EVOLUTION & DEVELOPMENT 13:5, 448–459 (2011)
DOI: 10.1111/j.1525-142X.2011.00502.x

Co-orthology of *Pax4* and *Pax6* to the fly *eyeless* gene: molecular phylogenetic, comparative genomic, and embryological analyses

Tereza Manousaki,[a,b,1] Nathalie Feiner,[a,c,1] Gerrit Begemann,[a] Axel Meyer,[a,b,c] and Shigehiro Kuraku[a,b,c,*]

[a] Laboratory for Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, Universitätsstrasse 10, 78464 Konstanz, Germany
[b] Konstanz Research School Chemical Biology (KoRS-CB), University of Konstanz, Universitätsstrasse 10, 78464 Konstanz, Germany
[c] International Max-Planck Research School (IMPRS) for Organismal Biology, University of Konstanz, Universitätsstrasse 10, 78464 Konstanz, Germany
*Author for correspondence (email: shigehiro.kuraku@uni-konstanz.de)
[1] Contributed equally to this work.

SUMMARY The functional equivalence of *Pax6/eyeless* genes across distantly related animal phyla has been one of central findings on which evo-devo studies is based. In this study, we show that *Pax4*, in addition to *Pax6*, is a vertebrate ortholog of the fly *eyeless* gene (and its duplicate, *twin of eyeless* [*toy*] gene, unique to Insecta). Molecular phylogenetic trees published to date placed the *Pax4* gene outside the *Pax6*/*eyeless* subgroup as if the *Pax4* gene originated from a gene duplication before the origin of bilaterians. However, *Pax4* genes had only been reported for mammals. Our molecular phylogenetic analysis, including previously unidentified teleost fish *pax4* genes, equally supported two scenarios: one with the *Pax4–Pax6* duplication early in vertebrate evolution and the other with this duplication before the bilaterian radiation. We then investigated gene compositions in the genomic regions containing *Pax4* and *Pax6*, and identified (1) conserved synteny between these two regions, suggesting that the *Pax4–Pax6* split was caused by a large-scale duplication and (2) its timing within early vertebrate evolution based on the duplication timing of the members of neighboring gene families. Our results are consistent with the so-called two-round genome duplications in early vertebrates. Overall, the *Pax6*/*eyeless* ortholog is merely part of a 2:2 orthology relationship between vertebrates (with *Pax4* and *Pax6*) and the fly (with *eyeless* and *toy*). In this context, evolution of transcriptional regulation associated with the *Pax4–Pax6* split is also discussed in light of the zebrafish *pax4* expression pattern that is analyzed here for the first time.

INTRODUCTION

Members of the *Pax* (paired box) gene family encode transcription factors that play crucial roles in development (Wehr and Gruss 1996). A milestone in the 1990s that promoted subsequent intensive studies on *Pax* genes was the ability of the *Drosophila melanogaster eyeless* gene as well as its mouse ortholog *Pax6* to induce eye formation when expressed ectopically in flies (Halder et al. 1995). *Pax6*/*eyeless* genes have thus been recognized as the master control gene for eye development (Gehring and Ikeo 1999). A recent report on secondary changes in the insect lineage shed light on a divergent aspect of the *Pax6*/*eyeless* orthology (Lynch and Wagner 2011). The aim of this article is to investigate possible changes in the gene repertoire and gene regulation in the chordate lineage.

Traditionally, nonphylogenetic classifications have grouped *Pax4* with *Pax6* because of the absence of a conserved octapeptide in both of them (Wehr and Gruss 1996). The other vertebrate *Pax* genes are divided into the classes Pax1/9, Pax3/7, and Pax2/5/8 depending on the completeness of the homeodomain (Chi and Epstein 2002). Recent studies suggested that the first wave of the diversification of the *Pax* gene family dates back to the early metazoan era (Matus et al. 2007). The second wave of the diversification of *Pax* genes later in the vertebrate lineage is marked by gene duplications between *Pax2*, *-5*, and *-8* (Kozmik et al. 1999; Bassham et al. 2008; Goode and Elgar 2009), between *Pax1* and *-9* (Holland et al. 1995; Ogasawara et al. 1999; Mise et al. 2008), and between *Pax3* and *-7* (Holland et al. 1999). These gene duplications occurred after invertebrate chordates branched off, but most likely before the split between gnathostomes and cyclostomes (McCauley and Bronner-Fraser 2002; O'Neill et al. 2007). This timing matches that of so-called two-round whole genome duplications (2R-WGDs; Lundin 1993; Holland et al. 1994; Sidow 1996; Spring 1997) implicated in early vertebrate evolution (Kuraku et al. 2009; reviewed in Panopoulou and Poustka

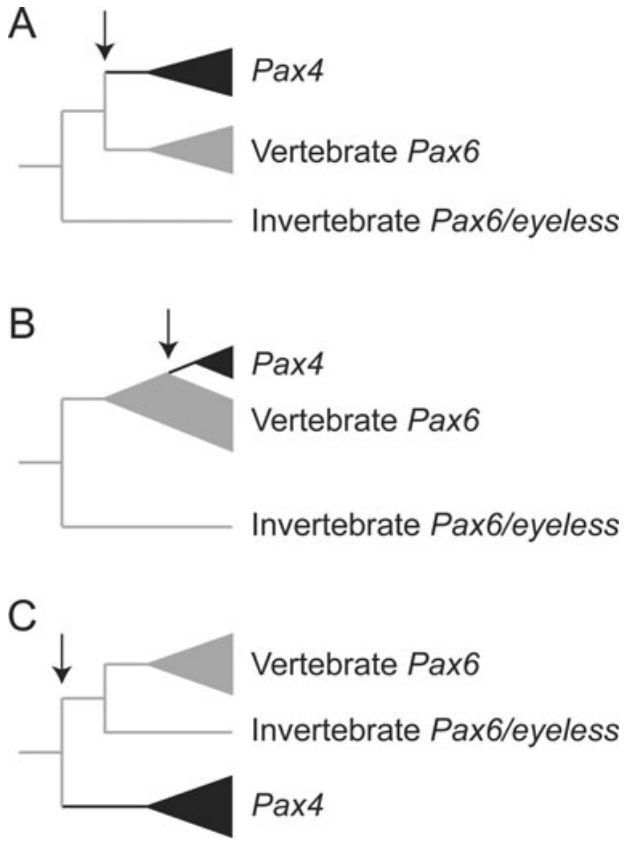

Fig. 1. Three possible scenarios of the timing of gene duplication between *Pax4* and *Pax6*. Arrows indicate the *Pax4–Pax6* split. (A) The *Pax4–Pax6* duplication took place in the vertebrate lineage, and both *Pax4* and *Pax6* are orthologous to invertebrate *Pax6/eyeless*genes. Inside other *Pax* classes, namely *Pax1/9*, *Pax3/7*, and *Pax2/5/8*, paralogs that share the same structural property were also duplicated at this timing (see Introduction). This scenario, however, has never been suggested by molecular phylogenetic analysis. (B) *Pax4* originated in a relatively recent gene duplication from mammalian *Pax6*. This scenario has been previously supported by the presence of *Pax4* genes only in mammals. (C) The *Pax4–Pax6* duplication predates the deuterostome–protostome split. Family-wide phylogenetic analyses usually support this scenario (see Introduction). However, no nonmammalian and invertebrate orthologs of *Pax4* have been reported.

2005). However, it has not been explored, in the modern framework of molecular phylogenetics and comparative genomics, whether the *Pax4–Pax6* split also coincided with this second wave of diversification (Fig. 1A).

The timing of the gene duplication has significant impacts on our understanding of evolutionary modification of gene repertoires and functions. In fact, *Pax4* genes have been reported only for human (Pilz et al. 1993), mouse (Sosa-Pineda et al. 1997), and rat (Tokuyama et al. 1998), suggesting that *Pax4* originated from a gene duplication unique to the mammalian lineage (Fig. 1B). However, family-wide phylogenetic analyses performed to date usually suggested an ancient origin of the *Pax4* gene early in metazoan evolution (Fig. 1C; Hoshiyama et al. 1998; Wada et al. 1998; Breitling and Gerber 2000). In these studies, invertebrate genes identified as *Pax6* orthologs, such as fly *eyeless* (Bopp et al. 1986) and *Caenorhabditis elegans vab-3* (Chisholm and Horvitz 1995; Zhang and Emmons 1995), were shown to be more closely related to vertebrate *Pax6* genes than to *Pax4* genes (Fig. 1C). Because critical phylogenetic signals may be obscured by divergent sequences from other *Pax* classes, the long-standing question regarding the timing of the *Pax4–Pax6* split should be addressed using a focused dataset aiming to resolve the *Pax4–Pax6* relationship.

Gene duplications are usually followed by interplay between duplicates in terms of their functional differentiation. Thus, a comparison of the regulation and functions of duplicates can also lead to better understanding of gene family evolution. In mammals, in addition to the aforementioned inductive role in eye development, *Pax6* is involved in development of the central nervous system (CNS), including the fore- and hindbrain, the neural tube, the pituitary, and the nasal epithelium (Walther and Gruss 1991). In mouse, *Pax6* is also expressed in all the four cell types (α, β, δ, and γ) in the islets of Langerhans, the endocrine part of the pancreas (St-Onge et al. 1997). In zebrafish, a composite expression pattern of *pax6a* and *pax6b* highly resembles that of its mouse ortholog (Kleinjan et al. 2008; also see Kinkel and Prince 2009 for a review on zebrafish pancreas development).

In contrast, *Pax4*, identified only in mammals, has not been implicated in eye development, but is rather expressed in the retinal photoreceptor cells (Rath et al. 2009a). *Pax4* is also expressed mainly in the β-cells of the pancreas, and is necessary for the differentiation of both β- and δ-cell lineages (Sosa-Pineda et al. 1997). A recent study revealed plasticity for pancreatic α-cells to transdifferentiate into β-cells (Thorel et al. 2010). Importantly, *Pax4* can trigger this transdifferentiation (Collombat et al. 2009; also see Liu and Habener 2009). This aspect of the *Pax4* function attracts attentions as a potential clinical target of diabetes therapy (Gonez and Knight 2010). It would be intriguing to reveal possible alterations or conservation in regulation of *Pax4* expressions during evolution in order to reveal the evolutionary history of partitioned or redundant roles between *Pax4* and *Pax6* genes. However, a thorough comparative picture has been obscured by the lack of our knowledge about nonmammalian *Pax4* orthologs.

In this study, we characterized the previously unidentified nonmammalian *Pax4* orthologs in teleost fish genomes and performed combinatorial analyses on molecular phylogeny, conserved synteny, and gene expression patterns. Our analysis favors a scenario that postulates the duplication between *Pax4* and *Pax6* genes in the 2R-WGDs (Fig. 1A). In light of this evolutionary scheme, we conclude that *Pax4*

secondarily lost its expression in the CNS after the 2R-WGD early in vertebrate evolution. This could have led to the highly asymmetric evolution between *Pax4* and *Pax6*.

MATERIALS AND METHODS

Reverse transcription polymerase chain reaction (RT-PCR)

Total RNA was extracted from a whole 52 hpf zebrafish embryo. The RNA was reverse transcribed into cDNA with SuperScript III (Invitrogen) using a 3′ RACE System (Invitrogen, Karlsruhe, Germany). This cDNA was used as template in the following 3′ RACE PCR. The first reaction was performed using the forward primer 5′-GACTGAGGGAATGAGA CCAT-3′, and the product of this PCR was used as template for the nested PCR with the forward primer 5′-CGCAGA GGAGACAAACCTTT-3′. These primers were designed based on zebrafish transcript sequences in Ensembl (ENSDART00000027919 and ENSDART00000078690). The middle fragment was amplified using the forward primer 5′-ATGATTGAGCTGGCGACTGA-3′ and the reverse primer 5′-TCAAACTTTCGCTCCCTCCT-3′ in the first PCR and the forward primer 5′-GACTGAGGGAATGAGACCAT-3′ and the reverse primer 5′-CCTCATCCTCGCTCTTG ATA-3′ in a nested PCR. The upstream fragment (covering the start codon) was amplified using the forward primer 5′-TTTCTAGGATGTTCAGCC-3′ and the reverse primer 5′-CTCTTGTGCTGAACTATG-3′ in the first PCR and the forward primer 5′-CAGCCAATTCTGCATGTA-3′ and the reverse primer 5′-TGATGGAGATGACTTCAG-3′ in a nested PCR. We concatenated the sequences of these three fragments into one with the full-length open reading frame (ORF) and deposited it in EMBL under the accession number FR727738.

For in situ hybridization to detect zebrafish *pax6b* transcripts, a fragment covering its 3′-end was isolated with 3′ RACE using the forward primer 5′-GTTTCACTG TTTTGCTCG-3′ in the first PCR, and the forward primer 5′-ACAGGACAACGGTGGTGAAAA-3′ in the nested PCR.

In situ hybridization

Two zebrafish *pax4* riboprobes were prepared separately using the middle and 3′ cDNA fragments described above. Whole-mount in situ hybridization using the *pax4* riboprobes labeled with digoxigenin (DIG)-UTP and the *pax6b* riboprobes labeled with Fluorescein (Roche Applied Science, Mannheim, Germany) was performed as previously described (Begemann et al. 2001). Hybridization was detected with alkaline phosphatase (AP)-conjugated anti-DIG antibody (Roche Applied Science) followed by incubation with nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl-phosphate (NBT/BCIP) for *pax4*, and with AP-conjugated anti-Fluorescein antibody (Roche Applied Science) followed by p-Iodonitrotetrazolium (INT)/BCIP-based detection for *pax6b*. In double in situ staining, *pax6b* transcripts were detected first, and after a washing step in 0.1 M glycine (pH 2.2), *pax4* transcripts were detected.

Fluorescent in situ hybridization was performed using the tyramide signal amplification (TSA) system (Invitrogen) as instructed by the manufacturer. DIG-labeled riboprobe was detected with horseradish peroxidase-conjugated anti-DIG antibody. After incubating with biotinyl-tyramide, fluorescent signal was detected with streptavidin-488 (Invitrogen).

Retrieval of sequences

Sequences for members of the *Pax* gene family were retrieved from the Ensembl genome database (version 58; Hubbard et al. 2009) and NCBI protein database by performing Blastp searches (Altschul et al. 1997) using mammalian Pax4 and Pax6 peptide sequences as queries. The zebrafish *pax4* sequence was curated by aligning the cDNA sequence we isolated in this study with the zebrafish genome assembly Zv8 (Fig. S1).

Molecular phylogenetic analysis

An optimal multiple alignment of 54 collected amino acid sequences (see Table S1) was constructed with the program MAFFT (Katoh et al. 2005). In tree inferences, we used amino acid residues unambiguously aligned with no gaps, which cover both paired domain and homeodomain. Optimal amino acid substitution models were selected by ProtTest (Abascal et al. 2005). The phylogenetic tree inference with the first dataset used the $LG + I + \Gamma_4$ model, whereas the inference with the second dataset (see below) used the $JTT + \Gamma_4$ model. Heuristic tree searches with the maximum-likelihood (ML) method were performed in PhyML (Guindon and Gascuel 2003) with 100 bootstrap re-samplings.

Exhaustive tree searches with the ML method were performed using Tree-Puzzle (Schmidt et al. 2002), where we input all 10,395 possible tree topologies consisting of eight operational taxonomic units (OTUs), namely (1) mammalian *Pax4*, (2) teleost *Pax4*, (3) gnathostome (jawed vertebrate) *Pax6*, (4) lamprey *Pax6*, (5) amphioxus *Pax6*, (6) tunicate *Pax6*, (7) protostome *Pax6/eyeless* orthologs (including *eyeless* and *twin of eyeless* [*toy*]), and (8) outgroup (putative *Nematostella vectensis Pax6* ortholog, *Ciona Pax3/7*, fly *paired*, human *Pax3*, and human *Pax7*) (for species names and accession IDs, see Table S1). Relationships within these individual OTUs were constrained according to generally accepted species phylogeny (Meyer and Zardoya 2003; Cracraft and Donoghue 2004; Tsagkogeorga et al. 2009; Philippe et al. 2005a; Wiegmann et al. 2009). To provide support values, we performed bootstrapping with 100 re-samplings by running Tree-Puzzle. Statistical tests to evaluate alternative tree topologies were performed using CONSEL (Shimodaira and Hasegawa 2001). Bayesian inferences were performed in MrBayes (Huelsenbeck and Ronquist 2001), where we ran 10,000,000 generations, sampled every 100 generations and excluded 25% of the sample as burnin.

Identification of conserved synteny

Via the BioMart interface, we downloaded a list of Ensembl IDs of 47 genes harbored in the genomic region spanning 20 Mb

both upstream and downstream of *Pax6* gene in human, together with IDs of paralogs of those genes. Our selection of genes in the *Pax6*-containing region that also had a paralog on chromosome 7 in a distance of 20 Mb up- and downstream of *Pax4* resulted in eight cases. For each of these eight cases, we collected homologous sequences in the Ensembl and NCBI protein databases, and inferred a molecular phylogenetic tree as described above (Fig. S5).

Survey of potential *cis*-regulatory elements

To identify conserved noncoding elements (CNEs) shared between *Pax4* and *Pax6*, we used two approaches. First, we aligned the genomic regions containing the two genes using mVISTA (Frazer et al. 2004; http://genome.lbl.gov/vista/) under the default conservation parameters (70% identity for 100 bp of alignment length). In the alignment, we included a number of vertebrate species, including human, mouse, cow, opossum, platypus, chicken, *Xenopus laevis*, and zebrafish. Second, we implemented an analysis to detect local similarity in noncoding regions that is obscured by translocation and inversion of *cis*-regulatory elements. We extracted the intronic as well as the intergenic sequences until the next genes or within a length of 200 kb surrounding the two genes on the human chromosomes. To detect local similarities between the two nonexonic regions, one of the sequences was used as a query in a Blastn search against the other.

To detect CNEs shared between *Pax4*-containing genomic regions of different species, we retrieved genomic sequences covering *Pax4* locus with 10 kb flanking sequences on both ends. When the next gene was located closer than 10 kb, only the intergenic region until the next gene was retrieved. Those sequences were compared in mVISTA. We also referred to VISTA Enhancer Browser containing experimentally validated noncoding fragments with transcriptional enhancer activity (Visel et al. 2007; http://enhancer.lbl.gov/), only to find that there is no *Pax4*-associated enhancer registered in this database.

RESULTS

Identification of teleost fish *Pax4* genes

As a result of Blastp searches using mammalian *Pax4* sequences, we identified Ensembl peptide sequences in the five teleost fish species with sequenced genomes that show higher similarity to *Pax4* than to *Pax6*. Of these, in Ensembl database, only the zebrafish ones (ENSDARP00000013792 based on the Ensembl gene ENSDARG00000021336 and ENSDARP00000073151 based on the gene ENSDARG00000056224) were not annotated as *pax4*. As in zebrafish, two peptides similar to *pax4* derived from two genes annotated separately were found in *Tetraodon nigroviridis* (ENSTNIG00000000660 and ENSTNIG00000011020).

We isolated cDNA fragments of zebrafish *pax4* by means of RT-PCR and compared a resultant concatenated cDNA sequence with those in Ensembl. Our sequence matched both of the two zebrafish Ensembl entries, suggesting that these two were split because of a misidentification of the ORF of a single *pax4* gene. We then aligned these sequences with the corresponding region in the genome assembly Zv8, and identified a putative full-length protein-coding sequence (Fig. S1). In this comparison, a presence of an exceptional splice donor site ("GC" instead of "GT") was revealed (Fig. S1), and this was confirmed with our genomic PCR (data not shown). Using its deduced amino acid sequence based on the curated zebrafish *pax4* ORF, we performed tBlastn searches in the genome assembly of other teleost fishes in Ensembl, and identified their putative pax4 peptide sequences (Fig. S2). Because the two aforementioned *Tetraodon* sequences do not share a region homologous to each other and are intervened by only a 66-bp stretch in the genome assembly, it is likely that they were also split because of a possibly wrong annotation of the ORF in the Ensembl database. Overall, in the five teleost fish species with sequenced genomes, we did not find any sequence that would represent the second *pax4* paralog derived from the teleost-specific genome duplication (TSGD; Kuraku and Meyer 2009).

Sequence alignment containing the five teleost *pax4* genes, other members of the *Pax4/6* class, and human paralogs revealed a high level of conservation in the paired domain and in the homeodomain (Fig. S2). Many of the amino acid residues conserved between *Pax6* sequences and their invertebrate orthologs were revealed to be altered in *Pax4* sequences (Fig. S2).

Expression analysis of zebrafish *pax4*

Expression patterns of zebrafish *pax4* were investigated by in situ hybridization for embryos spanning from 6 h post fertilization (hpf) to 5 days post fertilization (dpf). Identical expression patterns were observed with both probes (see Materials and Methods).

The earliest signals were detected in the developing pancreas at 13 hpf (Fig. 2A), where expression persisted until 30 hpf. The strongest expression was seen around 24 hpf (Fig. 2, B, C, E, and F). To examine the relative localization of the pancreatic expression signals of *pax4* to that of *pax6b*, a marker of early pancreatic endocrine cell development (Biemar et al. 2001), we conducted a double staining of these two genes in 24 hpf zebrafish embryos. We observed partial overlap of *pax4* and *pax6b* expressions (Fig. 2F). Expression of *pax4* was nested in the *pax6b*-expressing domain in the endocrine part of the developing pancreas (Fig. 2, D–F).

Expression of *pax4* in the stomodeum was detected from 57 to 96 hpf (Fig. 2, G–I and not shown). Between 57 and 72 hpf, the expression domain was strongest in the ventrolateral corners of the oral cavity and surrounds the future

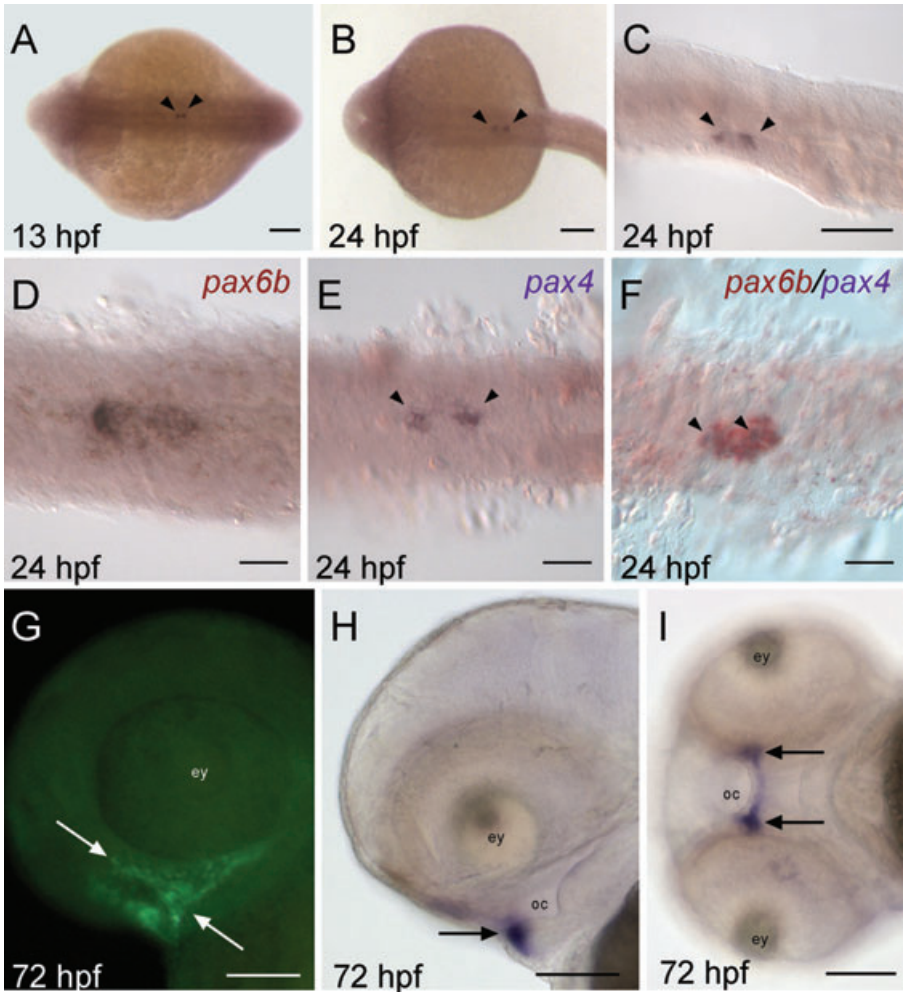

Fig. 2. Expression patterns of *pax4* in zebrafish embryos. All pictures except D (*pax6b*) and F (double staining of *pax4* in blue and *pax6b* in red) show expression of zebrafish *pax4*. The *pax4* riboprobe was synthesized with the 3′ cDNA fragment (see Materials and Methods). Expression of *pax4* in the pancreas is indicated by arrowheads (A–C, E, and F). (A, B) Dorsal views showing expression signals in the developing pancreas at 13 hpf (A) and 24 hpf (B). (C) A lateral view of the expression domain in the pancreas in a 24 hpf embryo. (D–F) Ventral views of *pax6b* (D) and *pax4* (E) and double staining of *pax6b* (red) and *pax4* (purple) (F) in pancreatic tissue of 24 hpf embryos. (G) Fluorescent expression signal in the developing stomodeum (arrows) in a lateral view of a 72 hpf embryo. (H, I) A lateral view of the *pax4* expression in the stomodeum at 72 hpf and a ventral view of the same embryo (arrows). Abbreviations: ey, eye; oc, oral cavity. Scale bars: 100 μm in A–C and G–I; 50 μm in D–F.

mouth (Fig. 2, G–I). More precisely, the signal in the region of the future lip was restricted to mesectodermal layers of the bilaminar stomodeum. The fluorescent in situ hybridization staining with the TSA system additionally showed that the signal in the 72 hpf embryo is not restricted to the outer region of the stomodeum, but elongates into the oral cavity along the pharynx (Fig. 2G). At 96 hpf, *pax4* expression was detected exclusively in the outer surface of the stomodeum, corresponding to the future lip (data not shown).

Survey of *Pax4* orthologs in nonmodel species

To search for *Pax4* orthologs outside the mammalian and teleost lineages, tBlastn searches were performed online using the human Pax4 peptide sequence as a query. First, we performed a search in NCBI dbEST and nr/nt databases of all vertebrates, specifying "Craniata" (taxon ID: 89593 in NCBI Taxonomy) while excluding mammalian (taxon ID: 40674) and teleost sequences (taxon ID: 32443)–note that the taxon "Craniata" adopted in NCBI Taxonomy is incompatible with molecular phylogenetic evidence supporting monophyly of cyclostomes (reviewed in Kuraku 2008). Second, we performed tBlastn searches against nucleotide genomic sequences of species included in Ensembl genome browser (http://www.ensembl.org). These searches resulted in no *Pax4* sequences in all available vertebrate species outside Teleostei and Mammalia, such as *X. tropicalis*, chicken, zebra finch, and anole lizard. Similarly, invertebrate species were revealed to have no other *Pax4/6* sequences other than those already recognized as *Pax6* orthologs.

Our additional search in Mammalia detected *Pax4* orthologs in noneutherians (platypus, ENSOANG-00000000819; opossum, ENSMODG00000015218) and early branching eutherians (two-toed sloth, ENSCHOG-00000009265; African elephant, ENSLAFG00000005297, and rock hyrax ENSPCAG00000016257). Overall, our effort to find additional *Pax4* orthologs, substantiated by available whole genome sequences, strongly suggested the restricted phylogenetic distribution of *Pax4* orthologs to Mammalia

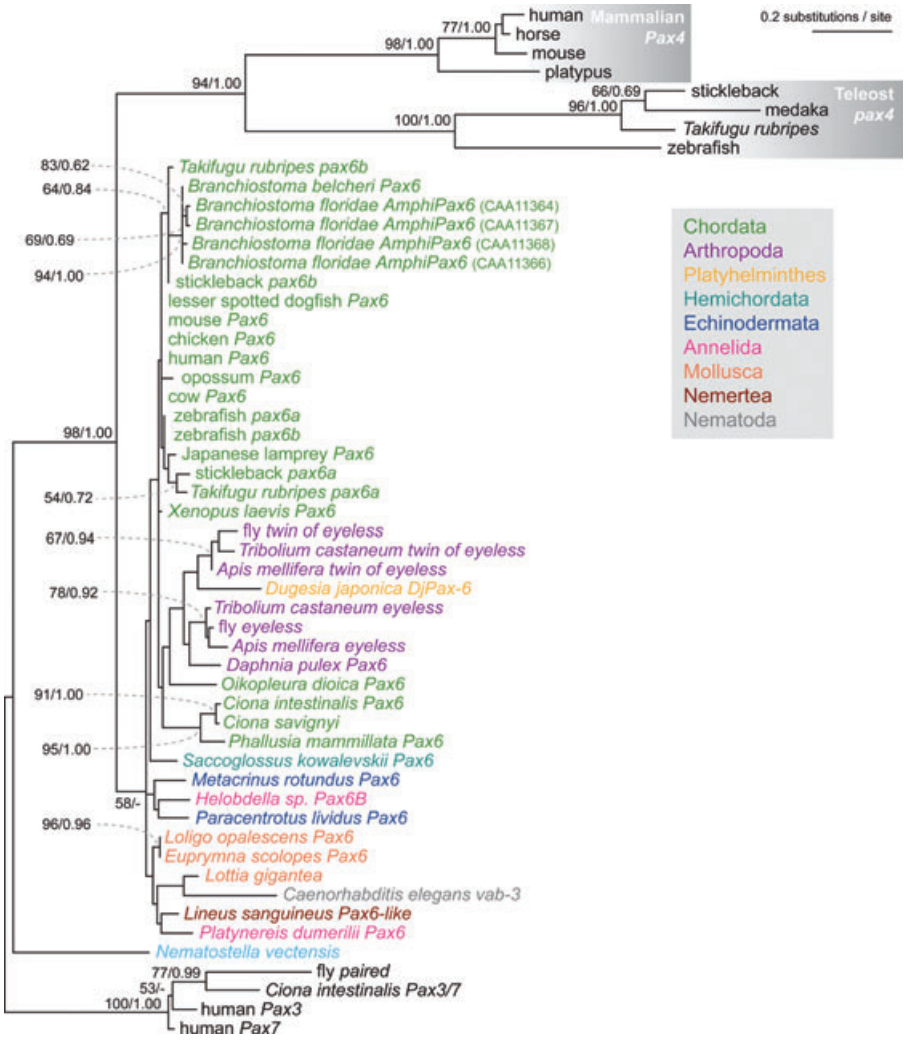

Fig. 3. Molecular phylogeny focusing on the Pax4/6 class of genes based on a broad taxon sampling. This tree was heuristically inferred with the maximum-likelihood (ML) method in PhyML (181 amino acid residues; shape parameter for the gamma distribution $\alpha = 0.88$). Support values at nodes are shown in order, bootstrap probabilities in the ML analysis, and Bayesian posterior probabilities. The support values are shown only when bootstrap probabilities are greater than 50. *Pax6* orthologs are colored for different animal phyla (see box at the right). See Table S1 for species names and accession IDs. Out of the five *B. floridae* sequence entries in GenBank, the deduced amino acid sequence of clone J2 (CAA11365) was not included because of unusual gaps.

and Teleostei. Our attempt with RT-PCR to identify *Pax4* in cyclostomes, chondrichthyans and nonteleost actinopterygian fishes resulted in no additional orthologs, which should be confirmed with anticipated whole genome sequences of species in those missing lineages.

Molecular phylogeny of *Pax4* and *Pax6*

Our molecular phylogenetic analysis used two sequence datasets. The first dataset included diverse invertebrates as well as vertebrates (see Table S1). Heuristic ML tree search and Bayesian inference produced consistent results on several points (Fig. 3). The putative *N. vectensis* (starlet sea anemone) *Pax6* ortholog was placed outside the monophyletic group of bilaterian sequences. Inside the *Pax6* group of bilaterians, however, the resultant tree topology with many low support values was largely inconsistent with generally accepted species phylogeny. For this reason, this phylogenetic analysis did not provide sufficient resolution to evaluate the alternative scenarios introduced in Fig. 1, although the overall tree topology vaguely supported the scenario that the gene duplication giving rise to *Pax4* occurred after the cnidaria–bilateria split, but before the deuterostome–protostome split (bootstrap probability in the ML analysis, 58). In contrast, the closest relationship between mammalian *Pax4* and teleost fish *pax4*, as well as monophylies of these two individual groups, was relatively strongly supported (Fig. 3; bootstrap probability in the ML analysis, 94; Bayesian posterior probability, 1.00). *toy* and *eyeless* (*ey*) genes of arthropods were closely related to each other, possibly because of a gene duplication in the insect lineage (Punzo et al. 2004; Lynch and Wagner 2011).

To perform a more focused assessment of the alternative scenarios, we prepared the second sequence dataset. In the previous dataset, there were four *Branchiostoma floridae* sequences (designated *AmphiPax6*) with polymorphic

Table 1. Result of maximum-likelihood analysis on *Pax4/6* phylogeny

Hypothesis	Tree topology	log*L*	*P*-value		RELL BP
			AU	SH	
1	(o,(pr,(am6,(tu6,(Lj,(g6,(t4,m4)))))))	−3240.37	0.55	0.78	0.00
2	(o,(pr,(am6,(tu6,((Lj,g6),(t4,m4))))))	−3240.37	0.55	0.99	0.21
3	(o,(pr,(am6,((tu6,(Lj,g6)),(t4,m4)))))	−3240.41	0.48	0.67	0.05
4	(o,(pr,((am6,(tu6,(Lj,g6))),(t4,m4))))	−3240.41	0.48	0.67	0.24
5	(o,((pr,(am6,(tu6,(Lj,g6)))),(t4,m4)))	−3240.63	0.45	0.47	0.42
6	(o,(pr,(am6,(tu6,(g6,(Lj,(t4,m4)))))))	−3240.37	0.55	0.83	0.09

Statistical supports of tree topologies corresponding to the six assumed hypotheses about the origin of *Pax4* are shown. Sequences were categorized into eight OTUs with constraints according to the generally accepted species phylogeny as follows: am6, amphioxus *Pax6*; g6, jawed vertebrate *Pax6*; Lj, lamprey (*Lethenteron japonicum*) *Pax6*; m4, mammalian *Pax4*; o, outgroup (putative *Nematostella vectensis Pax6* ortholog, *Ciona Pax3/7*, fly *paired*, human *Pax3*, and human *Pax7*); pr, protostome *Pax6*; t4, teleost *pax4*; tu6, tunicate *Pax6*. See Table S1 for species names and accession IDs of sequences in the dataset. Abbreviations: AU, approximately unbiased test; RELL BP, bootstrap probability based on re-sampling of log likelihood; log*L*, log-likelihood value; SH, Shimodaira–Hasegawa test. See Fig. S4 for hypotheses 1–6.

nonsynonymous changes (Glardon et al. 1998) as well as a *B. belcheri* sequence (Fig. 3). The differences between these sequences were thought to have been introduced in the amphioxus lineage, because the monophyly of them was strongly supported (Fig. 3; bootstrap probability in the ML analysis, 94; Bayesian posterior probability, 1.00). Of those, we selected only one *B. floridae* sequence (CAA11366) with no such lineage-specific substitution. We excluded *Dugesia japonica* and *C. elegans* because of long branches leading to these sequences (Fig. 3). As jawed vertebrates, we retained human, opossum, *X. laevis* and both *pax6a* and *pax6b* of zebrafish, *Takifugu rubripes*, and stickleback. *Loligo opalescent Pax6* was removed because its sequence was identical to *Euprymna scolopes Pax6*. We also excluded *Saccoglossus kowalevskii Pax6* and echinoderm *Pax6* (*Paracentrotus lividus* and *Metacrinus rotundus*) and medaka *pax4*. Using this second dataset including selected sequences, we performed a heuristic ML analyses. This analysis produced highly ambiguous results (data not shown) as in the analysis employing the first dataset (Fig. 3).

To statistically evaluate all possible tree topologies with this selected dataset, we performed an exhaustive ML analysis. To focus on the relationships of *Pax4* genes with *Pax6* and protostomes *Pax6* orthologs, we classify the sequences into eight OTUs with their internal relationships constrained according to generally accepted species phylogeny (see Materials and Methods).

This analysis resulted in the ML tree topology supporting a closer relationship of amphioxus *Pax6* to jawed vertebrate *Pax4* rather than to jawed vertebrate *Pax6* (Table S2; Fig. S3). It was also suggested that the *Pax4–Pax6* split occurred more recently than in the previous analysis (Fig. 3), namely in the chordate lineage. However, our comparison of the difference of the likelihood of each tree topology from that of the ML tree topology revealed as many as 360 tree topologies not rejected with 1σ of the log likelihood ($\Delta\log L/\sigma < 1$), twenty of which are listed in Table S2. Among the highly ranked tree topologies including the ML, no substantial difference was observed in the levels of support based on the approximately unbiased test (Shimodaira 2002), the Shimodaira–Hasegawa test (Shimodaira and Hasegawa 1999) and re-sampling of estimated log-likelihoods bootstrap probability (Kishino et al. 1990; Table S2). The clustering between teleost *pax4* and mammalian *Pax4* genes was relatively strongly supported (bootstrap probability in the ML analysis, 97; Bayesian posterior probability, 1.00; Fig. S3). The tree topology violating this cluster had a significantly lower likelihood ($\Delta\log L = 13.42 \pm 8.33$).

Notably, apart from the position of *pax4* genes, the ML tree topology as well as those supported with similar likelihood values (Table S2) are inconsistent with the generally accepted species phylogeny, when we assume orthology between *Pax6/eyeless* genes of diverse bilaterians. Thus, in order to assess alternative scenarios in a probabilistic framework based on the species phylogeny, we limited our targets of the statistical analysis with CONSEL to six tree topologies varying only the position of vertebrate *Pax4* (Fig. S4). These six included those introduced in Fig. 1 and the branching pattern with weak support in Fig. 3. As a result, these tree topologies were revealed to be almost equally probable (Table 1). It was also notable that when we compare these six tree topologies with the ML tree in the heuristic analysis, all of the six were ranked below 1σ in likelihood values (data not shown).

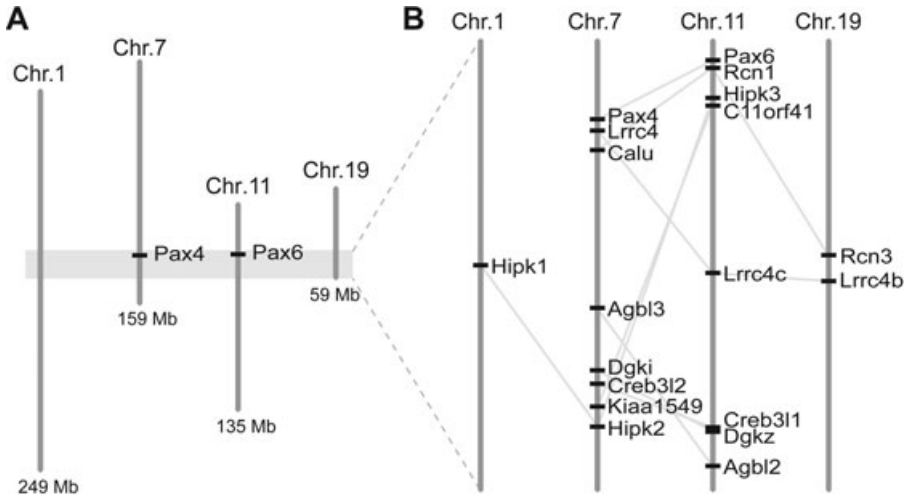

Fig. 4. Conserved synteny containing *Pax4* and *Pax6* genes. The 17.5 Mb stretches on human chromosomes 1, 7, 11, and 19 indicated with gray background in A are magnified in B. Seven gene families share paralogs commonly in the vicinity of *Pax4* and *Pax6* in the human genome (see Table S4 for their exact base positions). Members of the same gene family were connected with gray lines. Our phylogenetic analysis suggested that those members in individual gene families duplicated early in vertebrate evolution (Fig. S5).

Examination of the scale of the *Pax4–Pax6* duplication

If the *Pax4–Pax6* split took place in the vertebrate lineage (Fig. 1A), it is likely that it was part of the 2R-WGDs. In this scenario, similar arrays of genes should be found between genomic regions containing *Pax4* and *Pax6*. Analyzing phylogeny of those genes may allow us to date the timing of the duplication event. We performed a comprehensive search of conserved synteny by comparing gene compositions in 40 Mb genomic stretches (20 Mb on both ends) containing *Pax4* and *Pax6* in the human genome (see Materials and Methods). The search resulted in eight gene families whose members were shared between the two stretches (Fig. S5).

One of these eight gene families included the *mitochondrial inner membrane protease subunit* 1 (*IMMP1L*) gene on chromosome 11 and the *IMMP2L* gene on chromosome 7. This family experienced a gene duplication before the split between the animal and plant lineages (Fig. S5A). Except for this case, all the other seven shared genes were shown to have been duplicated in the vertebrate lineage, before the radiation of jawed vertebrates. In all cases where a cartilaginous fish sequence was available, it firmly clustered with a particular group of osteichthyan orthologs (e.g., cAMP responsive element binding protein 3-like 1 [*CREB3L1*], *LRRC4*; Fig. S5, B and C). Similarly, although not unambiguously supported, sea lamprey sequences also clustered with a particular group of jawed vertebrate orthologs (e.g., *LRRC4*, *HIPK2*, and diacylglycerol kinase zeta (*DGKZ*); Fig. S5, C, E, and F), suggesting that duplications of these genes occurred before the radiation of all extant vertebrates.

In spite of the wide scope (40 Mb) of our comparison, the seven genes spanned only 15.9 Mb (on chromosome 11) and 12.1 Mb (on chromosome 7), with both of *Pax6* and *Pax4* residing on the end of the shared gene arrays, respectively (Fig. 4). Our comprehensive survey of similar sequences in animals and molecular phylogenetic analysis detected additional paralogs that duplicated at the same evolutionary timing. *Leucine-rich repeat-containing 4B* (*LRRC4B*) and *reticulocalbin 3* both on chromosome 19 were revealed to be paralogs of the genes identified above on chromosomes 7 and 11 (Fig. 4; Fig. S5, C and D). In addition, *homeodomain interacting protein kinase 1* (*HIPK1*), paralogous to *HIPK2* and *HIPK3*, was found on chromosome 1 (Fig. 4; Fig. S5E).

Comparison of noncoding regions of *Pax4* and *Pax6* genes

It seemed possible that some of expression domains shared between *Pax4* and *Pax6* genes (see Table S3) are driven by *cis*-regulatory elements shared between these two genes. To examine this, we downloaded genome sequences containing *Pax4* and *Pax6* genes in diverse vertebrates. We used two different approaches to identifying noncoding sequences shared between *Pax4*-containing and *Pax6*-containing genomic regions (see Materials and Methods). However, both did not reveal any significant hit (data not shown).

We identified upstream noncoding sequences conserved within mammalian *Pax4* (Fig. S6A), and within teleost fish *pax4* (Fig. S6B). However, no noncoding sequences flanking *Pax4* was revealed to be conserved between mammal *Pax4* and teleost fish *pax4* (Fig. S6, A and B).

DISCUSSION

Pax4 and *Pax6* repertoires in vertebrates

Our survey based on available large-scale genomic and transcriptomic sequences indicated the absence of *Pax4* genes in sauropsids (birds and reptiles) and amphibians. It is very likely that *Pax4* genes were lost in these lineages independently. We also failed to identify *Pax4* genes in

chondrichthyans and cyclostomes, for which the *Pax6* gene has already been reported. Interestingly, our phylogenetic analysis did not necessarily rule out the possibility that the dogfish and lamprey *Pax6* sequences are orthologous to *Pax4* (Fig. 3; Table S2). However, expressions of these early vertebrate *Pax6* genes in the CNS (Murakami et al. 2001; Derobert et al. 2002), as well as a high level of conservation of amino acid sequences between them and osteichthyan *Pax6* (Fig. S2), suggest their orthology to osteichthyan *Pax6* genes. Taken together, *Pax4* genes have only been identified in mammals and teleost fishes.

Phylogenetic origin of *Pax4*

Identification of *Pax4* orthologs in teleost fishes supported the improbability of the scenario in Fig. 1B, namely a gene duplication specific to the mammalian lineage. It was recognized very early that *Pax6* sequences exhibit an extremely high level of sequence similarity among them, whereas those of *Pax4* are very divergent (Balczarek et al. 1997). To accommodate this rate heterogeneity in the dataset, we mainly adopted the ML method that is known to be less prone to artifacts such as long-branch attraction (Philippe et al. 2005b). The analysis significantly supported the orthology of teleost *pax4* to mammalian *Pax4* (Fig. 3; Fig. S3; also see Results). However, regarding the timing of the *Pax4–Pax6* split, our phylogenetic analysis did not provide unambiguous results (Table 1). It remained unclear which of the alternative hypotheses in Fig. S4 (including those in Figs. 1, A and C) delineates the timing of the *Pax4–Pax6* duplication. Because our dataset already contains representative species from the major chordate lineages, it does not seem likely that further identification of *Pax4/6*-related sequences will largely improve the resolution. The weakly supported molecular phylogeny described so far urged us to focus on a different aspect of the evolution of *Pax4* and *Pax6* genes.

Genomic background of the *Pax4–Pax6* duplication

To examine the timing of the duplication between *Pax4* and *Pax6*, we referred to the chromosomal locations of these genes and their neighbors. By detecting similar arrays of genes shared between chromosomes (conserved synteny) in a genome and reconstructing the evolutionary history of the harbored gene families, we can map the timing of large-scale duplications on the species phylogeny. In the human genome, several quartets of chromosomes showing conserved synteny have been detected (Kasahara et al. 1996). Some of these served as initial convincing evidence of intragenome duplications (Lundin 1993; Holland et al. 1994; Sidow 1996; Spring 1997). However, it is also expected that chromosomal rearrangements accelerated the decay of ancestral gene order during evolution. Although some effort has been made to reconstruct the ancestral vertebrate karyotype (Nakatani et al. 2007; Putnam et al. 2008), only a small fraction of all genes in sequenced genomes is implicated in those highly conserved syntenic regions.

Our analysis detected eight gene families whose members are co-localized inside 40 Mb genomic regions containing *Pax4* and *Pax6* on chromosomes 7 and 11, respectively (Fig. 4). Except for only one case, molecular phylogenetic analyses suggested that the duplications between genes on chromosomes 7 and 11 occurred early in vertebrate evolution (Fig. S5). This implies a large-scale duplication between these chromosomal regions. So far, no large-scale duplication event before the split between teleost and tetrapod lineages, other than the 2R-WGDs, has been documented (Van de Peer et al. 2009). Thus, it is likely that the *Pax4–Pax6* split was caused by the 2R-WGDs early in vertebrate evolution (Fig. 1A).

Role of Pax4 and its evolutionary change

We showed that zebrafish *pax4* is expressed in the developing pancreas and the stomodeum (Fig. 2). The *pax4* expression in the pancreas, nested in the broader *pax6b* expression (Fig. 2, D–F), is concordant with the pattern in mouse, where *Pax4* expression is restricted to β-cells, whereas *Pax6* is expressed in all the four cell types of the endocrine pancreas (St-Onge et al. 1997; Biemar et al. 2001; Delporte et al. 2008). This similarity indicates their common ancestry at the base of the Osteichthyes.

Our comparison of noncoding genomic sequences containing *Pax4* orthologs detected several conserved elements within mammals and within teleost fishes (Fig. S6). This included the only upstream enhancer characterized to date that is responsible for the pancreatic expression of *Pax4* in mouse (Brink et al. 2001). However, none of these potential *cis*-regulatory elements were shared between mammals and teleost fishes with a comparable level of similarity (Fig. S6). Our intensive search for CNE shared between *Pax4* and *Pax6* also failed to detect potential *cis*-regulatory elements commonly retained between these duplicates (see Materials and Methods).

Expression in the stomodeum, the other *pax4*-positive domain in zebrafish, has never been described for mammalian *Pax4* as well as for *Pax6* genes. Thus, this expression domain should have been gained in the teleost fish lineage. On the other hand, expression in the pineal gland and the retina, described for mammals (Rath et al. 2009a, 2009b), was not detected in zebrafish (Fig. 2). Expressions in the retina and the pineal gland have also been reported for *Pax6* in many vertebrates (Walther and Gruss 1991; Kawakami et al. 1997; Derobert et al. 2002; Navratilova et al. 2009). Interestingly, even the amphioxus *Pax6* ortholog, *AmphiPax6*,

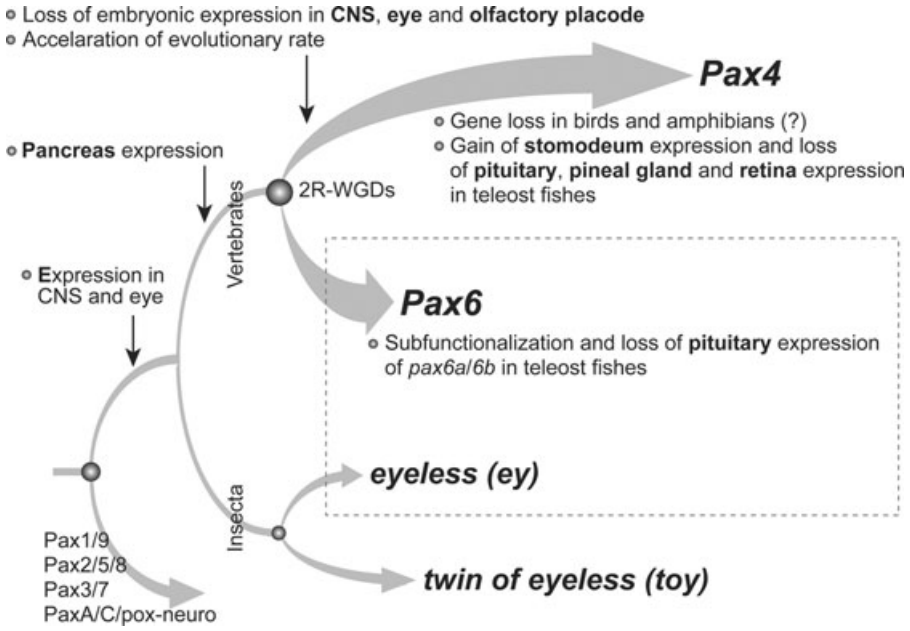

Fig. 5. A hypothesized scenario for phylogenetic and regulatory properties of *Pax4* and *Pax6*. The orthology between *Pax6* and *eyeless* (*ey*), which is usually referred to as functional equivalence, is highlighted with a dotted box. Including a duplicate in the vertebrate lineage, *Pax4*, and a duplicate in the insect lineage, *twin of eyeless* (*toy*), the relationship is 2:2 orthology between vertebrates and the fly. It should be noted that none of zebrafish *pax6a and pax6b* is expressed in the pituitary gland (Table S3). Although their expression was originally implicated in the pituitary (Puschel et al. 1992), no further studies, including ours, confirmed this interpretation. In zebrafish, two *Pax6* orthologs, *pax6a* and *pax6b*, are expressed in a complementary manner as a result of so-called subfunctionalization caused by the teleost-specific genome duplication (Kleinjan et al. 2008). Expression domains of *Pax4/6* genes in this figure are based on the literature included in Table S3. See Kammermeier et al. (2001) for functional divergences of *eyeless* and *twin of eyeless* in the insect lineage.

is expressed in the lamellar body, which is homologous to the pineal gland (Glardon et al. 1998). With a few exceptions (absence of zebrafish *pax4* expression in the retina and pineal gland and absence of *Xenopus Pax6* expression in the pineal gland [Hirsch and Harris 1997]), *Pax4* and *Pax6* genes are generally expressed in the retina and pineal gland, suggesting an ancient origin of these expression domains before the *Pax4–Pax6* duplication.

Although *Pax4* and *Pax6* seem to have retained a subset of expression domains, such as the pancreas, retina, and pineal gland after the gene duplication, one striking feature of *Pax4* is the absence of its expression in the CNS, including the eye and olfactory placode (Fig. 2; Table S3). *Pax4* genes seem to have evolved relatively rapidly, based on long branches in molecular phylogenetic trees (Fig. 3 and S3), experienced more dynamic secondary modification of expression patterns, and may have been lost in the birds and amphibian lineages (Fig. 5). In contrast, *Pax6* genes have highly conserved coding sequences (Fig. 3 and S3), experienced fewer changes in its highly pleiotropic expression, and have been retained in all species studied to date (Fig. 5). The asymmetry in gene retention, sequence conservation, and developmental regulation between *Pax4* and *Pax6* illustrates the extent to which gene duplications have contributed to the elaboration of gene regulatory networks that govern vertebrate embryogenesis.

Acknowledgments

This study was supported by the Young Scholar Fund, University of Konstanz to SK, the grants German Research Foundation (DFG) to SK (KU2669/1-1), Konstanz Research School Chemical Biology (KoRS-CB) to TM, and International Max Planck Research School (IMPRS) for Organismal Biology to NF. We thank Nicola Blum, Silke Pittlik, Adina J. Renz, Ursula Topel, and Elke Hespeler for technical support in cDNA cloning, handling of zebrafish embryos, and in situ hybridization. Our gratitude extends to two anonymous reviewers for their constructive suggestions.

REFERENCES

Abascal, F., Zardoya, R., and Posada, D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104–2105.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.

Balczarek, K. A., Lai, Z. C., and Kumar, S. 1997. Evolution of functional diversification of the paired box (Pax) DNA-binding domains. *Mol. Biol. Evol.* 14: 829–842.

Bassham, S., Canestro, C., and Postlethwait, J. H. 2008. Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. *BMC Biol.* 6: 35.

Begemann, G., Schilling, T. F., Rauch, G. J., Geisler, R., and Ingham, P. W. 2001. The zebrafish *neckless* mutation reveals a requirement for raldh2 in mesodermal signals that pattern the hindbrain. *Development* 128: 3081–3094.

Biemar, F., Argenton, F., Schmidtke, R., Epperlein, S., Peers, B., and Driever, W. 2001. Pancreas development in zebrafish: early dispersed appearance of endocrine hormone expressing cells and their convergence to form the definitive islet. *Dev. Biol.* 230: 189–203.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G., and Noll, M. 1986. Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of *Drosophila*. *Cell* 47: 1033–1040.

Breitling, R., and Gerber, J. K. 2000. Origin of the paired domain. *Dev. Genes Evol.* 210: 644–650.

Brink, C., Chowdhury, K., and Gruss, P. 2001. *Pax4* regulatory elements mediate beta cell specific expression in the pancreas. *Mech. Dev.* 100: 37–43.

Chi, N., and Epstein, J. A. 2002. Getting your Pax straight: Pax proteins in development and disease. *Trends Genet.* 18: 41–47.

Chisholm, A. D., and Horvitz, H. R. 1995. Patterning of the *Caenorhabditis elegans* head region by the Pax-6 family member vab-3. *Nature* 377: 52–55.

Collombat, P., Xu, X., Ravassard, P., Sosa-Pineda, B., Dussaud, S., et al. 2009. The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into alpha and subsequently beta cells. *Cell* 138: 449–462.

Cracraft, J., and Donoghue, M. J. 2004. *Assembling the Tree of Life*. Oxford University Press, Oxford, New York.

Delporte, F. M., Pasque, V., Devos, N., Manfroid, I., Voz, M., et al. 2008. Expression of zebrafish *pax6b* in pancreas is regulated by two enhancers containing highly conserved cis-elements bound by PDX1, PBX and PREP factors. *BMC Dev. Biol.* 8: 53.

Derobert, Y., Baratte, B., Lepage, M., and Mazan, S. 2002. *Pax6* expression patterns in *Lampetra fluviatilis* and *Scyliorhinus canicula* embryos suggest highly conserved roles in the early regionalization of the vertebrate brain. *Brain Res. Bull.* 57: 277–280.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32: W273–W279.

Gehring, W. J., and Ikeo, K. 1999. *Pax6*: mastering eye morphogenesis and eye evolution. *Trends Genet.* 15: 371–377.

Glardon, S., Holland, L. Z., Gehring, W. J., and Holland, N. D. 1998. Isolation and developmental expression of the amphioxus Pax-6 gene (*AmphiPax-6*): insights into eye and photoreceptor evolution. *Development* 125: 2701–2710.

Gonez, L. J., and Knight, K. R. 2010. Cell therapy for diabetes: stem cells, progenitors or beta-cell replication? *Mol. Cell. Endocrinol.* 323: 55–61.

Goode, D. K., and Elgar, G. 2009. The PAX258 gene subfamily: a comparative perspective. *Dev. Dyn.* 238: 2951–2974.

Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52: 696–704.

Halder, G., Callaerts, P., and Gehring, W. J. 1995. Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science* 267: 1788–1792.

Hirsch, N., and Harris, W. A. 1997. *Xenopus* Pax-6 and retinal development. *J. Neurobiol.* 32: 45–61.

Holland, L. Z., Schubert, M., Kozmik, Z., and Holland, N. D. 1999. *AmphiPax3/7*, an amphioxus paired box gene: insights into chordate myogenesis, neurogenesis, and the possible evolutionary precursor of definitive vertebrate neural crest. *Evol. Dev.* 1: 153–165.

Holland, N. D., Holland, L. Z., and Kozmik, Z. 1995. An amphioxus Pax gene, *AmphiPax-1*, expressed in embryonic endoderm, but not in mesoderm: implications for the evolution of class I paired box genes. *Mol. Mar. Biol. Biotechnol.* 4: 206–214.

Holland, P. W., Garcia-Fernandez, J., Williams, N. A., and Sidow, A. 1994. Gene duplications and the origins of vertebrate developement. *Dev. Sppl.* 125–133.

Hoshiyama, D., Suga, H., Iwabe, N., Koyanagi, M., Nikoh, N., et al. 1998. Sponge Pax cDNA related to Pax-2/5/8 and ancient gene duplications in the Pax family. *J. Mol. Evol.* 47: 640–648.

Hubbard, T. J., Aken, B. L., Ayling, S., Ballester, B., Beal, K., et al. 2009. Ensembl 2009. *Nucleic Acids Res.* 37: D690–D697.

Huelsenbeck, J. P., and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Kammermeier, L., Leemans, R., Hirth, F., Flister, S., Wenger, U., et al. 2001. Differential expression and function of the Drosophila Pax6 genes eyeless and twin of eyeless in embryonic central nervous system development. *Mech. Dev.* 103: 71–78.

Kasahara, M., Hayashi, M., Tanaka, K., Inoko, H., Sugaya, K., et al. 1996. Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proc. Natl. Acad. Sci. U S A* 93: 9096–9101.

Katoh, K., Kuma, K., Toh, H., and Miyata, T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33: 511–518.

Kawakami, A., Kimura-Kawakami, M., Nomura, T., and Fujisawa, H. 1997. Distributions of PAX6 and PAX7 proteins suggest their involvement in both early and late phases of chick brain development. *Mech. Dev.* 66: 119–130.

Kinkel, M. D., and Prince, V. E. 2009. On the diabetic menu: zebrafish as a model for pancreas development and function. *Bioessays* 31: 139–152.

Kishino, H., Miyata, T., and Hasegawa, M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* 30: 151–160.

Kleinjan, D. A., Bancewicz, R. M., Gautier, P., Dahm, R., Schonthaler, H. B., et al. 2008. Subfunctionalization of duplicated zebrafish *pax6* genes by *cis*-regulatory divergence. *PLoS Genet.* 4: e29.

Kozmik, Z., Holland, N. D., Kalousova, A., Paces, J., Schubert, M., and Holland, L. Z. 1999. Characterization of an amphioxus paired box gene, *AmphiPax2/5/8*: developmental expression patterns in optic support cells, nephridium, thyroid-like structures and pharyngeal gill slits, but not in the midbrain-hindbrain boundary region. *Development* 126: 1295–1304.

Kuraku, S. 2008. Insights into cyclostome phylogenomics: pre-2R or post-2R. *Zoolog. Sci.* 25: 960–968.

Kuraku, S. and Meyer, A. 2009. The evolution and maintenance of *Hox* gene clusters in vertebrates and the teleost-specific genome duplication. *Int. J. Dev. Biol.* 53: 765–773.

Kuraku, S., Meyer, A., and Kuratani, S. 2009. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol. Biol. Evol.* 26: 47–59.

Liu, Z. and Habener, J. F. 2009. Alpha cells beget beta cells. *Cell* 138: 424–426.

Lundin, L. G. 1993. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* 16: 1–19.

Lynch, V. J. and Wagner, G. P. 2011. Revisiting a classic example of transcription factor functional equivalence: are *eyeless* and *Pax6* functionally equivalent or divergent? *J. Exp. Zool. B Mol. Dev. Evol.*, 316B: 93–98.

Matus, D. Q., Pang, K., Daly, M., and Martindale, M. Q. 2007. Expression of Pax gene family members in the anthozoan cnidarian, *Nematostella vectensis*. *Evol. Dev.* 9: 25–38.

McCauley, D. W., and Bronner-Fraser, M. 2002. Conservation of Pax gene expression in ectodermal placodes of the lamprey. *Gene* 287: 129–139.

Meyer, A., and Zardoya, R. 2003. Recent advances in the (molecular) phylogeny of vertebrates. *Annu. Rev. Ecol. Evol. Syst.* 34: 311–338.

Mise, T., Iijima, M., Inohaya, K., Kudo, A., and Wada, H. 2008. Function of Pax1 and Pax9 in the sclerotome of medaka fish. *Genesis* 46: 185–192.

Murakami, Y., Ogasawara, M., Sugahara, F., Hirano, S., Satoh, N., and Kuratani, S. 2001. Identification and expression of the lamprey *Pax6* gene: evolutionary origin of the segmented brain of vertebrates. *Development* 128: 3521–3531.

Nakatani, Y., Takeda, H., Kohara, Y., and Morishita, S. 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res.* 17: 1254–1265.

Navratilova, P., Fredman, D., Hawkins, T. A., Turner, K., Lenhard, B., and Becker, T. S. 2009. Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. *Dev. Biol.* 327: 526–540.

O'Neill, P., McCole, R. B., and Baker, C. V. 2007. A molecular analysis of neurogenic placode and cranial sensory ganglion development in the shark, *Scyliorhinus canicula*. *Dev. Biol.* 304: 156–181.

Ogasawara, M., Wada, H., Peters, H., and Satoh, N. 1999. Developmental expression of Pax1/9 genes in urochordate and hemichordate gills: insight into function and evolution of the pharyngeal epithelium. *Development* 126: 2539–2550.

Panopoulou, G., and Poustka, A. J. 2005. Timing and mechanism of ancient vertebrate genome duplications—the adventure of a hypothesis. *Trends Genet.* 21: 559–567.
Philippe, H., Lartillot, N., and Brinkmann, H. 2005a. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* 22: 1246–1253.
Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., and Delsuc, F. 2005b. Heterotachy and long-branch attraction in phylogenetics. *BMC Evol. Biol.* 5: 50.
Pilz, A. J., Povey, S., Gruss, P., and Abbott, C. M. 1993. Mapping of the human homologs of the murine paired-box-containing genes. *Mamm. Genome* 4: 78–82.
Punzo, C., Plaza, S., Seimiya, M., Schnupf, P., Kurata, S., et al. 2004. Functional divergence between *eyeless* and *twin of eyeless* in *Drosophila melanogaster*. *Development* 131: 3943–3953.
Puschel, A. W., Gruss, P., and Westerfield, M. 1992. Sequence and expression pattern of pax-6 are highly conserved between zebrafish and mice. *Development* 114: 643–651.
Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453: 1064–1071.
Rath, M. F., Bailey, M. J., Kim, J. S., Coon, S. L., Klein, D. C., and Moller, M. 2009a. Developmental and daily expression of the *Pax4* and *Pax6* homeobox genes in the rat retina: localization of *Pax4* in photoreceptor cells. *J. Neurochem.* 108: 285–294.
Rath, M. F., et al. 2009b. Developmental and diurnal dynamics of Pax4 expression in the mammalian pineal gland: nocturnal down-regulation is mediated by adrenergic-cyclic adenosine 3′,5′-monophosphate signaling. *Endocrinology* 150: 803–811.
Schmidt, H. A., Strimmer, K., Vingron, M., and von Haeseler, A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18: 502–504.
Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51: 492–508.
Shimodaira, H., and Hasegawa, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16: 1114–1116.
Shimodaira, H., and Hasegawa, M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246–1247.
Sidow, A. 1996. Gen(om)e duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev.* 6: 715–722.
Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G., and Gruss, P. 1997. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. *Nature* 386: 399–402.
Spring, J. 1997. Vertebrate evolution by interspecific hybridization—are we polyploid? *FEBS Lett.* 400: 2–8.
St-Onge, L., Sosa-Pineda, B., Chowdhury, K., Mansouri, A., and Gruss, P. 1997. Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. *Nature* 387: 406–409.
Thorel, F., et al. 2010. Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. *Nature* 464: 1149–1154.
Tokuyama, Y., Yagui, K., Sakurai, K., Hashimoto, N., Saito, Y., and Kanatsuka, A. 1998. Molecular cloning of rat Pax4: identification of four isoforms in rat insulinoma cells. *Biochem. Biophys. Res. Commun.* 248: 153–156.
Tsagkogeorga, G., Turon, X., Hopcroft, R. R., Tilak, M. K., Feldstein, T., et al. 2009. An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. *BMC Evol. Biol.* 9: 187.
Van de Peer, Y., Maere, S., and Meyer, A. 2009. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10: 725–732.
Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L. A. 2007. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* 35: D88–D92.
Wada, H., Saiga, H., Satoh, N., and Holland, P. W. 1998. Tripartite organization of the ancestral chordate brain and the antiquity of placodes: insights from ascidian Pax-2/5/8, Hox and Otx genes. *Development* 125: 1113–1122.
Walther, C., and Gruss, P. 1991. Pax-6, a murine paired box gene, is expressed in the developing CNS. *Development* 113: 1435–149.
Wehr, R., and Gruss, P. 1996. Pax and vertebrate development. *Int. J. Dev. Biol.* 40: 369–77.
Wiegmann, B. M., Trautwein, M. D., Kim, J. W., Cassel, B. K., Bertone, M. A., et al. 2009. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol.* 7: 34.
Zhang, Y., and Emmons, S. W. 1995. Specification of sense-organ identity by a *Caenorhabditis elegans* Pax-6 homologue. *Nature* 377: 55–59.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. Accession IDs of the sequences included in molecular phylogenetic trees.

Table S2. Result of maximum-likelihood analysis on *Pax4/6* phylogeny.

Table S3. Expression domains of *Pax6* and *Pax4*.

Table S4. Chromosomal location of the human genes included in Fig. 4

Fig. S1. Identification of the open reading frame of the zebrafish *pax4* gene.

Fig. S2. Amino acid alignment of Pax4 and Pax6 sequences.

Fig. S3. Molecular phylogenetic analysis of *Pax4* and *Pax6* genes.

Fig. S4. Schematic tree showing alternative phylogenetic positions of *Pax4* genes.

Fig. S5. Maximum-likelihood trees of gene families whose members are shared between *Pax4*- and *Pax6*-containing regions in the human genome.

Fig. S6. Sequence comparison of *Pax4*-containing genomic regions.