Human Molecular Genetics, 2003, Vol. 12, No. 14 **1725–1735**
DOI: 10.1093/hmg/ddg180

A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly

Laura A. Lettice[1], Simon J.H. Heaney[1], Lorna A. Purdie[1], Li Li[2], Philippe de Beer[3], Ben A. Oostra[2], Debbie Goode[4], Greg Elgar[4], Robert E. Hill[1,*] and Esther de Graaff[2]

[1]MRC-Human Genetics Unit, Western General Hospital, Crewe Rd, Edinburgh EH4 2XU, UK, [2]Department of Clinical Genetics, Erasmus MC, PO Box 1738, 3000 DR Rotterdam, The Netherlands, [3]Centre for Human Genetics, University Hospital Gasthuisberg, University of Leuven, Leuven, Belgium and [4]Research Division, MRC-HGMP Resource Centre, Hinxton, Cambridge CB10 1SB, UK

Received March 27, 2003; Revised May 5, 2003; Accepted May 14, 2003

Unequivocal identification of the full composition of a gene is made difficult by the cryptic nature of regulatory elements. Regulatory elements are notoriously difficult to locate and may reside at considerable distances from the transcription units on which they operate and, moreover, may be incorporated into the structure of neighbouring genes. The importance of regulatory mutations as the basis of human abnormalities remains obscure. Here, we show that the chromosome 7q36 associated preaxial polydactyly, a frequently observed congenital limb malformation, results from point mutations in a *Shh* regulatory element. *Shh*, normally expressed in the ZPA posteriorly in the limb bud, is expressed in an additional ectopic site at the anterior margin in mouse models of PPD. Our investigations into the basis of the ectopic *Shh* expression identified the enhancer element that drives normal *Shh* expression in the ZPA. The regulator, designated ZRS, lies within intron 5 of the *Lmbr1* gene 1 Mb from the target gene *Shh.* The ZRS drives the early spatio-temporal expression pattern in the limb of tetrapods. Despite the morphological differences between limbs and fins, an equivalent regulatory element is found in fish. The ZRS contains point mutations that segregate with polydactyly in four unrelated families with PPD and in the *Hx* mouse mutant. Thus point mutations residing in long-range regulatory elements are capable of causing congenital abnormalities, and possess the capacity to modify gene activity such that a novel gamut of abnormalities is detected.

INTRODUCTION

The zone of polarising activity (ZPA) resides at the posterior margin of the developing tetrapod limb bud and is required for the anterio-posterior pattern of digits. The role of the ZPA and the significance of the asymmetric posterior location were demonstrated in chick explants. Manipulation of embryonic limb buds placing an additional ZPA source at the anterior margin results in supernumerary anterior-localized (preaxial) digits (1). Production of SHH (Sonic hedgehog) appears to be the ultimate role of the ZPA. Expression of *Shh* defines the limits of the ZPA in the posterior margin (2,3) and replacement of the ZPA with a source of SHH in explant experiments results in a normal A/P pattern. In addition SHH is sufficient to induce supernumerary digits when ectopically expressed in the anterior mesenchyme (2).

Corresponding to the chick studies, preaxial polydactyly in mouse mutants is attributable to ectopic anterior expression of *Shh* at early limb bud stages (4–6). One such mutant, Sasquatch (*Ssq*), arose during the production of transgenic mice and displays a preaxial polydactylous phenotype with no other associated anomalities (7). The transgenic insertion, responsible for the phenotype, carries the human placental alkaline phosphatase (HPAP) reporter gene driven by the *Hoxb1* rhombomere 4 enhancer (7). In the *Ssq* mouse, in addition to HPAP expression in hindbrain rhombomere 4, the transgene acquired expression in the ZPA of the limb bud. HPAP activity is also detected in the anterior limb margin overlapping the

*To whom correspondence should be addressed. Tel: +44 1314678410; Fax: +44 1314678456; Email: bob.hill@hgu.mrc.ac.uk

ectopic *Shh* expression responsible for generating the polydactyly phenotype.

Ssq maps to mouse chromosome 5 and corresponds to the preaxial polydactyly (PPD) locus on human chromosome 7q36 (8). PPD in human is one of the most frequently observed congenital hand malformations and patients present with a range of limb phenotypes from triphalangeal thumb to one, two or sometimes three extra digits on hands and/or feet (9–13). Analysis of both a *de novo* chromosomal translocation breakpoint in a patient with PPD and the site of transgenic insertion in *Ssq* identified lesions within intron 5 of the *Lmbr1* gene (8). *Lmbr1* is physically linked to *Shh* and intron 5 lies at a distance of 1 Mb (a genetic distance of 1.8 cM) upstream from *Shh*. Genetic analysis confirmed that, despite this distance, a *cis*-acting regulator is responsible for the preaxial polydactyly phenotype in *Ssq* and therefore PPD (8).

In our attempt to determine the molecular basis for preaxial polydactyly, we identified the regulatory element that is responsible for the initiation and spatially specific expression of *Shh* in the ZPA. As predicted by the genetic analysis this regulatory element is located 1 Mb upstream of *Shh*, embedded in a gene that resides in a cluster of unrelated genes. Point mutations were found that reside in the ZRS in both human and mouse. These mutations modify ZRS activity resulting in ectopic expression of *Shh* and, consequently, supernumerary preaxial digits.

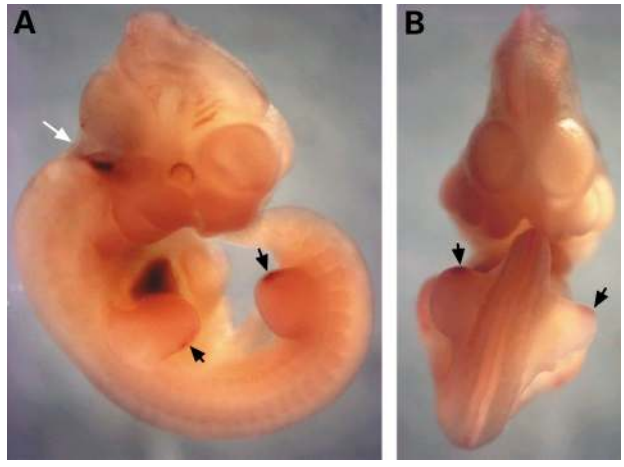

Figure 1. E10.5 transgenic embryo injected with Cos33S containing the *Ssq* insertion site and stained for HPAP expression. (**A**) Lateral view showing staining in rhombomere 4 (white arrow) and within the ZPA of the limb buds. (**B**) Dorsal side of the hindlimbs showing specific expression at posterior margin (black arrows).

RESULTS

Regulatory elements reside near the *Ssq* insertion site

The transgenic insertion site for the *Ssq* mutant mouse was isolated previously and was postulated to be near or within a regulatory element that affects expression of *Shh* (7,8). To investigate the proximity of a potential regulator, transgenic mice were made using cosmid clones from a *Ssq* genomic library containing the insertion site. The cosmid Cos33S contains at least two complete copies of the transgene construct (carrying the HPAP reporter gene) and 8.5 kb of flanking genomic DNA. G_0 embryos analysed at E10.5 showed expression of HPAP at the posterior margin of the limb contained within the predicted ZPA (Fig. 1). This indicated the presence of at least partial regulatory activity contained within the 8.5 kb of flanking genomic DNA in Cos33S.

Conservation of intronic regions within the *Lmbr1* gene

Interspecies conservation of nucleotide sequence may pinpoint important regulatory elements. The sequence of intron 5 of the *Lmbr1* gene in several vertebrate species was compared to search for conserved domains which would account for the limb expression displayed by Cos33S. The partial sequence available (accession no. AC0587880) for intron 5 in the mouse was used to further establish high quality sequence for this region and was compared to the available human sequence (accession no. AC007097). Comparison of the 31 kb mouse and 33 kb human intron 5 sequences using VISTA (14,15) showed a number of regions of significant similarity, the largest being a region of 1.3 kb (Fig. 2A). In addition the entire 18.5 kb chick intron 5 was sequenced. Comparison of this sequence to the mouse significantly reduced the overall similarity displayed by the VISTA plot and demonstrated a single conserved region of ~800 bp corresponding to the 1.3 kb region predicted in human and mouse (Fig. 2A).

Mapping the genomic flanking DNA contained in Cos33S (discussed above) established the region of the intron represented by the cosmid clone. Interestingly, the genomic DNA extended upstream and included about 800 bp of the 1.3 kb conserved domain (green arrow at nucleotide 400 in Fig. 2B). The data is consistent with the regulatory activity detected by transgenesis residing within the conserved domain of mouse, human and chick.

Identification of a ZPA-specific enhancer

A 1.7 kb *Hind*III fragment containing the mouse conserved region was incorporated into transgenic constructs containing the heterologous β-globin promoter and *lacZ* reporter gene (construct i in Fig. 3A) (16). Transgenic embryos showed similar β-gal staining patterns at the posterior margin of both the fore and hindlimbs (Fig. 3B) reminiscent of limb *Shh* expression in the ZPA. Much of the regulatory information for driving expression in the ZPA appears to be contained within this fragment, thus we refer to this conserved domain as the ZPA regulatory sequence (ZRS).

In addition to the *Lmbr1* sequences in species of land-based tetrapods, we attempted to detect similar ZRS sequences in fish. Cosmid and PAC clones from *Fugu rubripes*, the pufferfish, were identified. The compact genome of *Fugu* (17) facilitated the demonstration of the physical linkage of *Lmbr1* and *Shh* (manuscript in preparation). Genomic sequence was established and the region corresponding to intron 5 was compared to human, mouse and chick (the VISTA comparison with mouse sequence is shown in Fig. 2A). This highlighted the presence of a region of ~400 bp conserved between all four species. A 445 bp fragment of the mouse ZRS incorporating the *Fugu*

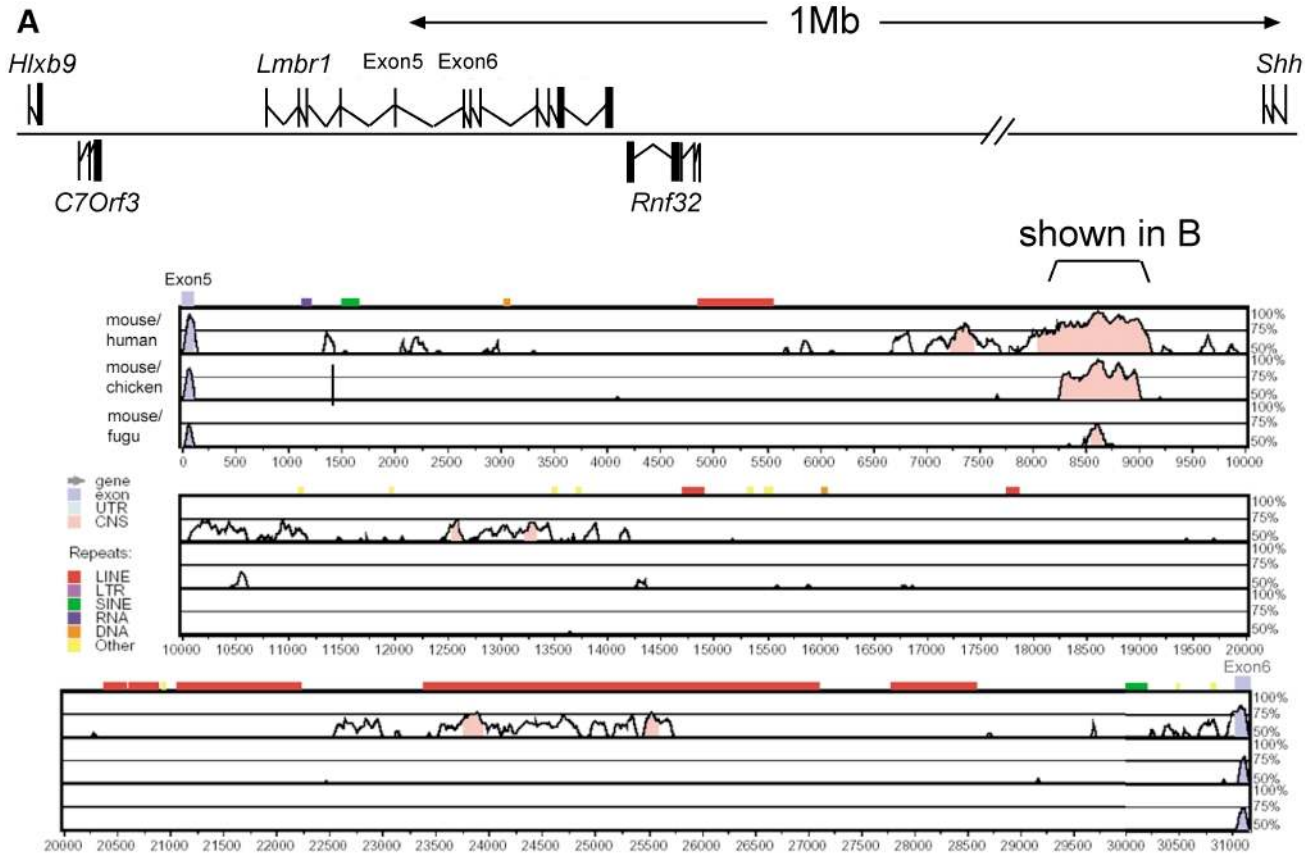

Figure 2. Interspecific sequence comparison of four vertebrate species for intron 5 of the *Lmbr1* gene. (**A**) Schematic depiction of the genomic organization in mouse from *Hlxb9* to *Shh*, reflecting the transcribed strand (taken from Ensembl and based on the mouse genome assemble V3). Exons 5 and 6 of *Lmbr1* are labelled and below a sequence comparison of the intervening intronic DNA (intron 5) predicted by VISTA is depicted. Species used for the comparison (mouse, human, chick and *Fugu rubripes*) are shown along the left side of the VISTA plot. Annotation along the top shows position of exons and predicted genomic repeats. Mouse was used as the base sequence for all the comparisons and predicted regions of significant identity are shown (50% identity in windows of 100 bp appear on the plot and those above 75% are shaded pink). The conserved region aligned in (B) is bracketed. (**B**) Alignment of the highly conserved region from all four species. Nucleotides conserved within all four species are shaded black, while those conserved within only three species are shaded grey. The end of the cosmid (Cos33S) is marked with a green arrow. The ends of the 445 bp constructs (constructs ii and iv) are marked with red arrows, and the end of construct iii is indicated with a blue arrow. The locations of the single base pair changes identified in the human PPD families and in the *Hx* mouse are marked with magenta stars.

region of similarity (nucleotides 28–568, the region between the red arrows in Fig. 2B) was used in a transgenic construct (construct ii, Fig. 3A). This construct drives *lacZ* expression within the putative domain of the ZPA of both the developing fore and hindlimb buds (Fig. 3C). Thus the fish sequence comparison has aided in identifying a limited region of activity.

An additional construct was made based on the endpoint (green arrow in Fig. 5B) of the genomic DNA in cosmid Cos33S. The 5′ end fragment from the 1.7 kb ZRS domain (5′ of the blue arrow in Fig. 2B) incorporating ~300 of the 445 bp in construct ii was used (construct iii; Fig. 3A). Construct iii drove expression in the posterior margin of fore and hind limbs (forelimbs shown in Fig. 3D), thus we suggest that most of the spatial activity is contained in a small fragment of 305 bp (nucleotides 20–417 in Fig. 2B).

Identification of the fin bud element

To determine the regulatory capacity of the fish conserved region, a construct containing the 502 bp *Fugu* DNA (construct iv, Fig. 3A) most similar to the mammalian ZRS was made. The resulting E10.5 transgenic embryos express a spatial pattern indistinguishable from the ZPA-like expression pattern (Fig. 3E) of the mouse (construct ii). The region of similarity predicts a functional regulatory unit in the fish. Since limb and fins are homologous structures and there is broad conservation of synteny in vertebrates (manuscript in preparation), we argue that this regulatory element is responsible for the expression of *Shh* observed in the posterior of the developing fin bud (18). In addition the basis for conservation of synteny may be a consequence of these distal regulatory elements.

Spatio-temporal activity of the ZRS

In order to examine the time course of activity, permanent transgenic lines were made with both the 1.7 kb, ZRS construct i, and the mouse 445 bp construct ii. Of 11 independent lines carrying construct i; five express β-gal with a consistent pattern

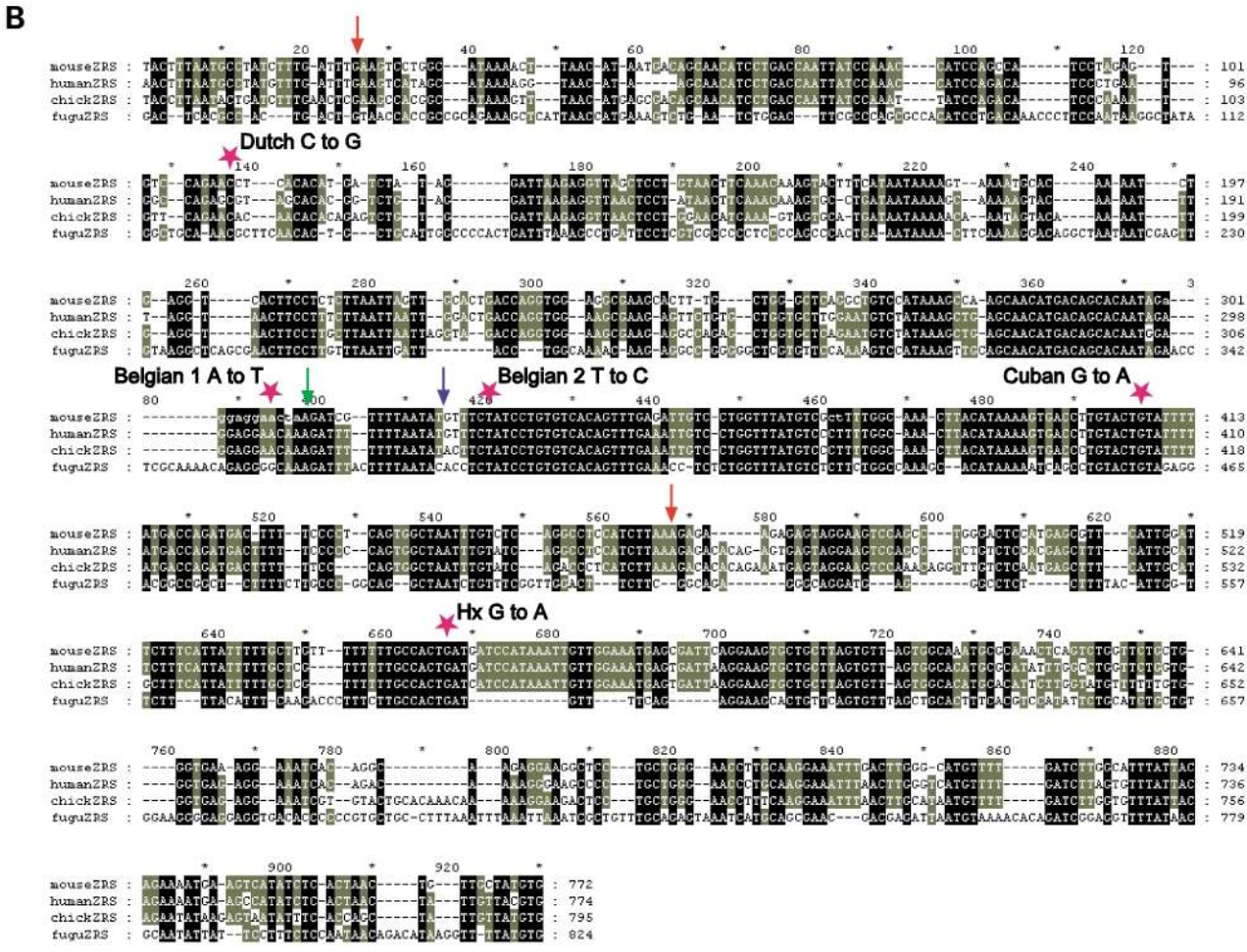

Figure 2 *Continued.*

in the limb buds. At E10.5 these lines show an accurate initiation of expression of β-gal firstly, within a small patch in the fore limb bud and, subsequently, in the hind limb (Fig. 4A). The expression expands into the posterior margin and very closely mimics the endogenous *Shh* pattern demonstrated by *in situ* hybridization (Fig. 4B). This correlation between β-gal activity and the *Shh* expression pattern continues to E11.5 (Fig. 4C). However, after this stage endogenous *Shh* expression is reduced and is undetectable in the limb buds at E12.5. In contrast the five ZRS lines continue to express the β-gal activity to at least E13.5 (Fig. 4E). *In situ* hybridization to examine *lacZ* mRNA showed that the transgene was being actively expressed at this stage (Fig. 4F), and that the β-gal activity does not simply reflect perdurance of the protein. However, the distribution of the β-gal activity shows some differences with that of the *lacZ* mRNA since the β-gal staining extends down the posterior shank of the limb (compare Fig. 4E and F at E13.5, and H and I at E12.5). In addition, at E12.5 while the *lacZ* RNA is expressed more highly at the limb margin and declines towards the centre of the limb bud (Fig. 4I), the protein has become localized to the developing digits, giving rise to a V-shaped staining pattern (Fig. 4H).

In contrast, the 445 bp element appears to be a weaker regulator in transgenics. Whereas in the transient transgenic embryos expression from the 445 bp construct ii is detectable at E10.5, the β-gal was difficult to detect until E11.5 in the two stable lines. Once again, as with the 1.7 kb construct, expression is then maintained through E12.5 (Fig. 4L) until at least E13.5 (Fig. 4M). This anomalous late-stage digit expression reflects the pattern observed for the HPAP activity reported for the *Ssq* mutant (Fig. 4G) (7), demonstrating that ZRS contains most, if not all, the activity detected by the *Ssq* HPAP reporter inserted in the *Lmbr1* gene.

Point mutations within the conserved domain associated with preaxial polydactyly

Having identified the functional ZRS domain, a total of seven families with preaxial polydactyly, mapping to chromosome 7q36, were analysed. A large Dutch family (9,12) was tested

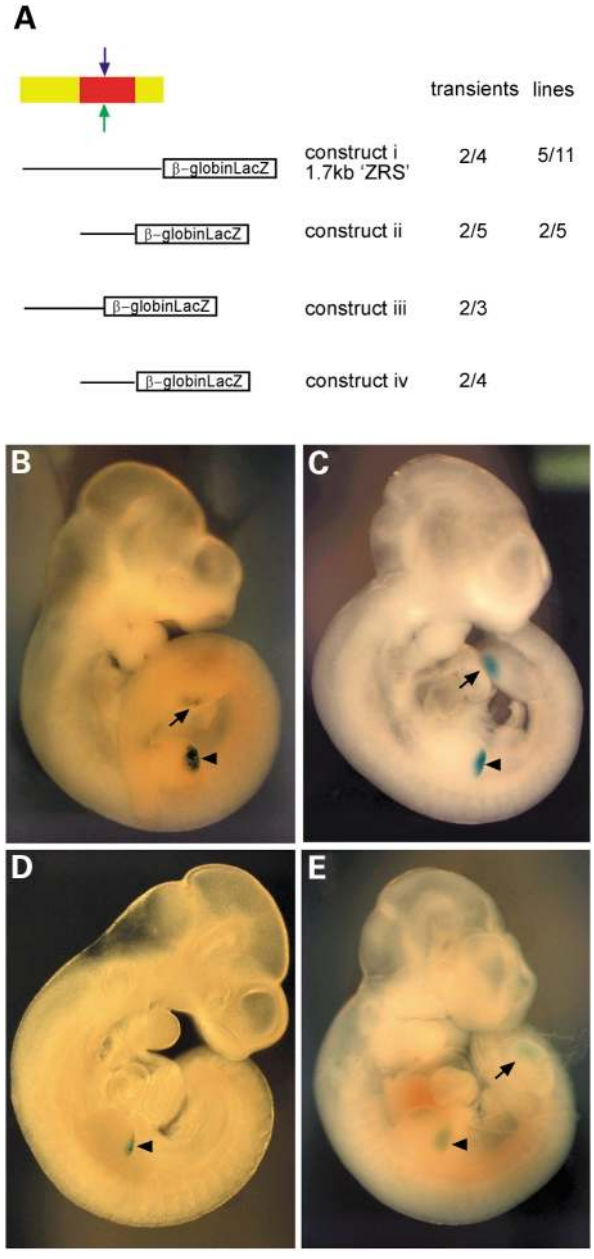

Figure 3. Summary of the constructs used to produce transgenic mice. (**A**) Schematic representation of the constructs used. The entire 1.7 kb ZRS is depicted by a yellow box, the 445 bp internal fragment by a red box, the end of the cosmid by a green arrow and the end of construct iii by the blue arrow. The region incorporated into each construct is depicted below. The table on the right depicts the number of G_0 (transients) embryos and lines as a ratio of expressing/transgenic embryos. (**B**–**E**) show β-gal expression in E10.5 G_0 embryos carrying (B) construct i, the 1.7 kb ZRS, (C) construct ii, the 445 bp mouse construct; (D) construct iii, the 305 bp fragment, and (E) construct iv, the 502 bp *Fugu* construct. Expression of the reporter gene in the posterior margin of the forelimbs indicated with arrowheads and hindlimbs with arrows.

and the 96 affected individuals were all found to be heterozygous for a C/G transversion at position 105 (Fig. 5E) of the human sequence as shown in Figure 2B. This in contrast to the 117 unaffected individuals (sibs or partners), who are homozygous C/C. As the large Dutch family is from a genetically isolated village, we tested 249 unaffected individuals from this village using ASO hybridization. In addition, we tested 128 controls from around the village, as well as 183 controls from The Netherlands (giving a total of 1354 tested control chromosomes) and all were found to lack this mutation.

In addition, we identified a DNA alteration in three smaller families. Two of these families were identified in Belgium and are designated as the Belgian 1 and Belgian 2 families (Fig. 5). In the Belgian 1 family, an A/T transversion (Fig. 5E) at position 305 of the human ZRS sequence segregated with the PPD phenotype (seven affected A/T, six unaffected A/A; pedigree shown in Fig. 5A) and was found to be absent in 183 Dutch and 95 Belgian controls. A T/C transition at position 323 of the human ZRS sequence (three affected T/C, four unaffected T/T, Fig. 5E), segregating with the PPD phenotype in the Belgian 2 family (pedigree shown in Fig. 5B), was absent in the same 183 Dutch and 95 Belgian controls. Finally, a G/A transition at position 404 (Fig. 5E) was detected in all six affected individuals in a Cuban family (12) and was absent in six unaffected relatives as well as 45 Cuban and 183 Dutch unaffected individuals. The sequence analysis also resulted in the identification of four other sequence alterations (C/G at position 3, C/T at position 236, T/C at position 295 and G/C at position 507 of the human ZRS sequence), which were detected in unaffected individuals (with frequencies of 10–30% in both Dutch controls and families), suggesting these are polymorphisms. We therefore suggest that the base pair differences found in the Dutch, Belgian and Cuban families are the pathogenic mutations leading to the PPD phenotype in these families. In the remaining three families we only found the C/G polymorphism at position 3 also present in unaffected individuals.

Previously, we showed that a region of ~24 kb of intron 5 of *Lmbr1* was duplicated in the *Ssq* genome (8). Here we show that the transgenic insertion is situated between the duplication endpoints with copies of the ZRS on each side (Fig. 5F); however, it is unclear how this contributes to ectopic expression. A second allelic mutation was examined and proved more informative. The *hemimelic extra toes* (*Hx*) mutation has a similar phenotype to *Ssq* (19) and a critical region that includes *Lmbr1* and the downstream *Rnf32* gene. No mutations were found in the coding region of these genes (20). The ZRS and three other highly conserved domains (data not shown) located within the critical genetic region were sequenced. A G to A transition was found at position 545 of the mouse ZRS sequence of the *Hx* mutant (Fig. 5G); all other regions sequenced were unaffected. The *Hx* mutation arose on the congenic line B10.D2/nSn (21) derived from the inbred strain C57BL/10Sn and the donor partner DBA/2J. Thus the ZRS regions from the B10.D2/nSn, C57BL/10Sn and DBA2J strains plus seven other control inbred strains (C57BL/6, C3H, CBA, AKR, 129SV, FVB and A/J) were sequenced and found to lack this G/A alteration. In addition, no other DNA alterations were found in these strains.

The base pair changes detected in mouse and human occur at highly conserved nucleotides. All are conserved in the ZRS of mouse, human and chick and all except the mutation found in the Belgian 1 family also include conservation with fish.

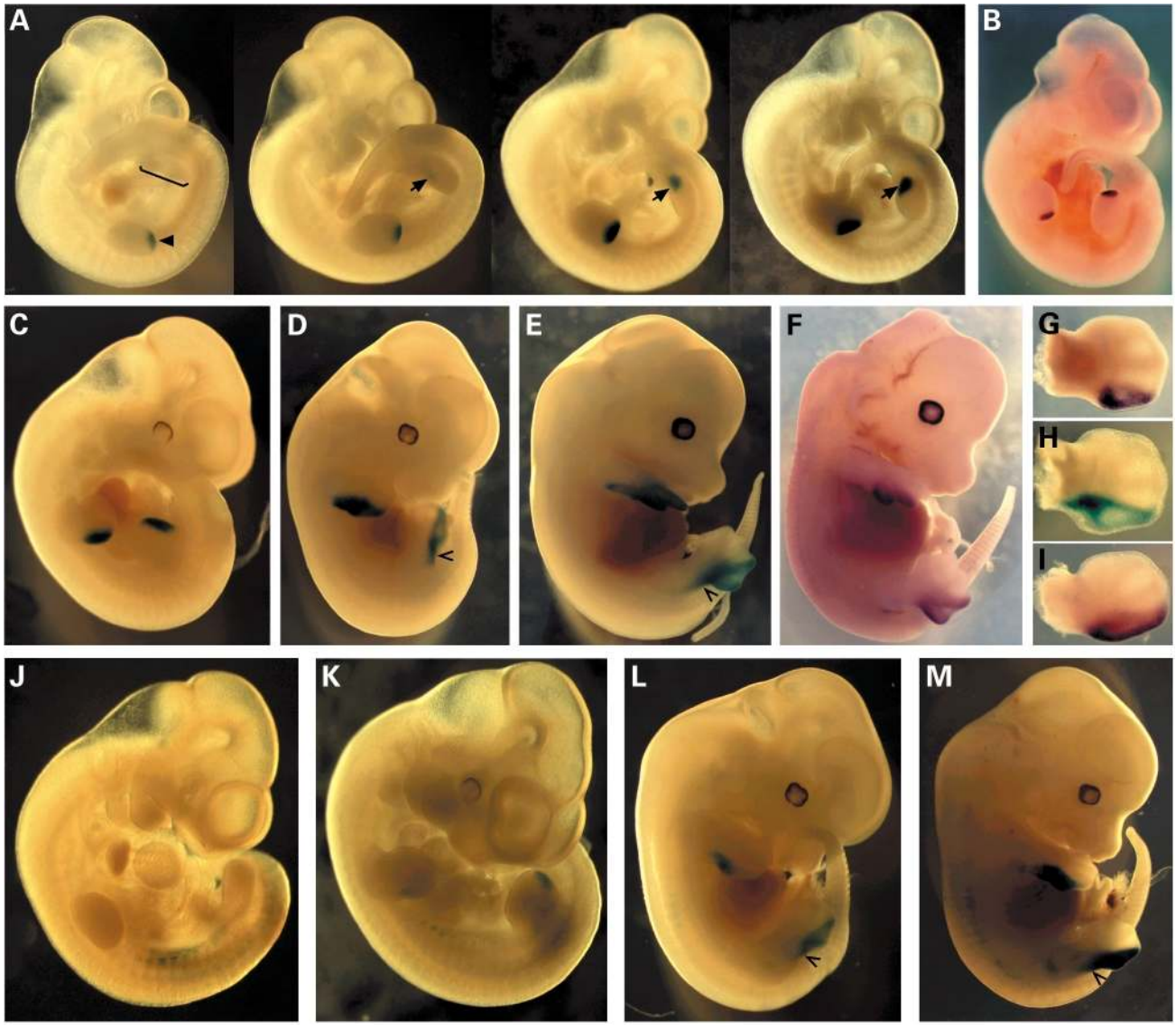

Figure 4. Analysis of spatio-temporal patterns in embryos from permanent transgenic lines. Embryos from lines carrying construct i (**A**–**F**, **H** and **I**) and construct ii (**J**–**M**). (A) Increasingly older embryos (from left to right, harvested at E10.5 and staged relative to each other by size of hindlimb bud) stained for β-gal expression show accurate activation of the transgene. The early hindlimb bud is indicated with a square bracket, hindlimb expression with an arrow and forelimb expression with an arrowhead. (B) Expression of *Shh* (by *in situ* hybridization at E10.5) for comparison. Later embryos at E11.5 (C) show β-gal expression comparable to Shh, whereas, embryos at E12.5 (D) and E13.5 (E) show persistence of β-gal staining. (F) *lacZ* mRNA *in situ* hybridisation on a transgenic littermate of (E) demonstrates that the transgene continues to be expressed. Forelimbs from E12.5 embryos are shown in (G and H). (**G**) A *Ssq*/+ mutant embryo stained for the HPAP reporter gene displays the posterior digital expression pattern. Expression from the transgene analysing for β-gal staining (H) and *lacZ* mRNA (I) shows that the transgene mRNA and HPAP staining are similar but that the protein perdures extending proximally down the shank. (J–M) β-gal staining in embryos carrying the 445 bp mouse construct ii at E10.5 (J), E11.5 (K), E12.5 (L) and E13.5 (M). Persistance of expression down the shank of the limb bud is indicated with an open arrowhead (D, E, L, M).

DISCUSSION

ZRS and polydactyly

Examples of suspected regulatory elements that operate over extreme distances are increasingly being recognized. A series of chromosomal aberrations that are purportedly regulatory mutations for the developmental genes *Sox9* (22) and *MAF* (23) lie at distances of about 1 Mb from the affected genes. In addition a number of studies indicate an association between chromosomal aberrations and congenital abnormalities in which the aberrations reside at moderate distances (reviewed in 24). In many of these cases the abnormalities are suggested to be due to chromosomal position effects. Here, we show that single basepair mutations residing within the ZRS regulatory element are associated with the congenital abnormality PPD. These mutations are located at a dramatic distance from the target gene, *Shh*. The identification of such subtle regulatory mutations is, of course, problematic. In the case of preaxial polydactyly the critical genetic region defined

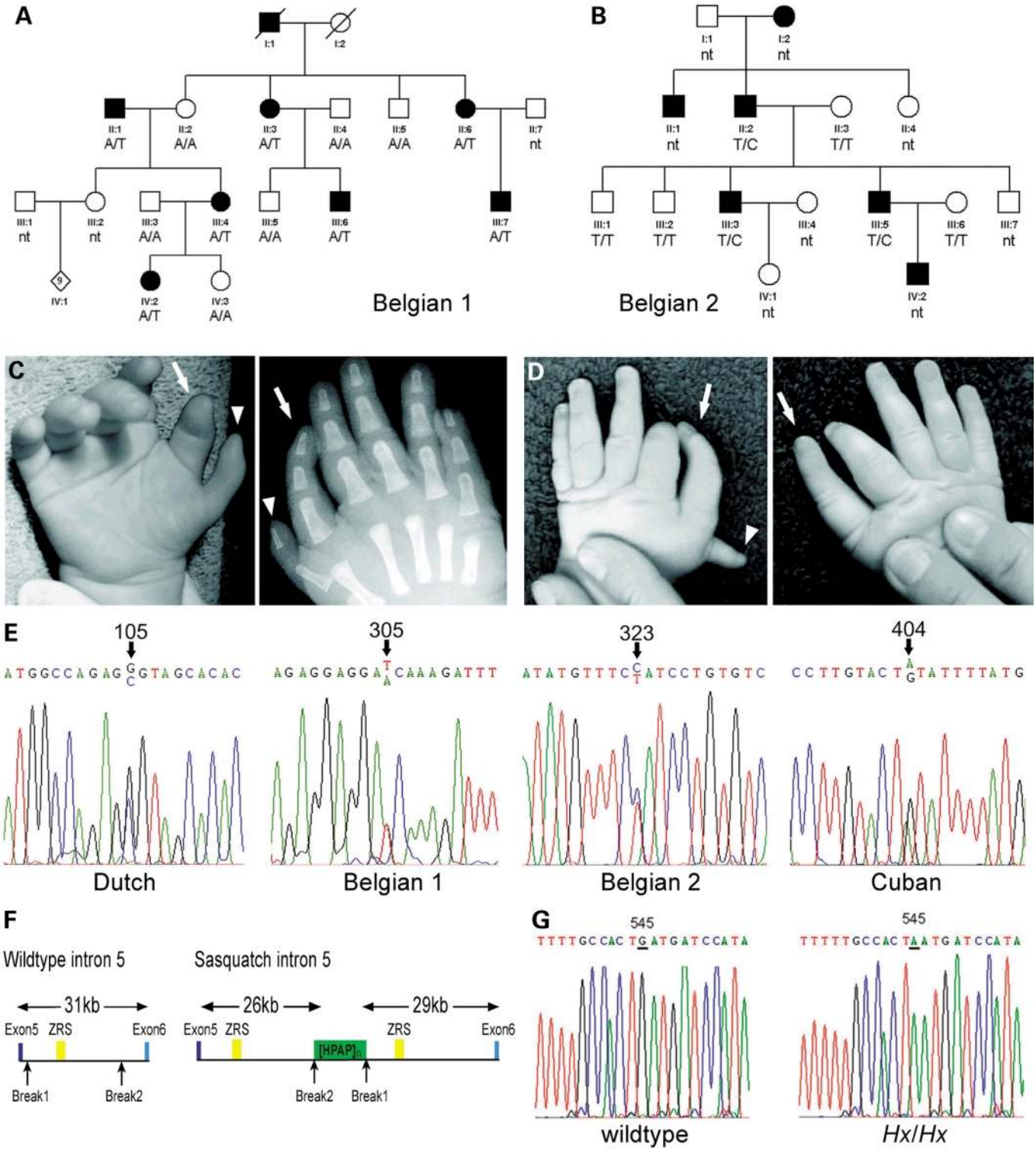

Figure 5. (**A**, **B**) Pedigree structure of part of a Belgian family 1 (A) and family 2 (B) with autosomal dominant preaxial polydactyly. Solid symbols represent affected individuals; open symbols represent unaffected individuals. Co-segregation of the A/T mutation in Belgian 1 and the T/C mutation in Belgian 2 with the PPD phenotype is indicated. (**C**) Normal and X-ray images of the right hand of affected individual IV-2. Note the extra biphalangeal thumb (arrowhead), with metacarpal, connected to a tri-phalangeal thumb (arrow). (**D**) Normal image of both hands of the affected individual IV-2 of Belgian family 2. Note the triphalangeal thumb on the left hand (arrow). (**E**) Sequence electropherogram of mutations found in the ZRS. Positions of the heterozygous mutations in the human sequence of the ZRS (Fig. 2B) are indicated above. (**F**) Schematic representation of *Lmbr1* intron5 in wild-type mice and in Ssq. Exons 5 and 6 are depicted by dark and light blue boxes, respectively. The ZRS is shown as a yellow box and the breakpoints of the duplication marked by arrows. In Ssq, the inserted multiple transgene elements are shown as the green box. (**G**) Sequence electropherogram of a wild-type and an Hx/Hx mouse. The G to A mutation is at position 545 of the mouse sequence of ZRS as indicated in Figure 2B.

for both mouse and human lies at a significant distance from the target gene and, in fact, the *Shh* gene and both the *Ssq* and *Hx* mouse mutations are at an approximate genetic distance of 1.8 cM (8,20). The task of identifying distantly located regulatory mutations required an initial search for functional regulatory elements within the critical region by a process of sequence comparisons in diverse species, and subsequent experimental verification of enhancer activity. Preaxial polydactyly is unlikely to be unique among congenital dysmorphologies and we suggest that a number of long-range regulatory mutations will be uncovered using similar approaches.

In mammals the ZRS is contained within a highly conserved 1.3 kb region. The minimal region required for driving ZPA expression in the transgenic constructs is found in a sequence of about 300 bp. The mutations associated with polydactyly are scattered through a region of 450 bases of the ZRS and map both inside and outside the minimal region required for expression activity. The lack of clustering of the mutations in the ZRS suggests there is no single subdomain responsible for mis-expression. In addition the mutations effect mis-expression of the *Shh* gene without disrupting the normal production at the posterior limb bud margin. Maintenance of normal ZRS regulatory activity is supported by studies in both mouse mutants, *Ssq* and *Hx*, in which posterior *Shh* expression in the homozygous mutants and patterning of the posterior digits is similar to wild-type. Thus we suggest that normal posterior activity is unaffected by ZRS mutations which, however, affect spatial specificity causing anterior, ectopic expression.

In our analysis multiple regulatory functions were assigned to the ZRS. The ZRS is responsible, firstly, for initiation of *Shh* expression at the appropriate time in limb bud development and, secondly, for driving the appropriate posteriorly restricted spatial expression pattern. However ZRS activity seems acutely sensitive to structural perturbations since single base pair changes appear to result in loss of expression asymmetry. The mechanism in which the base pair changes participate remains to be resolved. One approach to investigate this mechanism may entail the production of mutations or deletions in the ZRS/reporter gene constructs to generate ectopic reporter gene expression. However, preliminary evidence suggests expression levels of the reporter gene are not sufficiently high enough to be detected (data not shown). Based on the previous observation (8) that ZRS mutations (as in the *Ssq* mutation) when linked to an inactive *Shh* null allele suppress the ectopic extra digit, we suggest that the process of ectopic tissue expansion is requisite for the detection of anterior expression. Therefore straightforward ZRS/reporter constructs carrying mutations may not provide the experimental route to understand ectopic activity and further analysis incorporating *Shh* expression will be required.

A number of genetic resources are available to investigate the action of the ZRS in both normal posterior expression and anterior ectopic expression of *Shh.* The frequency of PPD in the human population opens the possibility of detecting a number of other mutations in the ZRS (currently under investigation), which may eventually aid in pinpointing functional domains within the ZRS. In addition in mouse other unlinked mutations, *lx*, *Dh*, *Rim4*, *Xpl* and *lst*, are known to produce preaxial polydactyly with associated ectopic *Shh* expression (4–6). These mutations are postulated to act by de-repressing *Shh* expression in the anterior margin and thus may function through the ZRS. The *lst* mutation is a particularly attractive candidate. The gene responsible is the *Alx4* gene (25,26), which is expressed in the anterior margin of the early limb bud. Examination of these polydactylous mouse mutations may be instrumental in pinpointing interacting genes which act at the ZRS.

Regulation at a distance

Chromatin or chromosomal structural elements are expected to play a part in regulating gene expression at such a distance. Here, we find that at least one major component of a long-range regulator is a distance-independent enhancer. Our data, however, question the necessity for the long-range separation of regulator and the target gene. Clearly, the genomic configuration of the ZRS situated inside the *Lmbr1* gene at a considerable distance from *Shh* is conserved over a significant evolutionary period. Does this long-distance separation play a role in the regulation? The transgenic approach poses an answer by effectively eliminating the long distance upon which the ZRS must act. The transgenes were constructed such that the enhancer abuts the β-globin heterologous promoter driving the reporter gene. The ZRS containing transgene initiates expression at a stage similar to that of endogenous *Shh*, and the subsequent spatial pattern is indistinguishable. Therefore, neither genomic distance nor chromosomal context appears to be important in defining the early parameters of Shh expression. In contrast, the transgenic lines do not 'turn off' expression at the normal endpoint (3). Resolution of the temporal pattern requires further input, leaving open the possibility that distance of the ZRS may play a crucial role in the appropriate resolution of the expression pattern.

Long-distance gene regulation must overcome a number of inherent problems. For example, local gene expression appears unperturbed by the presence of the ZRS. The ZRS lies within intron 5 of the ubiquitously expressed *Lmbr1*. In addition, *Lmbr1* is situated among a cluster of genes including *Rnf32* a testes specific gene (27) positioned between the ZRS and *Shh* and C7orf3, a widely expressed gene (13) and *Hlxb9*, a gene involved in neural tube closure (28) and pancreatic development (29) positioned directly upstream. The ZRS appears, therefore, to be a component of a complex of regulators that together drive the spatial and temporal pattern in the ZPA, while insulating the surrounding genes, and seeking out the *Shh* promoter specifically.

Evolution of *Shh* expression

The tetrapod limb was an essential morphological adaptation in vertebrate evolution that facilitated the transition from an aquatic to a land-based lifestyle. Expression of teleost HoxA and HoxD genes, orthologues of genes expressed in tetrapod limbs, suggest that the fin is homologous to proximal limb structures (30). Additional anatomical data supports the notion that in the fin-to-limb transformation advent of the autopodium (structure which gives rise to the carpus/tarsus and digits) was the distinguishing innovation (reviewed in 31). Generally in tetrapod limbs, *Shh* expression has acquired a function in the

autopod to pattern the chondrogenic elements. In teleosts, fin structure differs quite dramatically from limb and *Shh*, although expressed in the posterior margin, differs from the mouse pattern in that it does not extend to the distal edge (18). Given the substantial differences, is the fin *Shh* expression domain equivalent to the ZPA? Initially, the ZPA was defined functionally in chick explant studies (1). Since the characteristic ZPA properties are mediated by SHH (2,3), the ZPA has become synonymous with the limb expression pattern of *Shh*. At the molecular level much of the information for the specific *Shh* expression is contained in the ZRS. The data presented here strongly suggest that fin and limb *Shh* expression are under the same regulatory influences and thus the fish expression domain of *Shh* and the ZPA are equivalent.

Elaboration of *Shh* activity in producing the pattern particular to each type of appendage is different. Clearly, the ZRS of the fish, although driving a posterior but proximal pattern in the fin, has the capacity to regulate expression in a pattern remarkably similar to that of the mouse limb. Hence, within the vertebrate evolutionary scheme the ZRS regulatory element is ancient, having arisen prior to the fin-to-limb transition. The evolutionary period in which fin expression arose may have been subsequent to chondrichthyes since dogfish do not express *Shh* in the fin (32). The ZRS and consequently the ZPA arose, but not as part of a mechanism to pattern digits and therefore predate the advent of the tetrapod specific autopod. With the basic regulatory mechanism for *Shh* expression in place, it appears that the role of *Shh* changed within the context of the evolving limb bud to define digit number and pattern.

MATERIALS AND METHODS

Sequencing and analysis

High quality mouse intron 5 sequence was assembled from available unordered HTGS sequence of clone RP23-284A9 (accession no. AC0587880), in combination with λ and cosmid clone sequence containing the *Ssq* insertion site. This led to an ordered array of intron 5 sequences, and gaps were subsequently filled by PCR and sequencing. Finished human sequence was available in the database (accession no. AC007097). Database searching identified chicken *Lmbr1* ESTs (ChEST294A11, 382F7 and 771N8), which were used to screen the chicken BAC library filters (HGMP). Two clones were identified (10P15 and 75K5); subsequent characterization showed that 75K5 contained the entire *Lmbr1* gene and this was used as a template for long-range PCR (Roche) using primers from exons 5 and 6 (CTTGGAGTCGGAAGGATTTGC and TCTCCAGAATGCGTGCTCTG). The resultant PCR product was shotgun sequenced using standard methods. Fugu *Lmbr1* intron 5 sequence was obtained from a *Shh* containing PAC that was completely sequenced.

Initial sequence comparison was conducted using VISTA (www-gsd.lbl.gov/vista) (14,15) and the sequence displayed using PileUp (The GCG Wisconsin Package at the HGMP, Hinxton, UK) and ClustalX (HGMP) (33). The alignment in Figure 2 was prepared using GeneDoc (www.psc.edu/biomed/genedoc).

Transgenic analysis

The cosmid Cos33S containing the *Ssq* transgene and flanking DNA was digested with *Not*I to release the insert. Both insert and vector were injected. All other transgenic constructs made use of a β-globin minimal promoter and the bacterial *lacZ* reporter gene (16) (vector p1230, a kind gift from Robb Krumlauf). Construct i (1.7 kb, ZRS construct) was made by subcloning the 1.7 kb *Hind*III genomic fragment. Subfragments of this were generated by PCR and sequenced to ensure the absence of PCR errors. The *Fugu* construct iv, used *Fugu* cosmid 16E21 (HGMP) as a template. All PCR primers used had *Hind*III sites for ease of cloning and vectors were made with the PCR products in both orientations relative to *lacZ*. In all cases, the vector fragment was removed by *Not*I/*Sal*I digestion. Transgenic mice were made by pronuclear injection using standard protocols. G_0 embryos were harvested at E10.5 (assuming the day of transfer to be E0.5) or allowed to develop to term. Transgenic males were subsequently used as studs with CD1 females. All embryos harvested had their yolk sacs retained to allow for PCR genotyping.

HPAP, β-gal staining and *in situ* hybridizations were carried out using standard techniques (7,34,35 respectively). The probes used for *in situ* hybridization were *Shh* (a kind gift from Andy McMahon) or *lacZ*, transcribed from vector p1230.

Mutation detection

The Dutch and Cuban families were previously mapped to chromosome 7q36 (9,12). Affected individuals from the Belgian 1 family (Fig. 5A) display preaxial polydactyly II, with an extra thumb attached to the triphalangeal thumb (Fig. 5B). The phenotype of the Belgian 2 family (Fig. 5C) included triphalangeal thumbs with or without an additional biphalangeal thumb (Fig. 5B). The phenotype in both families was linked to chromosome 7q36.

Mouse strains tested for mutation analysis were: B10.D2/nSn-Hx/Hx (and Hx/+), B10.D2/nSn, C57BL/10SnJ, DBA2J, C57BL/6, C3H, CBA, AKR, 129SV, FVB and A/J (Jackson labs).

Genomic DNA was isolated from peripheral blood from affected and unaffected individuals according to standard procedures (36). For mutation analysis of the ZRS in both human and mouse, PCR reactions were performed in 50 µl containing 1 × GibcoBRL PCR buffer, 1.5 mM $MgCl_2$, 200 µm dNTPs, 200 µm primers, 2.5 units of *Taq* DNA polymerase (GibcoBRL) and 50 ng of genomic DNA. Cycling conditions were: 3 min at 94°C, 35 cycles of 15 s at 94°C, 15 s at 56°C and 35 s at 72°C followed by 5 min at 72°C. PCR products were purified using Amersham GFX purification columns, according to the products instructions.

Direct sequencing of both strands was performed using Big Dye Terminator version 3.0 chemistry (Applied Biosystems). Fragments were loaded on an automated sequencer and analysed with DNA Sequencing Analysis (version 3.7) and SeqScape (version 1.1) packages (Applied Biosystems). To ensure high quality sequencing we divided the ZRS and 200 nt up- and downstream in three overlapping regions: Hi5f/R1, Hi5f/R2 and Hi5f/R3 for human and mi5f/R1, mi5f/R2, mi5f/R3 for mouse. The primers used for amplification were also used for sequencing and are given in Table 1. For Allele

Table 1. Primers used for sequence analysis

	Forward 5′–3′	Position in ZRS[a]	Reverse 5′–3′	Position in ZRS[a]
Hi5f1/Hi5R1	ggaggtataacctctggccagtg	−221/−198	cgcttccacctggtcagtcc	217/237
Hi5f2/Hi5R2	ccagagcgtagcacacggtc	98/118	caatttatggatcatcagtggc	548/570
Hi5f3/Hi5R3	tcaggcctccatcttaaagag	449/470	atcttctgatccataaccatttc	877/901
mi5f1/mi5R1	gagggcaatgtactggccagtg	−223/−201	cttcctgaatcgctcatttcc	228/248
mi5f2/mi5R2	ggattaagaggttagctcctg	127/148	tgcttcgcctccacctggtc	571/592
mi5f3/mi5R3	tcaggcctccatcttaaagag	451/472	gcacacccacccgtcacagaag	919/941

[a]The position of the primers is given relative to the first A (human) or the first T (mouse) of the ZRS as shown in Figure 2.

Table 2. Primers used for allele specific oligo hybridization

	Wild-type oligo	Mutation oligo
Aso-Dutch	ccagagCgtagca	ccagagGgtagca
Aso-Belgian 1	ggaggaAcaaagatt	ggaggaTcaaagatt
Aso-Belgian 2	tatgtttcTatcctgt	atgtttcCatcctgt
Aso-Cuban	cttgtactGtattttat	cttgtactAtattttat
Aso-*Hx*	tgccactGatgatcc	tgccactAatgatcc

Specific Oligo (ASO) hybridization, PCR products similar to those sequenced were blotted onto Hybond-N+ membranes (Amersham Biosiences). Blots were hybridized for 1 h at 37°C in 5 × SSPE, 1% SDS and 50 µg/ml single-strand salmon sperm DNA with either the normal or mutated oligo, labelled with P^{32}-γATP. Filters were washed at 37°C until a final stringency of 0.3 × SSC/1% SDS. Oligonucleotides used for hybridization for the mutations can be found in Table 2.

ACKNOWLEDGEMENTS

We would like to thank the families for participating and M. Joosse, H. van der Linde, C. Huysmans, B. Doe, S. Rhodes and T. Chapman for expert technical assistance. We greatly appreciate the help of J. Zguricas, Professor S.E.R. Hovius, E. Morales-Peralta, M. Holvoet, E. Pykelsin and F. Stockmans in collecting blood samples and ascertaining family members and D. Lambrechts for collecting Belgian control samples. We also thank Professor Nick Hastie for critically reading the manuscript.

REFERENCES

1. Sanders, J.W. Jr and Gasseling, M.T. (1968) Ectodermal–mesenchymal interactions in the origin of limb symmetry. In Fleischmajar, R. and Billingham, R.F. (eds), *Ectodermal–Mesenchymal Interactions.* Williams and Wilkins, Baltimore, MD, pp. 78–97.
2. Riddle, R.D., Johnson, R.L., Laufer, E. and Tabin, C. (1993) Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell*, **75**, 1401–1416.
3. Echelard, Y., Epstein, D.J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J.A. and McMahon, A.P. (1993) Sonic hedgehog, a member of a family of putative signalling molecules, is implicated in the regulation of CNS polarity. *Cell*, **75**, 1417–1430.
4. Masuya, H., Sagai, T., Wakana, S., Moriwaki, K. and Shiroishi T. (1995) A duplicated zone of polarizing activity in polydactylous mouse mutants. *Genes Dev.*, **9**, 1845–1853.
5. Masuya, H., Sagai, T., Moriwaki, K. and Shiroishi, T. (1997) Multigenic control of the localization of the zone of polarizing activity in limb morphogenesis in the mouse *Dev. Biol.*, **182**, 42–51.
6. Lettice, L., Hecksher-Sørenson, J. and Hill, R.E. (1999) The dominant hemimelia mutation uncouples epithelial–mesenchymal interactions and disrupts anterior mesenchyme formation in mouse hindlimbs. *Development*, **126**, 4729–4736.
7. Sharpe, J., Lettice, L., Hecksher-Sorensen, J., Fox, M., Hill, R.E. and Krumlauf, R. (1999) Identification of Sonic hedgehog as a candidate gene responsible for the polydactylous mouse mutant Sasquatch. *Curr. Biol.*, **9**, 97–100.
8. Lettice, L.A., Horikoshi, T., Heaney, S.J.H., Van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N. *et al.* (2002) Disruption of a long range *cis*-acting regulator for *Shh* causes preaxial polydactyly. *Proc. Natl Acad. Sci. USA*, **99**, 7548–7553.
9. Heutink, P., Zguricas, J., van Oosterhout, L., Breedveld, G.J., Testers, L., Sandkuijl, L.A., Snijders, P.J., Weissenbach, J., Lindhout, D., Hovius, S.E. *et al.* (1994) The gene for triphalangeal thumb maps to the subtelomeric region of chromosome 7q. *Nat. Genet.*, **6**, 287–292.
10. Hing, A.V., Helms C., Slaugh R., Burgess A., Wang J.C., Herman, T., Dowton, S.B. and Donis-Keller, H. (1995) Linkage of preaxial polydactyly type 2 to 7q36. *Am. J. Med. Genet.*, **58**, 128–135.
11. Tsukurov, O., Boehmer, A., Flynn, J., Nicolai, J.P., Hamel, B.C., Traill, S., Zaleske, D., Mankin, H.J., Yeon, H., Ho, C. *et al.* (1994) A complex bilateral polysyndactyly disease locus maps to chromosome 7q36. *Nat. Genet.*, **6**, 282–286.
12. Zguricas, J., Heus, H., Morales-Peralta, E., Breedveld, G., Kuyt, B., Mumcu, E.F., Bakker, W., Akarsu, N., Kay, S.P., Hovius, S.E. *et al.* (1999) Clinical and genetic studies on 12 preaxial polydactyly families and refinement of the localisation of the gene responsible to a 1.9 cM region on chromosome 7q36. *J. Med. Genet.*, **36**, 32–40.
13. Heus, H.C., Hing, A., van Baren, M.J., Joosse, M., Breedveld, G.J., Wang, J.C., Burgess, A., Donnis-Keller, H., Berglund, C., Zguricas, J. *et al.* (1999) A physical and transcriptional map of the preaxial polydactyly locus on chromosome 7q36. *Genomics*, **57**, 342–351.
14. Dubchak, I., Brudno, M., Loots, G.G., Mayor, C., Pachter L., Rubin, E.M. and Frazer K.A. (2000) Active conservation of non-coding sequences revealed by 3-way species comparisons. *Genome Res.*, **10**, 1304–1306.
15. Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S. and Dubchak, I. (2000) VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics*, **16**, 1046–1047.
16. Yee, S.P. and Rigby, P.W. (1993) The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev.*, **7**, 1277–1289.
17. Elgar, G., Sandford, R., Aparicio, S., Macrae, A., Venkatesh, B. and Brenner, S. (1996) Small is beautiful: comparative genomics with the pufferfish (*Fugu rubripes*). *Trends Genet.*, **12**, 145–150.
18. Krauss, S., Concordet, J.P. and Ingham, P.W. (1993) A functionally conserved homolog of the Drosophila segment polarity gene *hh* is expressed in tissues with polarizing activity in zebrafish embryos. *Cell*, **75**, 1431–1444.
19. Knudsen, T.B. and Kochhar, D.M. (1981) The role of morphogenetic cell death during abnormal limb bud outgrowth in mice heterozygous for the dominant mutation Hemimelic-extra toes (*Hmx*). *J. Embryol. Exp. Morphol.*, **65** (suppl.) 289–307.
20. Clark, R.M., Marker, P.C. and Kingsley, D.M. (2000) A novel candidate gene for mouse and human preaxial polydactyly with altered expression in limbs of Hemimelic extra-toes mutant mice. *Genomics*, **67**, 19–27.
21. Dickie, M.M. (1968) *Mouse News Letter*, **38**, 24.

22. Bishop, C.E., Whitworth, D.J., Qin, Y., Agoulnik, A.I., Agoulnik, I.U., Harrison, W.R., Behringer, R.R. and Overbeek, P.A. (2000) A transgenic insertion upstream of sox9 is associated with dominant XX sex reversal in the mouse. *Nat. Genet.*, **26**, 490–494.
23. Jamieson, R.V., Perveen, R., Kerr, B., Carette, M., Yardley, J., Heon, E., Wirth, M.G., van Heyningen, V., Donnai, D., Munier, F. and Black, G.C. (2002) Domain disruption and mutation of the bZIP transcription factor, MAF, associated with cataract, ocular anterior segment dysgenesis and coloboma. *Hum. Mol. Genet.*, **11**, 33–42.
24. Kleinjan D.J. and van Heyningen, V. (1998) Position effect in human genetic disease. *Hum. Mol. Genet.*, **7**, 1611–1618.
25. Qu, S., Niswender, K.D., Ji, Q., van der Meer, R., Keeney, D., Magnuson, M.A. and Wisdom, R. (1997) Polydactyly and ectopic ZPA formation in Alx-4 mutant mice. *Development*, **124**, 3999–4008.
26. Qu, S., Tucker, S.C., Ehrlich, J.S., Levorse, J.M., Flaherty, L.A., Wisdom, R. and Vogt, T.F. (1998) Mutations in mouse Aristaless-like4 cause Strong's luxoid polydactyly. *Development*, **125**, 2711–2721.
27. van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Baarends, W.M., Rizzu, P., de Graaff, E., Oostra, B.A. and Heutink, P. (2002) A double RING-H2 domain in RNF32, a gene expressed during sperm formation. *Biochem. Biophys. Res. Commun.*, **292**, 58–65.
28. Ross, A.J., Ruiz-Perez, V., Wang, Y., Hagan, D.M., Scherer, S., Lynch, S.A., Lindsay, S., Custard, E., Belloni, E., Wilson, D.I. *et al.* (1998) A homeobox gene, *HLXB9*, is the major locus for dominantly inherited sacral agenesis. *Nat. Genet.*, **20**, 358–361.
29. Li, H., Arber, S., Jessell, T.M. and Edlund, H. (1999) Selective agenesis of the dorsal pancreas in mice lacking homeobox gene *Hlxb9. Nat. Genet.*, **23**, 67–70.
30. Sordino, P., van der Hoeven, F. and Duboule, D. (1995) Hox gene expression in teleost fins and the origin of vertebrate digits. *Nature*, **375**, 678–681.
31. Wagner, G.P., and Chui, C.-H. (2001) The tetrapod limb: a hypothesis on its origin. *J. Exp. Zool.*, **291**, 226–240.
32. Tanaka, M., Munsterberg, A., Anderson, W.G., Prescott, A.R., Hazon, N. and Tickle, C. (2002) Fin development in a cartilaginous fish and the origin of vertebrate limbs. *Nature*, **416**, 527–531.
33. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.*, **24**, 4876–4882.
34. Mackenzie, M.A., Jordan, S.A., Budd, P.S. and Jackson, I.J. (1997) Activation of the receptor tyrosine kinase Kit is required for the proliferation of melanoblasts in the mouse embryo. *Dev. Biol.*, **192**, 99–107.
35. Hechsher-Sørensen, J., Hill, R.E. and Lettice, L.A. (1998) Double labelling for wholemount *in situ* hybridization in mouse. *Biotechniques*, **24**, 914.
36. Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucl. Acids Res.*, **16**, 121.