Identification of *Cis*-regulatory Elements in the Mouse *Pax9/Nkx2-9* Genomic Region: Implication for Evolutionary Conserved Synteny

Fabio Santagati,*,[1] Kuniya Abe,[†] Volker Schmidt,[‡] Thomas Schmitt-John,[‡] Misao Suzuki,[§] Ken-ichi Yamamura and Kenji Imai[†,2]**

**GSF-National Research Center for Environment and Health, Institute of Developmental Genetics, D-85764 Neuherberg, Germany, [†]Technology and Development Team for Mammalian Cellular Dynamics, RIKEN Tsukuba Institute, BioResource Center, Ibaraki 305-0074, Japan, [‡]Developmental Biology and Molecular Pathology, University of Bielefeld, 33501 Bielefeld, Germany and [§]Center for Animal Resources and Development and **Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto 862-0976, Japan*

Manuscript received January 23, 2003
Accepted for publication April 28, 2003

ABSTRACT

We previously reported close physical linkage between *Pax9* and *Nkx2-9* in the human, mouse, and pufferfish (*Fugu rubripes*) genomes. In this study, we analyzed *cis*-regulatory elements of the two genes by comparative sequencing in the three species and by transgenesis in the mouse. We identified two regions including conserved noncoding sequences that possessed specific enhancer activities for expression of *Pax9* in the medial nasal process and of *Nkx2-9* in the ventral neural tube. Remarkably, the latter contained the consensus Gli-binding motif. Interestingly, the identified *Pax9 cis*-regulatory sequences were located in an intron of the neighboring gene *Slc25a21*. Close examination of an extended genomic interval around *Pax9* revealed the presence of strong synteny conservation in the human, mouse, and Fugu genomes. We propose such an intersecting organization of *cis*-regulatory sequences in multigenic regions as a possible mechanism that maintains evolutionary conserved synteny.

THE sequencing of genomes has opened new research perspectives for the full interpretation of genomic codes, not only by allowing the comprehensive recognition of protein coding and noncoding transcribed sequences, but also by facilitating the identification of regulatory sequences responsible for the control of correct gene expression in time and space. The identification of evolutionarily conserved noncoding sequences (CNSs) among orthologous genomic regions of different species has been proven to be a powerful guide for the localization of *cis*-acting transcriptional regulatory elements (Koop and Nadeau 1996; Hardison 2000; Wasserman *et al.* 2000). The elucidation of *cis*-regulatory elements in a genome is an important starting point for the identification of the specific binding factors and therefore for the subsequent discovery of the molecular pathways that control the mechanisms of cellular differentiation and physiology.

Pax9 belongs to the paired-box (Pax) transcription factor gene family. The Pax family in vertebrates includes nine members that are further subdivided into four paralogous subgroups. The products of the Pax genes have major roles during embryonic development in the processes of tissue patterning and organogenesis (Dahl *et al.* 1997). One subgroup consists of *Pax9* and its paralog *Pax1*. The two genes share a high sequence homology and show a very widely overlapping expression pattern. *Pax9* is expressed mainly in the sclerotome of the somites, in the pharyngeal pouch endoderm and its derivatives, in the developing limb buds, and in the facial mesenchyme of neural-crest cell origin, including nasal and jaw processes and tooth buds (Neubüser *et al.* 1995). A targeted mutagenesis approach has shown that *Pax9* exerts a fundamental function for the normal development of several facial structures, such as the secondary palate, teeth, and laryngeal cartilage, of the third and fourth pharyngeal pouch derivatives like parathyroid glands, ultimobranchial bodies, and the thymus, and of the limbs (Peters *et al.* 1998). A role of *Pax9* in vertebral column development was observed only in *Pax1;Pax9* double-mutant mice, which show a much more severe phenotype than that of single *Pax1* mutants, while single *Pax9* null mutants do not exhibit vertebral column defects (Peters *et al.* 1999).

No *cis*-regulatory mechanism that controls specific expression of *Pax9* in time and space has been described. In the framework of the search for *cis*-regulatory elements of *Pax9*, we previously described the exon-intron structure of *Pax9* in *Homo sapiens*, *Mus musculus*, and *Fugu rubripes* (Santagati *et al.* 2001). The analysis of the *Pax9* genomic region in the three species demon-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession no. AF267536.

[1]*Present address:* Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, Collège de France, 67404 Illkirch, Strasbourg, France.

[2]*Corresponding author:* GSF-National Research Center for Environment and Health, Institute of Developmental Genetics, Ingolstädter Landstrasse 1, D-85764 Neuherberg, Germany. E-mail: imai@gsf.de

strated a conserved physical association with another transcription factor gene, *Nkx2-9*. *Nkx2-9* codes for a member of the Nk2 transcription factor family, characterized by the Nk2 homeobox. *Nkx2-9* is initially expressed in the ventral half of the developing neural tube of the mouse embryo in the entire axis and, as development progresses, its expression becomes restricted to the brain region (Pabst *et al.* 1998). *Pax9* and *Nkx2-9* lie ~80 kb apart from each other in a head-to-head orientation on human chromosome (HSA) 14q13 and on mouse chromosome (MMU) 12, while they are only 10 kb apart in Fugu.

In this work, we investigated *Nkx2-9* and *Pax9* tissue-specific enhancers by interspecies sequence comparison and by transgenesis. Through an extended genomic analysis and the identification of more associated genes that enlarged the region of the conserved synteny, we propose a model for the regulatory sequence organization of this region and suggest a possible interpretation for the strong synteny conservation with evolutionary considerations.

MATERIALS AND METHODS

Genomic sequence comparison: Alignment and search for homologous regions between large genomic sequences of different species were carried out with the PipMaker program (Schwartz *et al.* 2000) available on the web site at Pennsylvania State University (http://bio.cse.psu.edu/pipmaker). The reference human sequence was masked against repetitive DNA with RepeatMasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html).

Construction of transgenic constructs with CNSs: A 1-kb fragment containing the CNS-6 segment was amplified from bacterial artificial chromosome (BAC) DNA by PCR with primers 5′-CATTTTGCCAGAGGCAGAGG-3′ and 5′-AAGGGAC AGTGAGCGGTCTG-3′ and was cloned in *Sma*I-linearized pASShsp68lacZpA (provided by H. Sasaki). Its original orientation with respect to the mouse *Nkx2-9* promoter was maintained in the construct. A 2.5-kb fragment containing the CNS+2 segment was amplified with primers 5′-GGACCAGG CCTTTGTATAAGGC-3′ and 5′-TGATTGTGACCCCTGGTT TAGC-3′ and cloned as above. The integrity of all constructs was verified by sequencing, and their inserts were excised from the vector sequence by *Sal*I digestion.

Generation of transgenic mice: Transgenic mice were generated by pronucleus injection of linearized transgene constructs into fertilized eggs that were subsequently transferred into the oviducts of pseudopregnant foster females. X-gal staining was performed as described (Peters *et al.* 1998) on founder transgenic embryos. To obtain sectioned views, X-gal-stained embryos were cut by scalpel at levels as indicated (Figure 3, A–D). Yolk sacs were used for DNA preparation for PCR genotyping (primer pairs 5′-AGGTGACACTATAGAAG GATCCG-3′, 5′-GGTGCTAGCTCAACTGGTGG-3′ and 5′-TCT CATGCTGGAGTTCTTCG-3′, 5′-AGCATTTGTATTTCTGAT GCCAAC-3′).

Whole-mount RNA *in situ* hybridization: A mouse *Nkx2-9* probe was amplified by RT-PCR from total RNA of embryonic day 11.5 mouse embryos and cloned in pCRII-TOPO (Invitrogen, San Diego). An antisense probe was transcribed with SP6 RNA polymerase on *Xho*I-linearized template. A sense probe was transcribed with T7 RNA polymerase on *Bam*HI-linearized template. Whole-mount *in situ* hybridization in mouse embryos was performed according to the protocol described by Spörle and Schughart (1998). After rehydration of paraformaldehide-fixed embryos in methanol/PBS, all incubation and washing steps were automatically processed by InsituPro (INTAVIS Bioanalytical Instruments AG, Köln, Germany).

RESULTS

Identification of candidates for *cis*-regulatory elements by comparative sequencing: The human sequence (GenBank accession no. AL079303) consisted of a nearly 200-kb genomic region from 18 kb downstream of *NKX2-9* to 84 kb downstream of *PAX9* (note their head-to-head organization as shown in Figure 1). The mouse sequence was derived from the combination of two overlapping *Pax9*-positive BAC clones, 136M3 (GenBank accession no. AC040982) and 327I21 (GenBank accession no. AC079959) from the RPCI-23 mouse BAC library. The 24.4-kb Fugu sequence (GenBank accession no. AF267536) was determined by sequencing a cosmid clone including *Nkx2-9* and *Pax9*. This Fugu sequence ranges from 1.9 kb downstream of *Nkx2-9* to 6.5 kb downstream of *Pax9*. Consistent with the fact that the Fugu genome is eight times more compact than the human genome (Elgar *et al.* 1996; Koop and Nadeau 1996), this Fugu genomic interval largely overlaps with the 200-kb human reference interval. A comparison between the human and mouse sequences by PipMaker analysis revealed that the entire genomic interval displayed a significant degree of homology, indicating a remarkable conservation even outside of the coding sequences (Figure 1). The BLASTZ alignment of PipMaker analysis found numerous (~2000) ungapped sequence segments that were similar between humans and mice (depicted as dots or bars in Figure 1). Of these conserved sequence segments, we focused on segments that were at least 100 bp long and at least 70% identical, according to the criteria suggested by Loots *et al.* (2000). In this way, we found 100 of the conserved segments as putative CNSs. Of these, 4 and 8 segments were found to correspond to exons of *Nkx2-9* and *Pax9*, respectively, and were excluded from further analyses. A BLAST search was performed for each of the remaining 88 putative CNSs to verify whether they really represented noncoding sequences. Surprisingly, 7 of them lying downstream of *PAX9* matched the cDNA sequence of *SLC25A21*. *SLC25A21* codes for a mitochondrial oxodicarboxylate carrier protein, which is ubiquitously present in the inner mitochondrial membrane and plays a central role in amino acid metabolism (Fiermonte *et al.* 2001). The mouse counterpart of *SLC25A21* (hereafter referred to as *Slc25a21*) lies ~2 kb from the 3′ end of *Pax9* in a tail-to-tail orientation. *SLC25A21/Slc25a21* consists of 10 exons that are distributed over a large genomic region of ~500 kb. These 7 conserved segments that corresponded to *SLC25A21* exon sequences

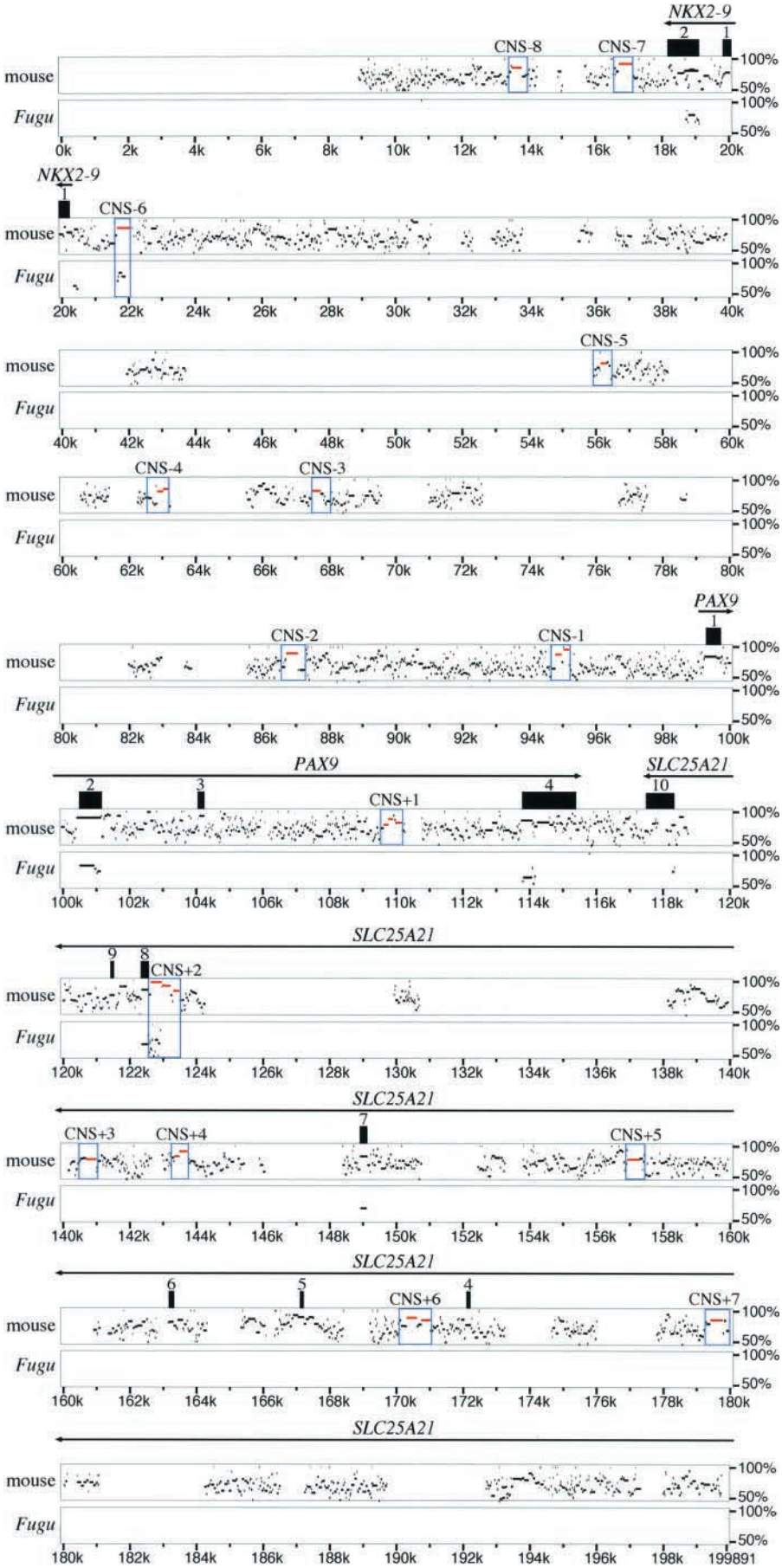

FIGURE 1.—PipMaker analysis of the human/mouse/Fugu sequences. The graph shows the percent identity plot output of ~200 kb of the human genomic sequence including *Nkx2-9*, *Pax9*, and part of *Slc25a21*. The human sequence was used as reference. Arrows above the gene designations indicate the transcriptional direction of the three genes. Homology matches with the mouse and Fugu sequences are represented as dots and dashes, where the length corresponds to that of the matching sequence and the height in the plot to the homology degree (scaling between 50 and 100% on the right side of each row). Boxed segments (CNS-8–CNS+7, outlined in blue) contain one or more of major CNSs (highlighted in red). The CNS segments upstream and downstream of *Pax9* have negative and positive numbers, respectively.

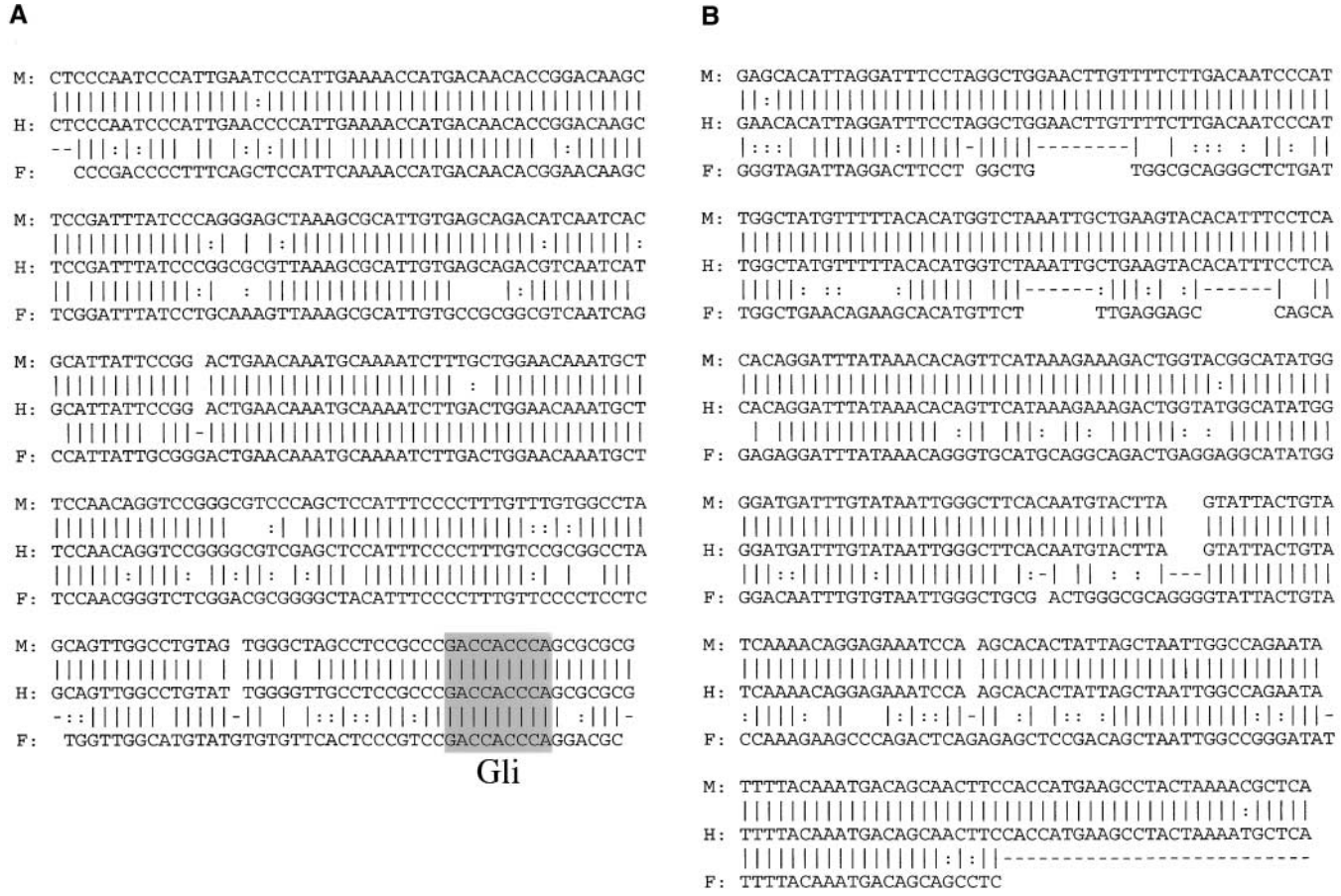

FIGURE 2.—Sequence alignment of CNS-6 (A) and CNS+2 (B). Within the 200-kb interval analyzed by PipMaker, these regions are the only two segments of noncoding sequences that show a significant degree of homology among humans, mice, and Fugu. Vertical bars, double dots, and dashes indicate perfect matches, transitions, and gaps, respectively. In CNS-6, the potential Gli-binding sites with the complete consensus motif are shaded. M, mouse; H, human; and F, Fugu.

were excluded from further investigations. Thus, the rest of 81 sequence segments were regarded as true CNSs. These CNSs often appeared in clusters, and we found 15 ~1-kb genome segments that contained multiple CNSs, including very long ones, which were distributed over the whole region examined (CNS-8–CNS+7; Figure 1).

In contrast, in the alignment of the Fugu and human sequences, only a small number of homologous segments were detected by PipMaker analysis (Figure 1). In the intergenic region between *Pax9* and *Nkx2-9*, sequences within the CNS-6 region also were found conserved in Fugu (Figures 1 and 2A). Downstream of *Pax9*, four distinct hits were identified by PipMaker analysis. Interestingly, three of these conserved segments fell into three exons of *SLC25A21* (exons 7, 8, and 10) in the corresponding human sequence. Although *Slc25a21* in the mouse and in Fugu has not been described, this result provides evidence for the presence of the orthologous gene of *SLC25A21* in the two species. This finding was very intriguing, because it extended the region of conserved synteny among the three vertebrate species, already described for *Pax9* and *Nkx2-9* (SANTAGATI *et al.* 2001), to include an additional gene, *Slc25a21*. The fourth hit was detected inside *Slc25a21* intron 7 in the close vicinity of exon 8 (Figure 2B). This region coincided with the mouse-human CNS+2.

Transient transgenesis with CNS constructs: To test the potential regulatory activity of the selected CNS segments, we performed a transient transgenic assay in founder embryos. The general design of the experiment was to place a test genomic fragment upstream of a *lacZ* reporter cassette with a minimal promoter from mouse *Hsp68*. (KOTHARY *et al.* 1989; SASAKI and HOGAN 1996).

CNS-6 segment drives Nkx2-9 expression in the ventral neural tube: CNS-6 and CNS+2, which are conserved in all three species, were tested by this enhancer assay. CNS-6 is only ~1.5 kb from the 5′ end of *Nkx2-9* and consists of a 422-bp sequence with an overall 85% identity between humans and mice. Within this sequence a 244-bp segment had 80% conservation with the Fugu DNA (Figure 2A). Five independent mouse embryos at E10.5 carried the CNS-6 transgene, and all of them showed *lacZ* expression upon X-gal staining. As compared to endogenous *Nkx2-9* expression (Figure 3, D and H), expression of the CNS-6 transgene in the transgenic embryos varied

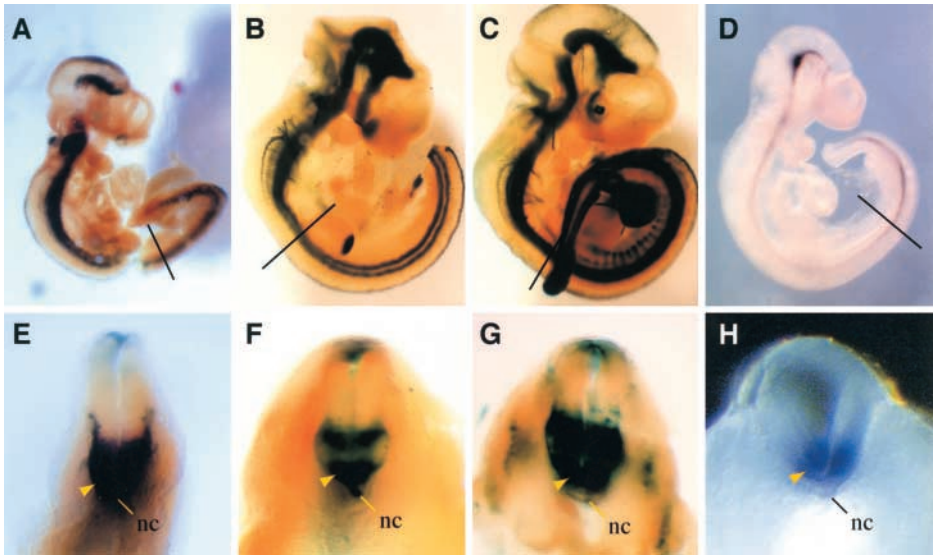

FIGURE 3.—X-gal staining of CNS-6 transgenic embryos. (A–C) Whole-mount overview of transgenic founder embryos at E10.5-E11. *lacZ* expression is clearly visible along the whole neural tube. (D) RNA *in situ* hybridization of *Nkx2-9* at E10.5 as control. E–H show transversal dissections of the whole embryos depicted in A–D, respectively. The section level is indicated in A–D with a line. Note that X-gal staining is consistently seen in the ventral part of the neural tube (yellow arrowheads in E–G). The location of the notochord is indicated (nc in E–H).

in both levels and tissue-specific patterns, including irreproducible ectopic staining in diverse structures (Figure 3, A–C and E–G). Nevertheless, all of the samples shared a common positive domain in the ventral neural tube (Figure 3, E–G). This *lacZ* expression in the ventral neural tube corresponds well to that of endogenous *Nkx2-9* (Figure 3, D and H). Thus, our data strongly suggest that the CNS-6 segment includes *cis*-acting sequences that regulate *Nkx2-9* expression in the ventral neural tube. This *Nkx2-9* expression pattern has been shown to be dependent on Sonic hedgehog (Shh) signaling from the floor plate (PABST *et al.* 2000), which is mediated by Gli transcription factors as the final effectors. Consistent with this model, a fully conserved consensus sequence of nine bases for the Gli-binding site (GACCACCCA), originally found in *Foxa2* (SASAKI *et al.* 1997), was recognized within the CNS-6 sequence in all three species (Figure 2A).

CNS+2 segment possessed the enhancer activity of Pax9 expression in the medial nasal process: The second fragment tested was CNS+2. This segment was located very close to exon 8 of *Slc25a21* inside the preceding intron. The human/mouse homologous region was 594 bp long with an overall identity of 95%, 300 bp of which were homologous to the Fugu sequence with 64% identity (Figure 2B). A 2.2-kb segment of the mouse sequence, including CNS+2 and part of *Slc25a21* exon 8, was used for the transgenic construct. Three transgenic founder embryos at E11.5 were produced with this construct. Besides inconsistent X-gal staining in several domains that did not match with the endogenous *Pax9* expression pattern, a very specific expression was observed in all of the embryos in the ventromedial region of the medial nasal processes (Figure 4, A–C). This domain overlapped with the two stripes of endogenous *Pax9* expression in the same structures (Figure 4D), suggesting that CNS+2 includes *cis*-regulatory elements that drive *Pax9* expression in this domain. However, endogenous *Pax9* expression can also be detected in the more internal medial nasal processes as well as in the lateral nasal process (Figure 4D). Thus the CNS+2 segment does not appear to contain all the elements required for the complete expression of *Pax9* in the nasal processes.

DISCUSSION

Identification of enhancers for expression of *Nkx2-9* and *Pax9*: This work provides the first evidence for the elucidation of *cis*-regulatory elements that control expression of the physically linked but distinctly regulated genes *Pax9* and *Nkx2-9*. The CNS-6 segment directed the transcription of the reporter gene in the ventral part of the neural tube. The analysis of the CNS-6 sequence revealed not only a high degree of sequence conservation among the three species, but also a putative binding site for Gli proteins with a complete match to the consensus sequence identified in *Foxa2* (SASAKI *et al.* 1997). In the 200-kb *Nkx2-9/Pax9* genome interval of the mouse, we found three of the nine-base Gli-binding motif, but the one included in CNS-6 is the only one that is conserved in humans, mice, and Fugu. Whether the identified sequence in CNS-6 functions as a Gli-binding site remains to be confirmed in future studies.

Our finding that CNS+2 is able to direct transgene expression selectively in the oral edge of the medial nasal processes but not in the remaining mesenchymal tissues suggests that each of the *Pax9*-expressing structures is independently regulated. Thus, the regulatory mechanisms that control the whole cranio-facial development may be very complex and diverse, consistent with the composite distribution of signaling molecules and transcription factors involved in facial development (FRANCIS-WEST *et al.* 1998). CNS+2 is located only 8 kb downstream of the last exon of *Pax9*, but remarkably this region is in the seventh intron of the neighboring gene *Slc25a21*. The entire genomic interval that con-

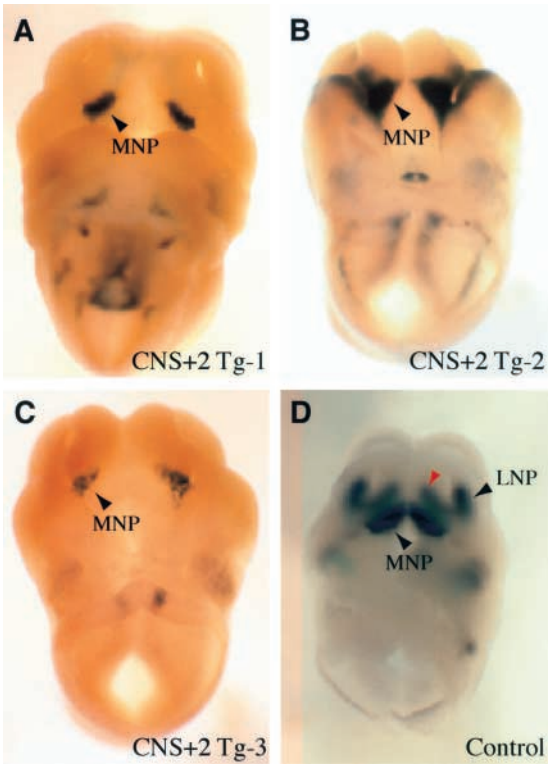

Figure 4.—X-gal staining of CNS+2 transgenic embryos. A–D show reporter (*lacZ*) expression in the ventral part of the facial structures from three independent transgenic founder embryos at E11.5 (A–C) in comparison to endogenous *Pax9* expression detected by RNA *in situ* hybridization at E11.0 (D). (A–C) The reporter expression is consistently observed in the oral edge of the medial nasal processes (MNP) as indicated by black arrowheads. (D) *Pax9* is very strongly expressed in the corresponding part of the MNP. Note that *Pax9* is also expressed in the dorsal part of the MNP (red arrowhead) and in the lateral nasal processes (LNP).

tains all *cis*-regulatory elements of *Pax9* has not been determined. By BAC transgenesis we tested whether a 195-kb BAC clone, including 130 kb upstream and 50 kb downstream of *Pax9*, contained all *cis*-regulatory elements of *Pax9*. Our preliminary result showed that the *Pax9*-BAC transgene could only partially reproduce endogenous expression of *Pax9* (see the supplemental figures available online at http://www.genetics.org/supplemental). This observation suggests that additional *cis*-regulatory elements of *Pax9* are located outside this BAC interval. It should be noted that this BAC clone contained not only *Pax9* but also its flanking genes, *Nkx2-9* and part of *Slc25a21*, on either side. Taken together, our results suggest that *cis*-regulatory elements of *Pax9* are interspersed in a wide genomic interval, even in the regions of neighboring genes. The correct expression of *Pax9* in a particular domain may be determined by the concomitant function of separate control elements, even at a long genomic distance from each other, as exemplified in the case of mouse *Myf5* and *Mrf4* (Carvajal *et al.* 2001). The regulatory elements of both *Myf5* and *Mrf4* trespass the limits of gene boundary and, within the introns of a neighboring gene, localize the protein tyrosine phosphatase-RQ gene. Recently, another remarkable example of a complex organization of genes in mammals has been reported for *Shh* (Lettice *et al.* 2002). The genomic region containing the limb-specific control elements of *Shh* was identified to be ~800 kb away from the gene itself inside the introns of the *Lmbr1* gene in humans and mice. Moreover, at least one more gene has been located between *Shh* and *Lmbr1*, suggesting that these elements can exert their specific function on the target gene despite large genomic distances and the presence of other intervening genes.

Conserved synteny around *Pax9*: Our finding that *Slc25a21* physically links to *Pax9* in the human, mouse, and Fugu genomes has extended the region of conserved synteny in the three species that we described earlier (Santagati *et al.* 2001). Moreover, close examination of the mouse and human genome sequences from GenBank (mouse, NW_000053; human, NT_025892) has revealed that this region of conserved synteny extends even further with at least four additional genes included. In the mouse and human genomes, *Nkx2-1* (or *Titf1*) and *Mbip* (Fukuyama *et al.* 2000) are found on the centromeric side of *Nkx2-9*, and *Mipol1* (Kondoh *et al.* 2002) and *Foxa1* are located on the telomeric side of *Slc25a21*. Thus, the order of the genes in the synteny group is: *Mbip—Nkx2-1—Nkx2-9—Pax9—Slc25a21—Mipol1—Foxa1*. In the Fugu genome (http://www.ensembl.org), five of the seven genes also physically link together: *Mbip*, *Nkx2-1*, *Nkx2-9*, *Pax9*, and *Slc25a21*. Whether Fugu counterparts of *Mipol1* and *Foxa1* are also associated with these linked genes cannot be determined on the basis of the currently available Fugu sequence data.

Interestingly, a similar physical linkage is encountered for mouse and human paralogs of many of these genes. The paralogous genes of *Pax9*, *Nkx2-9*, and *Nkx2-1* are *Pax1*, *Nkx2-2*, and *Nkx2-4*, respectively, and they are known to physically link in MMU2 and in HSA20p11 (Wang *et al.* 2000; Santagati *et al.* 2001). By close examination of genome sequence data (mouse, NW_000178; human, NT_011387), we found that *Foxa2*, the paralog of *Foxa1*, also fell into this synteny group. Furthermore, within the human genome sequence between *PAX1* and *FOXA2*, a computer-predicted gene with homology to *SLC 25A21* (named *LOC200259* or *SLC21A5L*) was present. Thus, the locus order in this synteny group is: *Nkx2-4—Nkx2-2—Pax1—Slc25a5l—Foxa2*. Similarly, in Fugu, we could confirm the physical association of *Nkx2-4*, *Nkx2-2*, and *Pax1* in the genome annotation data (http://www.ensembl.org). Together, gene order in the two paralogous genome segments including *Pax9* or *Pax1* is highly conserved in the human, mouse, and Fugu genomes.

In lower chordates, only one member for each paralogous gene pair is present, strongly suggesting that the genomic regions evolved through duplication of a common ancestral genome (Figure 5). This genomic region

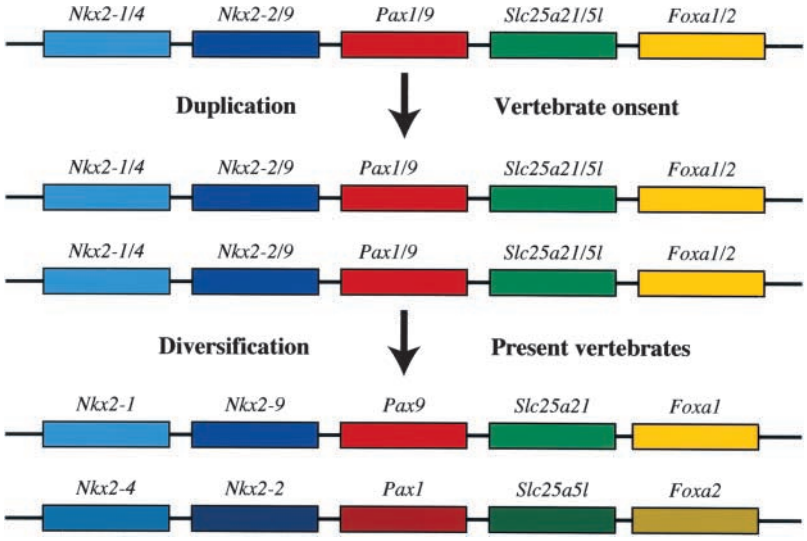

FIGURE 5.—A model of the origin and conservation of the syntenic regions including *Nkx2-2/9* and *Pax1/9* paralogous gene pairs. From the conserved synteny of the *Pax1/Pax9* genomic regions and the genetic information about lower chordates such as amphioxus, it can be assumed that in the ancestral genome of chordate progenitors a set of at least five genes (ancestors of *Nkx2-1/4*, *Nkx2-2/9*, *Pax1/9*, *Slc25a21/5l*, and *Foxa1/2*) were tightly associated. A series of genome duplication events brought about the evolutionary burst that caused the origin of vertebrates. The *Pax1/9* syntenic region also underwent duplication, originating two sets of paralogous genes that afterward independently diversified, acquiring distinct functions up to the present time. The presence of interspersed regulatory *cis*-elements throughout this genomic region may have been working as the driving force that has kept these genes tightly associated through evolution.

surrounding the *Pax1/9* gene may include at least four other linked genes that have maintained their physical association from the primitive situation predating the locus duplication up to the present time. Following this duplication event, these paralogous gene pairs diversified their roles, contributing to the variety of functionalities in vertebrates (Figure 5). To our best knowledge, such a synteny conservation between mammals and fish involving five to seven genes maintaining the same gene order has never been described in detail before. Moreover, the comparison between two paralogous genomic segments including *Pax9* or *Pax1* allowed us to speculate on the organization of a putative ancestral genome. This physical association might reflect a fixed multigenic transcriptional domain whose members (at least the five genes in Figure 5) cannot be separated without compromising their normal function. The nature of this functional bond has still to be investigated, but the strong indication from our study that at least some of the *Pax9* regulatory elements are interspersed or interdigitated over a widespread multigenic region can already account for a decisive factor to maintain their physical linkage. As mentioned above, the tight physical linkage between *Myf5* and *Mrf4* is likewise conserved in all the vertebrate species so far analyzed (CARVAJAL *et al.* 2001). It is logical to deduce that this kind of genomic organization of genes and their regulatory elements represents an irresolvable constraint for neighboring genes.

In conclusion, interpretation of our results in light of similar examples in the literature leads to some new general considerations. The presence of intersecting regulatory sequences in multigenic genomic regions conceivably represents a key point in genome evolution. The onset and fixation of regulatory elements inside the territory of neighboring genes constitutes a functional bond resulting in the physical association between the genes. These associations in time might have extended, involving entire blocks of genes. Thus, it logically follows that the plasticity of the genome in the events of shuffling and reorganization during evolution has been inevitably limited to some extent. The rearrangement units may not have been single genes but the blocks of physically linked genes. Hence, in the future, the search for *cis*-regulatory elements not only will lead to a better understanding of the molecular mechanisms that control gene expression, but also will provide insights into genome evolution.

We acknowledge H. Sasaki, P. A. Iannou, K. Araki, and F. Stewart for providing plasmid constructs as indicated in the text. We thank B. Rey, A. Krause, and G. Michel for their contribution to the generation of transgenic mice; M. Wahl for providing informatics tools for genomic sequence analyses; and R. Balling, L. Bally-Cuif, and J. Favor for critical reading of the manuscript and helpful comments. This work was supported in part by the Deutsche Forschungsgemeinschaft and the GSF-Forschungszentrum and in part by the Ministry of Education, Culture, Sports, Science and Technology of Japan, with Special Coordinating Funds for Promoting Science and Technology (K.A.).

LITERATURE CITED

CARVAJAL, J. J., D. COX, D. SUMMERBELL and P. W. RIGBY, 2001 A BAC transgenic analysis of the *Mrf4/Myf5* locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development. Development **128:** 1857–1868.

DAHL, E., H. KOSEKI and R. BALLING, 1997 Pax genes and organogenesis. Bioessays **19:** 755–765.

ELGAR, G., R. SANDFORD, S. APARICIO, A. MACRAE, B. VENKATESH *et al.*, 1996 Small is beautiful: comparative genomics with the pufferfish (*Fugu rubripes*). Trends Genet. **12:** 145–150.

FIERMONTE, G., V. DOLCE, L. PALMIERI, M. VENTURA, M. J. RUNSWICK *et al.*, 2001 Identification of the human mitochondrial oxodicarboxylate carrier. Bacterial expression, reconstitution, functional characterization, tissue distribution, and chromosomal location. J. Biol. Chem. **276:** 8225–8230.

FRANCIS-WEST, P., R. LADHER, A. BARLOW and A. GRAVESON, 1998 Signalling interactions during facial development. Mech. Dev. **75:** 3–28.

FUKUYAMA, K., M. YOSHIDA, A. YAMASHITA, T. DEYAMA, M. BABA *et al.*, 2000 MAPK upstream kinase (MUK)-binding inhibitory protein, a negative regulator of MUK/dual leucine zipper-bear-

ing kinase/leucine zipper protein kinase. J. Biol. Chem. **275:** 21247–21254.

Hardison, R. C., 2000 Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet. **16:** 369–372.

Kondoh, S., H. Sugawara, N. Harada, N. Matsumoto, H. Ohashi *et al.*, 2002 A novel gene is disrupted at a 14q13 breakpoint of t(2;14) in a patient with mirror-image polydactyly of hands and feet. J. Hum. Genet. **47:** 136–139.

Koop, B. F., and J. H. Nadeau, 1996 Pufferfish and new paradigm for comparative genome analysis. Proc. Natl. Acad. Sci. USA **93:** 1363–1365.

Kothary, R., S. Clapoff, S. Darling, M. D. Perry, L. A. Moran *et al.*, 1989 Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. Development **105:** 707–714.

Lettice, L. A., T. Horikoshi, S. J. Heaney, M. J. Van Baren, H. C. Van Der Linde *et al.*, 2002 Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. USA **99:** 7548–7553.

Loots, G. G., R. M. Locksley, C. M. Blankespoor, Z. E. Wang, W. Miller *et al.*, 2000 Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science **288:** 136–140.

Neubüser, A., H. Koseki and R. Balling, 1995 Characterization and developmental expression of Pax9, a paired-box-containing gene related to Pax1. Dev. Biol. **170:** 701–716.

Pabst, O., H. Herbrand and H. H. Arnold, 1998 Nkx2–9 is a novel homeobox transcription factor which demarcates ventral domains in the developing mouse CNS. Mech. Dev. **73:** 85–93.

Pabst, O., H. Herbrand, N. Takuma and H. H. Arnold, 2000 NKX2 gene expression in neuroectoderm but not in mesendodermally derived structures depends on sonic hedgehog in mouse embryos. Dev. Genes Evol. **210:** 47–50.

Peters, H., A. Neubüser, K. Kratochwil and R. Balling, 1998 *Pax9*-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. Genes Dev. **12:** 2735–2747.

Peters, H., B. Wilm, N. Sakai, K. Imai, R. Maas *et al.*, 1999 Pax1 and Pax9 synergistically regulate vertebral column development. Development **126:** 5399–5408.

Santagati, F., J. K. Gerber, J. H. Blusch, C. Kokubu, H. Peters *et al.*, 2001 Comparative analysis of the genomic organization of *Pax9* and its conserved physical association with *Nkx2–9* in the human, mouse, and pufferfish genomes. Mamm. Genome **12:** 232–237.

Sasaki, H., and B. L. Hogan, 1996 Enhancer analysis of the mouse HNF-3 beta gene: regulatory elements for node/notochord and floor plate are independent and consist of multiple sub-elements. Genes Cells **1:** 59–72.

Sasaki, H., C. Hui, M. Nakafuku and H. Kondoh, 1997 A binding site for Gli proteins is essential for HNF-3beta floor plate enhancer activity in transgenics and can respond to Shh in vitro. Development **124:** 1313–1322.

Schwartz, S., Z. Zhang, K. A. Frazer, A. Smit, C. Riemer *et al.*, 2000 PipMaker—a web server for aligning two genomic DNA sequences. Genome Res. **10:** 577–586.

Spörle, R., and K. Schughart, 1998 Paradox segmentation along inter- and intrasomitic borderlines is followed by dysmorphology of the axial skeleton in the *open brain* (*opb*) mouse mutant. Dev. Genet. **22:** 359–373.

Wang, C., C. T. Brodnicki, N. G. Copeland, N. A. Jenkins and R. P. Harvey, 2000 Conserved linkage of NK-2 homeobox gene pairs Nkx2–2/2–4 and Nkx2–1/2–9 in mammals. Mamm. Genome **11:** 466–468.

Wasserman, W. W., M. Palumbo, W. Thompson, J. W. Fickett and C. E. Lawrence, 2000 Human-mouse genome comparisons to locate regulatory sites. Nat. Genet. **26:** 225–228.

Communicating editor: C. Kozak