic embryos carrying the ($Nh$)-$\beta$Lac transgene were produced, containing the C1 region in the absence of C2. Four out of five did not show any distal limb expression, whereas the remaining embryo showed some patches of weakly stained cells in distal forelimb buds (Fig. 4D; right panel). This expression could not be convincingly compared with the CsC pattern. However, three transgenic embryos showed a clear staining at the posterior margin of the proximal limb bud, indicating that the C1 element might nonetheless provide some regulatory activity in limb buds, yet outside its most distal part. Altogether, these results suggested that the C2 region, despite its apparent sequence conservation restricted to amniotes, is crucial for the CsC/Prox-mediated distal limb expression. By contrast, the C1 element, present also in amphibians, did not appear to confer distal limb expression on its own.

Discussion

Patterning of the distal tetrapod limb is controlled in part by the activities of posterior Hoxd genes, that contribute to both the growth of the future handplate and the number and identities of digits. Several experiments have shown that their coordinated expression is achieved by sharing a common regulatory mechanism, imposed by sequences localized upstream (5' of) the HoxD cluster (Spitz et al., 2005; Van der Hoeven et al., 1996). A transgenic screen and the resolution of the Ulnaless mutation have revealed the important function endorsed by a large evolutionary conserved element identified as the GCR (Spitz et al., 2001, 2005). The GCR contains distinct global enhancer elements, which are active both in distal limbs and in neural tissues (Spitz et al., 2003) and a chromosomal rearrangement that separates the GCR from the HoxD cluster demonstrated the functional importance of this sequence, not only to control posterior Hoxd gene transcription during limb morphogenesis, but also for the expression of the neighboring Evx2 and Lnp genes in the developing CNS (Spitz et al., 2003). However, whenever the 40 kb large GCR sequence was associated with a reporter gene in a transgenic assay, the observed patterns did not fully recapitulate those of endogenous Hoxd genes; expression was indeed not homogeneously distributed in distal limb mesenchyme, showing a preferential activity in cells close to the dorsal or ventral sides, with an important reduction in the medial compartment. Also the expression onset, in posterior limb buds, and progressive extension in anterior domains were slightly delayed, when compared to endogenous Hoxd genes.

Two enhancers for the late phase of expression in digits

These differences could be explained by the lack, in the transgenic context, of additional cis-acting elements, such as for example an enhancer-tethering sequence linked to Hoxd13 (Beckers and Duboule, 1998; Kmita et al., 2002a), or another remote enhancer element. As the split of the cluster into two independent pieces had no major impact upon the expression of Hoxd genes in distal limb (Spitz et al., 2005), it was unlikely that telomeric-located enhancer sequences, including the ELCR (Tarchini and Duboule, 2006; Zakany et al., 2004), would be involved in this late phase of expression. Here, we identified a novel element Prox/CsC, localized within the Lnp-Evx2 intergenic region, 35 kb from Evx2 and 40 kb from Lnp, in the mouse genome. When linked to a reporter transgene, this enhancer, was independently, from the GCR, to direct expression in distal limbs. In this work, we show that the presence of both Prox and the GCR sequences significantly improved the limb expression of the associated reporter genes, in particular for expression in deep distal mesenchyme. Interestingly, the GCR and Prox elements displayed distinctly restricted patterns, whenever their activities were tested separately. Therefore, the GCR and Prox elements likely involve different transcription factors and/or signaling pathways, each sequence contributing to part of the required regulation. Although several mouse stocks have been described carrying mutations in various genes involved in limb morphogenesis, posterior Hoxd genes were remarkably resistant to these perturbations, being generally expressed correctly as long as a distal limb element was formed (e.g. Litingtung et al., 2002; Michos et al., 2004; Yamaguchi et al., 1999). On the other hand, gain-of-function mutations of the same genes or pathways are generally able to turn on Hoxd genes ectopically (e.g. Duprez et al., 1996), providing support for a regulatory mechanism that
relies on the integration of distinct but redundant pathways, at least for this aspect. The observation that *Hoxd* genes are controlled by different, but complementary limb elements harboring distinct specificities, re-enforces the suggestion that they may integrate overlapping inputs from different signaling pathways. In physiological conditions, these elements may both serve a single regulatory function, by acting together as part of the same mechanism. Alternatively, they may contribute differently to *Hoxd* gene expression during limb development. For instance, the GCR might regulate a broad distal domain, required for the growth of the autopod, whereas the Prox element would preferentially set up the position and/or differentiation of digits by reinforcing *Hoxd* genes expression in and around the condensing digits. Such a multi-modal control of *Hoxd* genes during limb bud development should in principle provide some robustness to their expression patterns. The further identification of the core enhancer sequences within CsC and CsB should help identify those relevant transcription factors and discriminate between these possibilities.

**Cooperation for long distance activation**

The complementarity between the Prox and the GCR elements is not limited to tissue-specificity, and some other important functional properties appeared to be distinctly achieved by the two sequences. Remarkably, in the absence of the GCR, the Prox element was unable to activate *Hoxd* genes in a normal context. This was particularly obvious on a BAC transgene, or when the *Ulnaless* chromosome was studied. In this latter case, a large inversion separated *Hoxd* genes from the GCR, leading to an almost complete disappearance of *Hoxd* expression in distal limb buds, despite the presence of the Prox element at its normal position with respect to *HoxD* (Spitz et al., 2003; Herault et al., 1997). This inability of Prox to act alone at such a distance may be intrinsic, CsC being a short range enhancer, or instead, being limited by the presence of boundary elements that may prevent this enhancer to autonomously interact with *Hoxd* genes, as observed previously for some CNS enhancers (Kmita et al., 2002b).

This limited range of action of CsC was nevertheless relieved by the addition of the GCR, showing that this latter element plays a central role in the activation of the locus, not only by acting itself as a global long-range enhancer, but also by integrating the output of other enhancers located at its vicinity and extending their spectrum of activity. Recent data indicate that cis-acting elements can contribute to the organization of the spatial architecture of different loci, such as the β-globin, or the Interleukin gene clusters by promoting, through their associated trans-acting factors, the formation of large loops that bring distantly located DNA sequences in close vicinity (Cai et al., 2006; Patrinos et al., 2004; Spilianakis et al., 2005; Splinter et al., 2006). It is still unclear whether or not such interactions take place between the GCR, CsC and the neighboring genes (*Lnp*, *Evx2*, *Hoxd*), nor what is the role of the GCR in this respect, but our functional transgenic analyses suggest that the GCR might play an important role in building the regulatory architecture of the *HoxD* locus.

The functional dissection of the GCR by transgenic approach showed that the CsB element seemed to contain critical sequences required for the limb-specific activity of the GCR. Yet, generally speaking and despite a somewhat stronger signal seen with the CsB element, this distal digit pattern was better recapitulated by the entire GCR, as if CsB was lacking one particular component. This may illustrate that, even though CsB and the GCR may share the distal limb specific regulation, CsB might lack some “coordinating” activity located somewhere else in the GCR that could also be involved in extending the intrinsic short-range action of CsC. Remote cis-acting elements may act as genuine enhancers, and thus work at shorter distances, for example in a classical transgenic assay. However, at their endogenous position, in a different context, one would expect these elements to work equally well over long distances and in interaction with other elements, as the synergy and cooperation between various dispersed sequences are likely key factors in the global regulatory outcome. Classical transgenic approaches, while providing positive and valuable information about the intrinsic, short-range enhancer activity of one particular element (e.g. Gould et al., 1998; Vogels et al., 1993), are perhaps less adapted to those cases where long-range regulations are at work. Activities that would bridge distant elements and coordinate their action in a coherent output potentially different from the mere additive effect of individual elements remain difficult to analyze with currently available transgenic tools (see Kleinjan et al., 2006).

**Evolution of the regulations; “regulatory priming”**

We previously reported that the overall organization of the GCR-*Lnp-Evx2-HoxD* region had been conserved from teleost fishes to mammals. Evolutionary modifications leading to the recruitment of *Hoxd* genes to build the autopods of extant tetrapods, thus likely occurred through reshuffling of local elements rather than via major rearrangements of the locus (Shubin et al., 1997; Sordino et al., 1995). If alterations of *Hox* gene function through changes in cis-regulatory elements have been proposed to underlie morphological variations along the vertebral axis (e.g. Shashikant et al., 1998), the compact organization of the vertebrate *Hox* clusters has also been seen as a restriction to their evolvability for other purposes (Wagner et al., 2003). Our observations that *HoxD* distal limb expression is controlled by elements localized far outside of the complex could correspond to the solution used to accommodate this new function while preserving ancestral ones along the main body axis (see also Spitz et al., 2001). Several lines of evidence suggest that the existence of the GCR preceded the duplication of *Hox* clusters, i.e. before distal limbs emerged, and that a part of this sequence subsequently evolved along with the occurrence of specialized tetrapod appendages. For example, the orthologous GCR sequence from the puffer fish genome showed a reproducible CNS specificity when introduced into mice, yet expression in limb buds was not scored (Spitz et al., 2003).
Another observation arguing in favor of an ancient function of the GCR in the CNS, that was subsequently the ground for evolving a limb functionality, is provided by the analysis of the HoxA cluster, where a remnant of the GCR/CsB sequence was found upstream of Evx1 (Lehoczky et al., 2004), indicating that HoxA and HoxD probably derive from the same duplication event. Transgenic analysis of this sequence showed some activity in the CNS, but failed to uncover a clearly reproducible enhancer activity in developing distal limb, even though the closely linked Hoxa13, Evx1, Tax and Hibadh genes share some expression specificity there (Lehoczky et al., 2004). It suggests that these distal limb regulations, quite distinct in space between the HoxA (Hoxa13, Evx1, Tax, Hibadh) and HoxD (Hoxd13 to Hoxd10, Evx2 and Lwp) limb regulatory landscapes, may have evolved subsequently, either independently, or triggered by the presence of this ancestral GCR. In this last scenario, the intrinsic properties (e.g. accessibility to specific transcription factors; interaction with pleiotropic signaling pathways) of an already functional enhancer sequence may have facilitated a convergent evolution of regulations, after genomic duplications. We propose to refer to this process as “regulatory priming”, for such a facilitated regulatory co-option may have further stabilized, in turn, the system by adding yet another constraint. This may be important to consider when comparing paralogous enhancers and their regulatory properties, as some of their shared specificities may illustrate this convergent process, rather than the conservation of ancestral functions, after (full or partial) genomic duplications.

In addition to the GCR, the evolution of a global regulation in the HoxD cluster, to accompany the emergence of tetrapod digits, involved the subsequent addition of the Prox region. This sequence indeed likely appeared more recently than the GCR, since no related sequences are found in the available actinopterygians sequences. In addition, even its core enhancer region (CsC) is only partially conserved in amphi- bians and Prox-like sequences are not found in the HoxA cluster. From these considerations, we infer a model whereby a limb enhancer would have first evolved inside the GCR, due to the intrinsic capacity of this large DNA fragment to impose long-distance global regulation. As a result of this first step, a broad distal expression domain may have accompanied limb development in primitive tetrapods or sarcopterygians. Subsequently, the CsC/Prox element evolved due to its potential to both reinforce limb expression and refine it, by concentrating Hoxd gene expression around those regions developing into fingers and toes.

The alternative scenario whereby the Prox sequence appeared first and was subsequently re-enforced by the “extender” function of the GCR cannot be completely excluded, even though we consider it unlikely for those reasons mentioned above. Furthermore, our functional dissection of the Prox DNA fragment, based upon regions that showed strong conservation in tetrapods only, indicated that the C2 region of the CsC element was likely the element required for the limb specificity (this work and FS, unpublished), even though this latter sequence is apparently conserved only in amniotes. The CsC region found in Xenopus tropicalis is quite different, with only a C1-related element, and no C2 (according to the current assembly, which comprises no gap 10 kb around C1), despite the fact that posterior Hox genes are expressed during amphibian limb development (Christen et al., 2003). This suggests either that frogs and amniotes use different elements to drive similar expression domains, related to the case of c-er in fishes and mammals (Fisher et al., 2006), or that an enhancer function can be kept within the same DNA region without any obvious sequence conservation, as detected by current analytical tools. It is also possible that the very different strategies to develop limbs in amniotes and amphibians, due to the occurrence of metamorphosis, also marked by differences in Hox genes limb expression patterns (Christen et al., 2003; Satoh et al., 2006; Torok et al., 1998), imposed different requirements in terms of regulatory controls, thus making the Prox region obsolete in amphibians.

Two highly conserved DNA segments, specific for tetrapods, were identified within the CsB element of the GCR (B1 and B2). Sequence analysis of these regions revealed the presence of conserved consensus binding sites for different transcription factors, notably Gli, Smad and Left/Tcf, especially in the B2 segment, suggesting important functions associated with the Shh, BMP or Wnt signaling pathways. However, none of these motifs appeared critical for proper limb expression mediated by the CsB element, though these regions were functionally important in the forebrain. Discrepancies between an apparent highly constraint conservation and a tolerance to experimentally induced mutations were reported also for a Dach1 gene enhancer element (Poulin et al., 2005). Such unexpected results both highlight the difficulty to translate a mere evolutionary sequence conservation into a precise functional role, as often postulated, and underscore the need for extensive functional studies of enhancers in vivo.

Acknowledgments

We thank Carole Herkenne for her help to generate some of transgenic animals used in this study. This work was supported by funds from the canton de Genève, the Louis-Jeantet and Claraz foundations, the Swiss National Research Fund, the National Center for Competence in Research (NCCR) “Frontiers in Genetics” and the EU programme “Cells into Organs” to D.D., as well as by an HFSP Postdoctoral Fellowship to F.S.

References


Patrinos, G.P., et al., 2004. Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. Genes Dev. 18, 1495–1509.


