



## Coelacanth genome sequence reveals the evolutionary history of vertebrate genes

James P. Noonan, Jane Grimwood, Joshua Danke, et al.

*Genome Res.* 2004 14: 2397-2405

Access the most recent version at doi:[10.1101/gr.2972804](https://doi.org/10.1101/gr.2972804)

---

**References** This article cites 27 articles, 11 of which can be accessed free at:  
<http://genome.cshlp.org/content/14/12/2397.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

# Coelacanth genome sequence reveals the evolutionary history of vertebrate genes

James P. Noonan,<sup>1,4</sup> Jane Grimwood,<sup>2</sup> Joshua Danke,<sup>3</sup> Jeremy Schmutz,<sup>2</sup> Mark Dickson,<sup>2</sup> Chris T. Amemiya,<sup>3</sup> and Richard M. Myers<sup>1,2,5</sup>

<sup>1</sup>Department of Genetics, Stanford University School of Medicine, Stanford, California 94305-5120, USA; <sup>2</sup>Stanford Human Genome Center, Stanford University School of Medicine, Palo Alto, California 94304, USA; <sup>3</sup>Benaroya Research Institute at Virginia Mason, Seattle, Washington 98101, USA

The coelacanth is one of the nearest living relatives of tetrapods. However, a teleost species such as zebrafish or *Fugu* is typically used as the outgroup in current tetrapod comparative sequence analyses. Such studies are complicated by the fact that teleost genomes have undergone a whole-genome duplication event, as well as individual gene-duplication events. Here, we demonstrate the value of coelacanth genome sequence by complete sequencing and analysis of the protocadherin gene cluster of the Indonesian coelacanth, *Latimeria menadoensis*. We found that coelacanth has 49 protocadherin cluster genes organized in the same three ordered subclusters,  $\alpha$ ,  $\beta$ , and  $\gamma$ , as the 54 protocadherin cluster genes in human. In contrast, whole-genome and tandem duplications have generated two zebrafish protocadherin clusters comprised of at least 97 genes. Additionally, zebrafish protocadherins are far more prone to homogenizing gene conversion events than coelacanth protocadherins, suggesting that recombination- and duplication-driven plasticity may be a feature of teleost genomes. Our results indicate that coelacanth provides the ideal outgroup sequence against which tetrapod genomes can be measured. We therefore present *L. menadoensis* as a candidate for whole-genome sequencing.

[Supplemental material is available online at [www.genome.org](http://www.genome.org) and <http://www-shgc.stanford.edu/myerslab/>. The BAC sequence data from this study have been submitted to GenBank under accession nos. AC150283, AC150284, and AC150308–AC150310.]

A primary aim of vertebrate comparative genomics is to reconstruct the evolutionary history of vertebrate genomes. To this end, genome sequence is currently being obtained from organisms at critical positions in the vertebrate phylogeny. Notably absent from the list of planned or completed genomes, however, is a species on a lineage arising between ray-finned fishes and tetrapods, when vertebrates were about to undergo an adaptive transformation. The sarcopterygian fishes, coelacanths and lungfish, are the only extant taxa that occupy this unique phylogenetic position (Gorr et al. 1991; Zardoya and Meyer 1996, 1997). As the nearest living relatives of the tetrapod ancestor, these species provide access to the phenotypic and genomic transitions leading to the emergence of tetrapods.

Despite substantial sequence analysis, it is unclear whether coelacanths or lungfish are more closely related to tetrapods (Takezaki et al. 2004). However, lungfish have very large genomes (>100 Gb), making them poor candidates for genomic sequencing. Coelacanths, although abundant in the fossil record, were believed to be extinct before a living specimen was identified in 1938 by Marjorie Latimer and J.L.B. Smith (Smith 1939). The two modern coelacanth species that are known, *Latimeria chalumnae* and *Latimeria menadoensis*, are remarkably similar to their fossil relatives, showing little morphological change over 360 million years (Smith 1939; Forey 1998; Holder et al. 1999). In

addition, the *L. menadoensis* genome is smaller than the human or mouse genomes, and is therefore as amenable to whole-genome sequencing (Danke et al. 2004; J.P. Noonan, J. Grimwood, J. Danke, J. Schmutz, M. Dickson, C.T. Amemiya, and R.M. Myers, unpubl.). The phenotypic stability of coelacanths over long evolutionary time scales suggests that the coelacanth genome may also be stable over long periods, evolving neutrally with few major rearrangements. This is in stark contrast to the contortions of genome evolution accompanying the teleost radiation (Postlethwait et al. 1998; Taylor et al. 2003). Analysis of 33 *L. menadoensis* *Hox* cluster cDNAs indicates that 32 of these genes have orthologs in the four mammalian *HOX* clusters (Koh et al. 2003). This suggests that the organization of coelacanth *Hox* cluster genes is similar to that in mammals, and unlike that in zebrafish, which has seven *hox* clusters as a consequence of the teleost whole-genome duplication (Amores et al. 1998). However, the order and orientation of coelacanth *Hox* genes and their distribution in the genome is not known. No detailed consideration of gene arrangement in coelacanth using genomic sequence has been attempted.

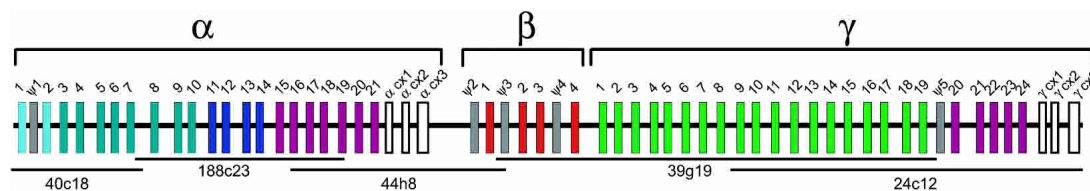
To evaluate the utility of coelacanth genome sequence in inferring the evolutionary history of tetrapod genomes, we isolated and sequenced clones from the VMRC-4 *L. menadoensis* BAC library that span the coelacanth protocadherin cluster (*LmPcdh*) (Danke et al. 2004). We generated ~609 kb of contiguous coelacanth genome sequence, and found that the cluster contains 49 tandemly arrayed protocadherin genes (Fig. 1). Protocadherin cluster proteins are cadherin-like synaptic cell-adhesion molecules required for normal brain development, and may provide a combinatorial molecular code for the generation

<sup>4</sup>Present address: Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

<sup>5</sup>Corresponding author.

E-mail [myers@shgc.stanford.edu](mailto:myers@shgc.stanford.edu); fax (650) 725-9689.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.2972804>. Article published online before print in November 2004.



**Figure 1.** The *Latimeria menadoensis* protocadherin cluster. Genes in each paralog subgroup are indicated by color. Constant exons are in white and pseudogenes ( $\psi$ ) are in gray. Sequenced VMRC-4 BAC clone names and their positions are shown below the cluster.

of synaptic complexity (Kohmura et al. 1998; Wu and Maniatis 1999; Wang et al. 2002). In mammals, protocadherins are single exon genes organized into three subclusters,  $\alpha$ ,  $\beta$ , and  $\gamma$ . Each exon encodes an extracellular domain consisting of six cadherin ectodomains, a transmembrane segment, and a short cytoplasmic tail. At the 3' end of both the  $\alpha$  and  $\gamma$  subclusters are an additional three short exons that are alternatively *cis*-spliced to each  $\alpha$  and  $\gamma$  exon, providing a common cytoplasmic domain that mediates intracellular signaling (Kohmura et al. 1998; Wu and Maniatis 1999).

As a tandem array of partially redundant genes, the protocadherin cluster is especially susceptible to lineage-specific modification by whole-genome or individual gene duplications. We previously identified 66 zebrafish protocadherin genes, widely diverged from mammalian protocadherins and organized in two unlinked clusters, *DrPcdh1* and *DrPcdh2*, arising from the teleost whole-genome duplication (Noonan et al. 2004). We have since identified an additional 31 *DrPcdh2* $\gamma$  exons in high-quality finished sequence from the most recent zebrafish genome assembly. Whole-genome and tandem gene duplication, followed by diversification of the duplicated genes, have also produced teleost-specific protocadherin paralogs with no counterpart in any other lineage. In this study, we show that the coelacanth protocadherin cluster is more similar to protocadherin clusters in mammals, both in organization and gene content, than are the zebrafish protocadherin clusters. In particular, coelacanth protocadherins have been spared multiple recent tandem or whole-genome duplications. Unlike zebrafish or mammalian protocadherins, coelacanth protocadherins are not subject to frequent homogenizing gene-conversion events, indicating that the coelacanth genome could be less prone to recombination-driven rearrangements than teleost or mammalian genomes. These results suggest that the coelacanth protocadherin cluster is likely very similar to the protocadherin cluster of the true tetrapod ancestor. We therefore propose that *L. menadoensis* provides the ideal reference genome for comparative sequence analyses in tetrapods.

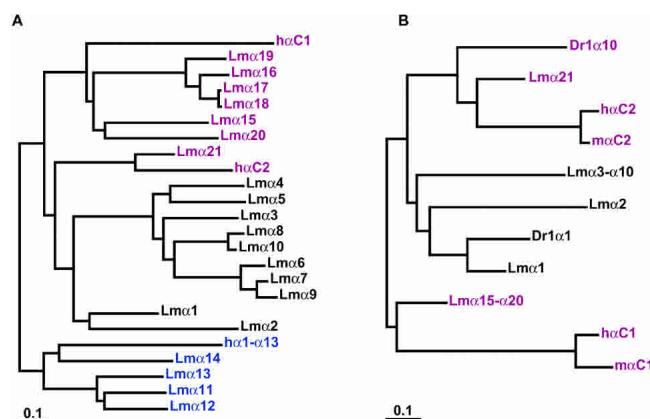
## Results and Discussion

### Comparison of coelacanth, human, and zebrafish protocadherin clusters identifies modifications specific to tetrapods and teleosts

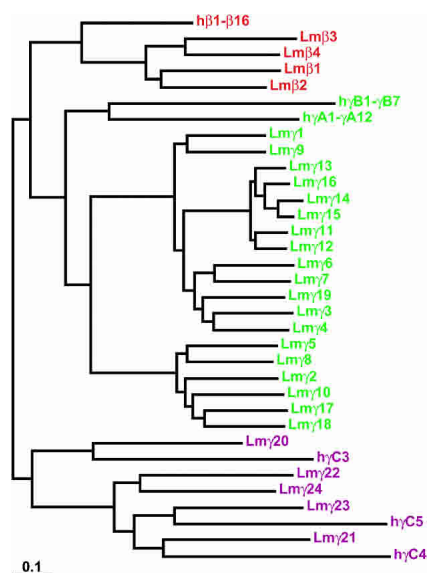
To determine the phylogenetic relationship of coelacanth, zebrafish, and human protocadherin cluster proteins, we aligned protocadherin extracellular domain sequences from all three species by using CLUSTALW, and built maximum likelihood (ML) trees in SEMPHY (see Methods for details). On the basis of these analyses and our library screening, coelacanth has one protocadherin cluster comprised of 49 variable exons arrayed in  $\alpha$ ,  $\beta$ , and  $\gamma$  subclusters, in the same order and orientation as mammalian protocadherins (Fig. 1). The coelacanth protocadherin cluster

shows no evidence of whole-genome duplication, consistent with the explanation that the coelacanth genome has not experienced a recent polyploidization event. The coelacanth  $\alpha$  and  $\gamma$  subclusters also have the characteristic arrangement of multiple variable and three constant region exons seen in mammals and zebrafish. Zebrafish, coelacanth, and human variable exon splice donor sites are essentially identical (data not shown). Few orthologous relationships are evident among individual coelacanth, human, and zebrafish protocadherins, but all three lineages share conserved *Pcdh $\alpha$*  and *Pcdh $\gamma$*  paralog subgroups (Figs. 2–4). The phylogenies of these paralog subgroups follow the species tree for these organisms, in which zebrafish is the outgroup to both human and coelacanth (Fig. 5). These subgroups are likely to be conserved in all vertebrates. However, the diversity of modern teleost and mammalian species is the product of an adaptive radiation in each lineage, which is driven by the lineage-specific adaptive evolution of genomes. Coelacanth sequence can be used to identify these tetrapod- and teleost-specific changes in gene content.

The 15 human protocadherin  $\alpha$  genes belong to one of two paralog subgroups, *Pcdh $\alpha$ 1- $\alpha$ 13* and *Pcdh $\alpha$ C1* and  $\alpha$ C2. This arrangement is conserved in mouse and, most likely, in all mammals (Wu et al. 2001; J.P. Noonan, J. Grimwood, J. Danke, J. Schmutz, M. Dickson, C.T. Amemiya, and R.M. Myers, unpubl.). The coelacanth *Pcdh $\alpha$*  subcluster is more complex than its mammalian counterparts, consisting of 21 variable exons organized into three divergent paralog subgroups as follows: *LmPcdh $\alpha$ 1- $\alpha$ 10*,  $\alpha$ 11 $\alpha$ -14, and  $\alpha$ 15 $\alpha$ -21 (Figs. 2A, 4). These subgroups are physically contiguous, indicating that they arose from three diverse, single-copy ancestors and have since resisted rearrange-



**Figure 2.** Ancestral *Pcdh $\alpha$*  paralog subgroups present in coelacanth have been lost in mammals. (A) Maximum likelihood phylogenies of coelacanth and human protocadherin  $\alpha$  cluster proteins. Colors indicate paralog subgroups as shown in Figure 1. The subtree of human *Pcdh $\alpha$ 1- $\alpha$ 13* is shown collapsed for clarity. (B) Phylogenetic relationship of zebrafish, coelacanth, human, and mouse *Pcdh $\alpha$ C* and coelacanth *Pcdh $\alpha$ 1- $\alpha$ 10* proteins.



**Figure 3.** Maximum likelihood phylogeny of coelacanth and human protocadherin  $\beta$  and  $\gamma$  proteins. Colors indicate paralog subgroups as shown in Figure 1. The subtrees of human  $Pcdh\beta 1$ - $\beta 16$ ,  $Pcdh\gamma A 1$ - $\gamma A 12$ , and  $Pcdh\gamma B 1$ - $\gamma B 7$  are shown collapsed for clarity.

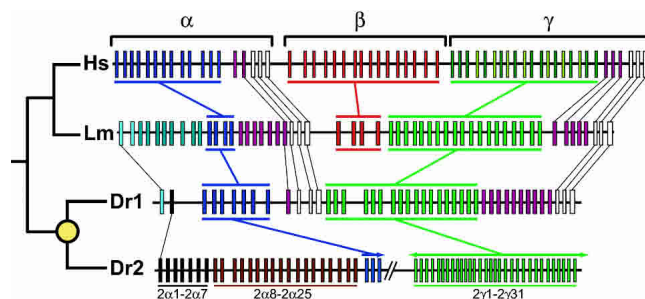
ment. *LmPcdh $\alpha 11$* - $\alpha 14$  and human *Pcdh $\alpha 1$* - $\alpha 13$  are derived from a common ancestral paralog, as are *LmPcdh $\alpha 15$* - $\alpha 21$  and human *Pcdh $\alpha C 1$* . *LmPcdh $\alpha 2$*  and human *Pcdh $\alpha C 2$*  are clearly orthologs (Fig. 2A). The order of these groups is also conserved between human and coelacanth protocadherin clusters (Fig. 4).

The majority *Pcdh $\alpha$*  paralog subgroup in coelacanth, *LmPcdh $\alpha 1$* - $\alpha 10$ , has no equivalent in mammals. On the basis of the tree topologies in Figure 2, *LmPcdh $\alpha 1$* - $\alpha 10$  appear to be descended from the same ancestral protocadherin as *LmPcdh $\alpha 21$*  and human *Pcdh $\alpha C 2$* . However, *LmPcdh $\alpha 1$* - $\alpha 10$  and *LmPcdh $\alpha 21$*  are considerably diverged in sequence, copy number, and physical position within the coelacanth *Pcdh $\alpha$*  cluster (Figs. 1, 4). In addition, coelacanth *Pcdh $\alpha 1$*  and zebrafish *Pcdh1 $\alpha 1$*  are orthologous, indicating that this paralog subgroup predates the divergence of ray-finned and lobe-finned fishes (Figs. 2B, 4). Given their physical proximity, coelacanth *Pcdh $\alpha 2$* - $\alpha 10$  may have arisen through tandem duplication and subsequent diversification in the coelacanth lineage, or else their zebrafish orthologs have not been maintained. Coelacanth *Pcdh $\alpha 8$*  and *Pcdh $\alpha 10$*  are clearly the products of a recent duplication (Fig. 2A), indicating that gene duplication continues to generate new coelacanth protocadherin paralogs. We searched the *Xenopus tropicalis* genome assembly (v1.0; <http://genome.jgi-psf.org/xenopus/>) for paralogs in this class by using TBLASTN, and found one predicted protein that is ~70% identical to coelacanth *Pcdh $\alpha 1$* . This predicted protein appears to be orthologous to coelacanth *Pcdh $\alpha 1$*  when included in a maximum likelihood phylogeny of zebrafish, coelacanth, and human *Pcdh $\alpha$*  proteins (data not shown). The protocadherin paralog subgroup including coelacanth *Pcdh $\alpha 1$* - $\alpha 10$  and zebrafish *Pcdh1 $\alpha 1$*  was evidently lost subsequent to the emergence of tetrapods, but prior to the mammalian radiation. Functional differences between these two paralog subgroups may contribute to adaptive differences in coelacanth and mammalian brain development.

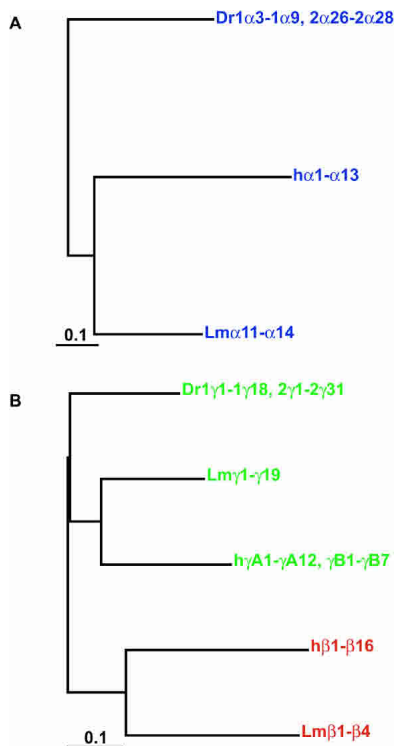
The phylogeny of coelacanth *Pcdh $\gamma$*  proteins is straightforward (Figs. 3, 4). Coelacanth *Pcdh $\gamma 1$* - $\gamma 19$  and human *Pcdh $\gamma A 1$* -

$\gamma A 12$  and  $\gamma B 1$ - $\gamma B 7$  are derived from the same ancestral *Pcdh $\gamma$*  paralog, as are coelacanth *Pcdh $\gamma 20$* - $\gamma 24$  and human *Pcdh $\gamma C 3$* - $\gamma C 5$ . Coelacanth *Pcdh $\gamma 20$* ,  $\gamma 21$ , and  $\gamma 23$  are orthologous to human  $\gamma C 3$ ,  $\gamma C 4$ , and  $\gamma C 5$ , respectively. These orthologous relationships are not evident between human and zebrafish *Pcdh $\gamma$*  proteins (Fig. 4). The division of mammalian *Pcdh $\gamma$*  proteins into  $\gamma A$  and  $\gamma B$  subtypes is absent in coelacanth. Coelacanth nevertheless has a diverse *Pcdh $\gamma$*  repertoire. On the basis of their physical distribution and their relationships in the protein tree, coelacanth *Pcdh $\gamma$*  paralogs arose through multiple tandem-duplication events. The most recent duplications were physically localized and generated *Pcdh $\gamma 13$*  through *Pcdh $\gamma 16$* , which are very similar to one another. However, most related coelacanth *Pcdh $\gamma$*  paralogs, such as *Pcdh $\gamma 1$*  and *Pcdh $\gamma 9$*  or *Pcdh $\gamma 5$*  and *Pcdh $\gamma 8$* , are much more divergent and are dispersed throughout the *Pcdh $\gamma$*  cluster, indicating that they are the product of ancient duplications (Fig. 3).

Protocadherin  $\beta$  genes likely arose by duplication and sequence divergence of existing *Pcdh $\gamma$*  genes at some undetermined point in vertebrate evolution (Wu et al. 2001; Noonan et al. 2004). Coelacanth and humans each have a *Pcdh $\beta$*  cluster, whereas zebrafish apparently do not (Figs. 3–5). However, there are only four functional coelacanth *Pcdh $\beta$*  paralogs, versus 16 in human and 22 in mouse. The four *LmPcdh $\beta$*  paralogs appear to be the products of ancient tandem duplications (Fig. 3). There are nearly as many *Pcdh $\beta$*  pseudogenes as functional *Pcdh $\beta$*  genes in coelacanth, suggesting that the common ancestor of coelacanth and tetrapods had a substantial *Pcdh $\beta$*  repertoire that has expanded in tetrapods and decayed in coelacanth. The most obvious effect of this expansion is that the protocadherin repertoire in mammals is more diverse than that in coelacanth. The maintenance and diversification of so many duplicate genes in mammals could provide some adaptive benefit, most likely by allowing a greater number and variety of *Pcdh $\beta$* -mediated interactions between neurons, both at synapses and elsewhere, in the developing brain. The absence of  $\beta$  protocadherins in zebrafish could be due to secondary loss in teleosts, or the *Pcdh $\beta$*  cluster could have arisen after the separation of teleosts from lobe-finned fishes. In either event, depending solely on zebrafish sequence as a reference would lead to the faulty conclusion that  $\beta$  protocadherins are purely tetrapod-specific genes.



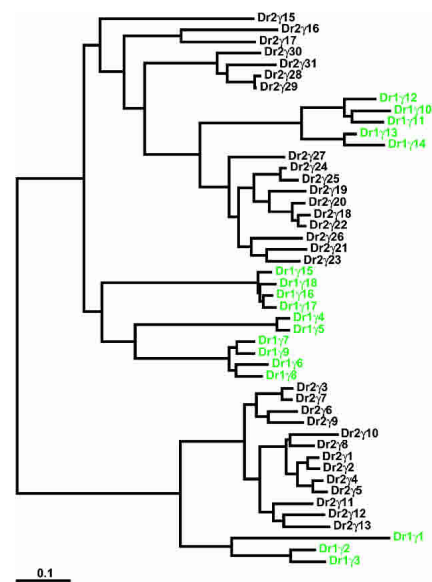
**Figure 4.** Comparison of the human (Hs), coelacanth (Lm), and zebrafish (Dr1 and Dr2) protocadherin clusters. The known phylogeny is shown at left. The yellow circle indicates the teleost whole-genome duplication. Genes in each paralog subgroup are indicated by color and connecting bars. Individual orthologous relationships are indicated by narrow lines. Zebrafish-specific paralog subgroups in *DrPcdh2 $\alpha$*  are labeled. Sequencing of *DrPcdh2 $\gamma$*  is sufficiently complete to determine that  $\beta$  protocadherins are absent. *DrPcdh2 $\gamma$*  exons are numbered on the basis of their order in the genomic contig BX005294, not on their order in the genome.



**Figure 5.** Phylogenetic relationship of coelacanth (Lm), zebrafish (Dr), and human Pcdh $\alpha$  (A) and Pcdh $\gamma$  (B) proteins. Subtrees have been collapsed for clarity, so terminal branch lengths are approximate. Colors refer to paralog subgroups shown in Figures 1 and 2.

The teleost whole-genome duplication event has radically altered the overall protocadherin complement in zebrafish relative to coelacanth and mammals. Zebrafish has two highly divergent protocadherin clusters: *DrPcdh1*, consisting of 38  $\alpha$  and  $\gamma$  protocadherins, and *DrPcdh2*, which has at least 59 genes (Noonan et al. 2004; Fig. 4). Sequencing of *DrPcdh2* is incomplete, but sufficient sequence is available to indicate that more  $\alpha$  and  $\gamma$  protocadherins, and no  $\beta$  protocadherins, will be identified in the finished sequence. Ohno postulated that, following gene duplication, one duplicate would evolve rapidly relative to the other, because only one copy is necessarily under functional constraint (Ohno 1970). The massive expansion and diversification of protocadherins in zebrafish relative to coelacanth is an excellent illustration of this concept. Comparison of coelacanth and zebrafish protocadherin cluster sequences suggests that *DrPcdh2* is considerably more divergent relative to the preduplication ancestor than is *DrPcdh1* (Fig. 4). In addition, two *DrPcdh2* protocadherin  $\alpha$  paralog subgroups, *Pcdh2 $\alpha$ 1-2 $\alpha$ 7* and *Pcdh2 $\alpha$ 8-2 $\alpha$ 25*, have no counterpart in coelacanth or human protocadherin clusters (Fig. 4). These subgroups are ultimately descended from redundant ancestral paralogs generated by the whole-genome duplication. The absence of functional constraint on these duplicates evidently allowed them to accumulate diversifying substitutions, and additional tandem duplications have greatly increased the number of paralogs in each subgroup. Many of the paralogs in the *Pcdh2 $\alpha$ 8-2 $\alpha$ 25* subgroup are also ancient duplicates, and the maintenance of these paralogs indicates that they have acquired a teleost-specific function (Noonan et al. 2004).

Zebrafish also has at least 49 *Pcdh $\gamma$*  genes, although all of these are paralogous to human and coelacanth  $\gamma$  protocadherins (Figs. 4–6). The maintenance of so many paralogs since the whole-genome duplication represents a massive expansion in *Pcdh $\gamma$*  gene content in zebrafish relative to other species. Although many of these genes appear to be the products of recent duplications (Fig. 6), they engage in frequent gene conversion events, as we discuss below. Gene conversion reduces sequence diversity among paralogs and can thereby make ancient paralogs appear to be recent duplicates. It is therefore difficult to estimate the age of zebrafish *Pcdh $\gamma$*  paralogs on the basis of sequence similarity. In addition, there are no *Pcdh2 $\gamma$*  pseudogenes evident in zebrafish and very few *DrPcdh2* pseudogenes overall. This implies that many *Pcdh2 $\gamma$*  genes are under purifying selection, despite their apparent similarity. However, gene conversion between highly similar paralogs could also be acting to repair inactivating substitutions, thus causing both paralogs to be maintained when only one is constrained. Interestingly, although there is some indication that groups of *Pcdh1 $\gamma$*  and *Pcdh2 $\gamma$*  paralogs are descended from a common preduplication ancestor, there are apparently no remaining *Pcdh1 $\gamma$*  and *Pcdh2 $\gamma$*  paralogs that arose directly from the whole-genome duplication (Fig. 6). Therefore, many modern zebrafish protocadherin paralogs seem to be the products of tandem duplications occurring subsequent to the whole-genome duplication event. Two possibilities arise from this observation. Zebrafish *Pcdh $\gamma$*  genes may be generated through a continuous process of gene duplication, in which case, many of the current, highly similar zebrafish *Pcdh $\gamma$*  genes will be lost and subsequently replaced by new, equally disposable duplicates. Alternatively, many zebrafish *Pcdh $\gamma$*  genes may have arisen in a burst of tandem duplication occurring after the whole-genome duplication, with additional duplications occurring at a lower frequency since that event. In this scenario, many zebrafish *Pcdh $\gamma$*  genes are functionally constrained. A comparison of zebrafish and pufferfish *Pcdh $\gamma$*  genes would help re-



**Figure 6.** Massive expansion of the Pcdh $\gamma$  paralog subgroup in zebrafish. *DrPcdh1 $\gamma$*  paralogs (see Fig. 4) are shown in green. *DrPcdh2 $\gamma$*  paralogs from BX005294 are shown in black. *Pcdh2 $\gamma$*  exons are numbered relative to the clone sequence. There are additional zebrafish *Pcdh2 $\gamma$*  paralogs in draft sequence 5' of *Pcdh2 $\gamma$ 1* and 3' of *Pcdh2 $\gamma$ 31*.

solve this issue. However, whole-genome shotgun sequence, which comprises the current *Fugu* and tetraodon genome builds, is inadequate for assembling contiguous protocadherin cluster sequences, due to the highly repetitive nature of the genes (data not shown). Any inferences as to protocadherin cluster organization or paralog content that we could make from current *Fugu* and tetraodon genome assemblies would therefore be unreliable.

### Low frequency of gene conversion events in coelacanth protocadherin cluster genes

We recently determined that paralogous protocadherin cluster genes in zebrafish and mammals undergo gene conversion events, resulting in ectodomain-specific sequence homogenization (Noonan et al. 2004). To compare the frequencies of gene conversion in human, coelacanth, and zebrafish protocadherin subgroups, we aligned the full-length extracellular domains and individual ectodomains of all paralogs in each subgroup and estimated the total number of synonymous substitutions per codon (dS) in the gene tree generated from each alignment (Table 1; see Methods). Our results indicate that coelacanth protocadherins rarely engage in gene conversion. Coelacanth *Pcdhγ1-γ19* and zebrafish *Pcdh2γ15-2γ31* have similar overall estimated neutral substitution rates (Table 1). However, the distribution of substitutions among the ectodomains in each subgroup is very different (Fig. 7). The ratio of substitution rates for the most divergent and least divergent ectodomains in *LmPcdhγ1-γ19* is 1.59, indicating a uniform distribution of synonymous-site diversity among the ectodomains in this class. In contrast, the estimated total neutral substitution rate in *DrPcdh2γ15-2γ31* ectodomain 6 is zero. The estimated substitution rate for the most divergent ectodomain in this class is 53.32 total substitutions per codon. *DrPcdh2α8-2α25* shows a similar substitution rate distribution. At this level of diversity, the variance in these estimates is considerable. Nevertheless, these results indicate that gene conversion

**Table 1.** Total number of synonymous substitutions per codon in full-length extracellular domains and the most divergent and least divergent ectodomains from various human, coelacanth and zebrafish protocadherin paralog subgroups

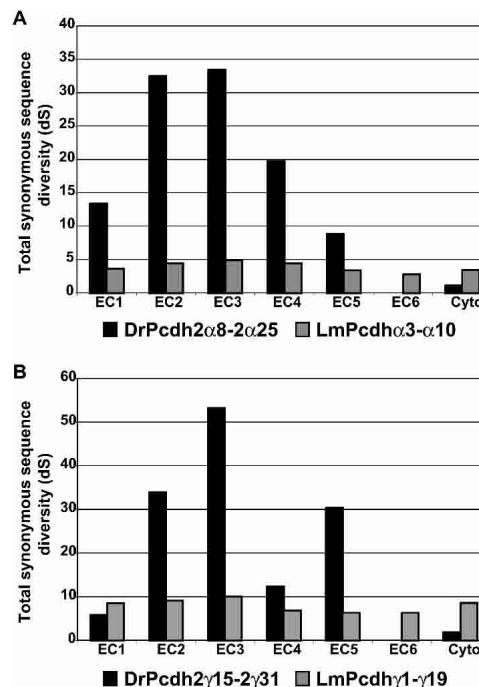
Subgroup	dS <sub>EC</sub> <sup>a</sup>	dS <sub>EC/branch</sub> <sup>b</sup>	dS <sub>high</sub> <sup>c</sup>	dS <sub>low</sub> <sup>d</sup>	Ratio (dS <sub>high</sub> /dS <sub>low</sub> )
hα1-α13	4.68	0.20	22.29	1.56	14.29
hβ2-β15	4.64	0.17	9.41	1.27	7.40
hγA1-γA12	5.43	0.26	9.55	2.53	3.78
mα1-α12	5.40	0.26	23.90	1.34	17.77
mβ2-β22	8.02	0.21	13.63	5.77	2.36
mγA1-γA12	7.23	0.34	12.70	4.42	2.87
Dr1γ4-1γ18	5.44	0.20	14.32	0.15	94.62
Dr1γ19-1γ27	5.24	0.35	20.37	0.53	38.42
Dr2α8-2α25	9.37	0.28	33.46	0	INF
Dr2γ15-2γ31	6.02	0.19	53.32	0	INF
<b>Lmα3-α10</b>	<b>3.64</b>	<b>0.28</b>	<b>4.89</b>	<b>2.80</b>	<b>1.75</b>
<b>Lmγ1-γ19</b>	<b>7.76</b>	<b>0.22</b>	<b>10.04</b>	<b>6.33</b>	<b>1.59</b>

<sup>a</sup>Total synonymous substitutions per codon (dS) calculated for the alignment of full-length extracellular domain sequences for each paralog subgroup.

<sup>b</sup>The average number of substitutions per codon for each branch in the gene tree of each subgroup.

<sup>c</sup>Total synonymous substitutions per codon calculated for the alignment of the most divergent ectodomain in each subgroup.

<sup>d</sup>Total synonymous substitutions per codon calculated for the alignment of the least divergent ectodomain in each subgroup.



**Figure 7.** Coelacanth protocadherins rarely undergo gene conversion events. Distribution of synonymous substitution rates across ectodomains 1 through 6 and the cytoplasmic domain for zebrafish *Pcdh2α* and coelacanth *Pcdhα* (A) and zebrafish *Pcdh2γ* and coelacanth *Pcdhγ* (B) paralog subgroups.

causes much of the overall sequence diversity among zebrafish protocadherins, but not coelacanth protocadherins, to be sequestered in particular ectodomains.

Because the effective substitution rate in each ectodomain is inversely proportional to the rate of gene conversion, the conversion rate in coelacanth protocadherins must be very low. The high rate of gene conversion in zebrafish protocadherins may reflect a recombination-driven propensity for frequent tandem duplication and gene conversion in teleosts in the wake of the whole-genome duplication. Unlike zebrafish, and to a lesser extent, mammals, coelacanth appears to tolerate repetitive gene sequences without frequent tandem duplication, gene loss, or conversion. The coelacanth genome may also have a lower overall substitution rate compared with mammalian and teleost genomes. Coelacanth protocadherins with identifiable zebrafish or human orthologs have accumulated fewer amino acid substitutions per site relative to the common ancestor than their zebrafish or human counterparts (Figs. 2, 3). For example, in the ML protein tree of coelacanth *Pcdhα21*, human *PcdhαC2*, and zebrafish *Pcdh1α10* (Fig. 2B), coelacanth *Pcdhα21* has diverged less from its ortholog in the last common ancestor of mammals and lobe-finned fishes than has human *PcdhαC2* (0.14 vs. 0.36 substitutions per site). Coelacanth *Pcdhα21* shows a similar low substitution rate relative to zebrafish *Pcdh1α10* (0.20 vs. 0.32 substitutions per site since the last common ancestor). A recent study comparing *Rag1* and *Rag2* genes from coelacanth and many other vertebrate species also found a low substitution rate in coelacanth relative to teleosts and mammals (Brinkmann et al. 2004). In this study and in our results, zebrafish proteins show a relatively high rate of amino acid replacement compared with

their orthologs in other vertebrate species, consistent with a relaxation of selective constraint following the whole-genome duplication.

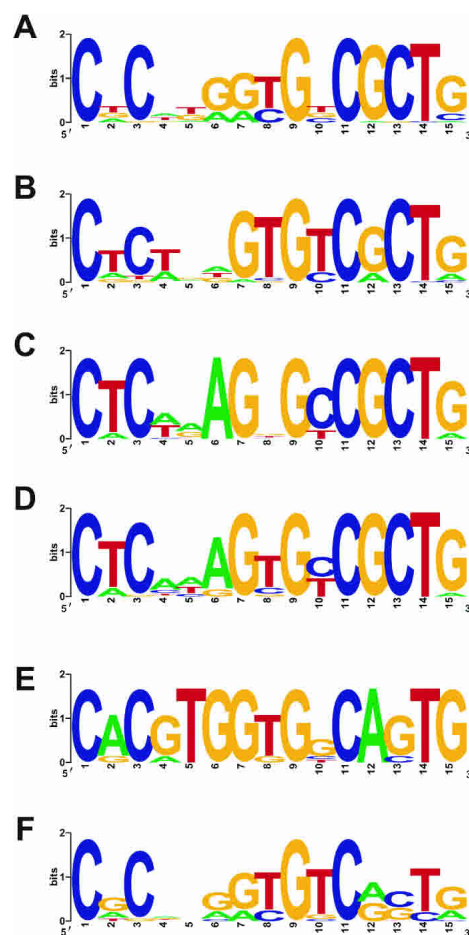
In addition, numerous diversifying speciation events, and episodes of adaptive evolution associated with the teleost radiation, have occurred on the teleost lineage since the divergence of teleosts and lobe-finned fishes. Coelacanth, however, appear to be evolving largely under purifying selection, with relatively few recent adaptive changes in the lineage. These historical differences undoubtedly contribute to the apparent substitution rate difference between the lineages. Differences in generation time and effective population size between coelacanth and teleost species will also yield different evolutionary rates. Unlike zebrafish, which have external fertilization, a generation time of 3 mo, and which produce hundreds of offspring per mating, coelacanths give birth to live pups and may have long generation times and small brood sizes as a result (Heemstra and Greenwood 1992). Long generation times mean fewer opportunities for introducing genetic change into the population, and thus, a slower rate of molecular evolution. The high substitution rate in mammalian protocadherins relative to their coelacanth orthologs may be due to similar processes operating on the lineage leading to placental mammals. The relative stability of the coelacanth genome makes it an ideal reference against which the derived genomes of mammals and teleosts can be compared.

Our results also suggest that gene conversion can act to increase diversity among protocadherin cluster genes. Protocadherin proteins are believed to be homophilic adhesion molecules, a mechanism that requires sequence diversity among paralogs. Particular ectodomains could determine specificity in homophilic interactions, and would therefore show the greatest sequence diversity. The substitution rate variation we observe in both zebrafish and mammalian protocadherins is consistent with a model of gene conversion in which long conversion events completely homogenize ectodomains where diversity is not required, whereas short conversion events shuffle sequences among necessarily diverse domains, thereby inflating the substitution rate. This may explain the extreme diversity of some zebrafish *Pcdh2α8-2α25* and *Pcdh2γ15-2γ31* ectodomains and the complete absence of substitutions in other ectodomains in these subgroups (Fig. 7; Table 1). The accumulation of gene conversion tracts eventually leads to the deterioration of sequence similarity between orthologs. Whole-genome or tandem duplications, which introduce redundant paralogs free to accumulate conversion tracts, would greatly accelerate this process. However, the coelacanth protocadherin cluster, subject mostly to individual tandem duplications and neutral substitutions, is likely to have remained similar to the protocadherin cluster in the true tetrapod ancestor.

#### Regulatory element conservation in coelacanth and human protocadherin clusters

A major aim in comparative sequence analysis is to use sequence conservation to identify functionally constrained regulatory elements. All mammalian protocadherin cluster promoters share a 15-bp core sequence element (Wu et al. 2001). This element is also present in all zebrafish protocadherin promoters (Noonan et al. 2004). However, comparing consensus protocadherin core promoter element sequences from coelacanth, human, and ze-

brafish illustrates the effect of whole-genome and tandem duplication on regulatory element evolution. The 15-bp core promoter motif is well conserved among human and coelacanth protocadherin promoters (Fig. 8A,B). The coelacanth and human motifs are virtually identical, particularly the CGCT element, which has been implicated in promoter function (Tasic et al. 2002). This motif is highly conserved in zebrafish *Pcdh1γ* and *Pcdh2γ* promoters (Fig. 8C,D). However, the CGCT motif in *DrPcdh1α* and *DrPcdh2α* promoters is divergent (Fig. 8E,F). As mentioned above, zebrafish  $\alpha$  protocadherins are considerably more diverse than coelacanth or mammalian protocadherins due to whole-genome and tandem duplication. Many of these zebrafish *Pcdhα* genes are the product of ancient duplications and are clearly under purifying selection (Noonan et al. 2004). Zebrafish  $\alpha$  protocadherins that have a novel, teleost-specific function may also have acquired a new expression pattern through substitutions in the core promoter element. These substitutions are not likely to inactivate the promoters, as the genes are constrained. Alternatively, complementary mutations in the promoters of duplicated genes can lead to subfunctionalization and the maintenance of both duplicates by limiting the expression of each duplicate to a



**Figure 8.** Divergence of the zebrafish *Pcdhα* core promoter element following whole-genome and tandem duplication. WebLogo plots of consensus human *Pcdh* (A), coelacanth *Pcdh* (B), zebrafish *Pcdh1γ* (C), *Pcdh2γ* (D), *Pcdh1α* (E), and *Pcdh2α* (F) core promoter motifs identified by using MEME.

subset of cells or conditions relative to the ancestor. These processes make it difficult to call regulatory elements tetrapod specific on the basis of their absence from the zebrafish genome, as there may be many cases in which the divergent copy of an ancient element has been retained and the original lost. Comparisons that use coelacanth genome sequence are not subject to this limitation, and will therefore capture more ancient regulatory elements and reliably identify elements specific to tetrapods.

To determine the overall level of intergenic conservation between human, mouse, coelacanth, and zebrafish protocadherin clusters, we aligned protocadherin cluster sequences from each species by using multi-LAGAN and visualized this alignment with VISTA (Mayor et al. 2000). We found many noncoding elements conserved between human and mouse that are clustered around *Pcdh $\alpha$*  and *Pcdh $\gamma$*  constant region exons. However, there is little intergenic conservation evident between human and coelacanth or human and zebrafish protocadherin cluster sequences (Supplemental Fig. S1). Human, coelacanth, and zebrafish exons show some sequence similarity, and there are a few instances of weak conservation visible between coelacanth and human that appear to correlate with regions of human–mouse conservation, but none of the noncoding elements conserved between human and mouse are detectable in coelacanth or zebrafish by BLAST search (data not shown). These elements are likely to be tetrapod or mammalian innovations, a conclusion made more certain by their absence from a basal lineage—coelacanth—as opposed to the highly derived teleost lineage, in which ancestral sequences are more subject to frequent secondary loss or adaptation.

## Conclusions

Our results demonstrate that *L. menadoensis* is a vitally important species for understanding the evolution of tetrapod genomes, particularly in regard to the identification of tetrapod-specific genomic features. Using the protocadherin cluster as a measure of genome stability, it appears that coelacanth has little history or propensity for whole-genome duplication or frequent tandem gene duplications. Coelacanth protocadherin proteins have also accumulated fewer amino acid substitutions relative to their zebrafish and human orthologs. The modern coelacanth genome therefore provides access to the state of the sarcopterygian genome just prior to the emergence of tetrapods. However, whole-genome duplication and the subsequent adaptive radiation of teleost species have radically altered teleost genome content relative to the common ancestor of coelacanths and ray-finned fishes. Analyses that use the zebrafish or *Fugu* genomes alone as references to identify tetrapod-specific regulatory and coding elements will founder on this divergence. The absence of a particular regulatory or coding feature in teleosts may reflect teleost-specific rearrangements or secondary loss of an ancestral element, rather than indicate the rise of a tetrapod-specific functional element. The coelacanth genome, although subject to lineage-specific sequence changes like every genome, nevertheless appears to be much more stable than any of the teleost genomes, and we can therefore be more confident that a regulatory element found in tetrapods and not found in coelacanth and zebrafish is a tetrapod innovation. In this regard, the most informative analysis is the comparison of coelacanth, tetrapod, and teleost genome sequences. Coelacanth provides the reference against which tetrapod and teleost genomes can be measured,

allowing the identification of genomic features that drive tetrapod and teleost species diversity. A complete genome sequence of *L. menadoensis* would therefore be an extremely valuable tool for understanding vertebrate evolution.

## Methods

### *Latimeria* BAC isolation and sequencing

We designed degenerate PCR primers against zebrafish protocadherin  $\gamma$  variable exons and amplified two distinct protocadherin sequences from *L. chalumnae* genomic DNA. We sequenced these amplicons and designed overgo oligonucleotides to probe the VMRC-4 *L. menadoensis* BAC library (Danke et al. 2004). The overgo labeling and hybridization conditions were as previously described (McPherson et al. 2001). We identified five minimally overlapping clones as follows: 40c18, 188c23, 44h8, 39g19, and 24c12 (Fig. 1). The Stanford Human Genome Center Sequencing Group sequenced these clones as described previously (Noonan et al. 2004).

### Protocadherin cluster gene prediction and sequence annotation

We used TBLASTN (Altschul et al. 1997) and OrfFinder (<http://www.ncbi.nlm.nih.gov/>) to identify large single-exon genes encoding proteins similar to human *Pcdh* variable and constant region protein sequences in assembled coelacanth BAC sequence as described previously (Noonan et al. 2004). We identified coelacanth *Pcdh $\alpha$*  and *Pcdh $\gamma$*  constant region exons by using TBLASTN with human *Pcdh $\alpha$*  and *Pcdh $\gamma$*  constant region protein sequence, and we estimated splice sites by visual inspection according to the GT-AG rule in a manner that yielded an in-frame predicted transcript. We extracted, managed, and translated protocadherin exon sequences and annotated genomic sequence by using custom Perl scripts. We searched zebrafish finished and whole-genome shotgun sequence ([http://www.sanger.ac.uk/Projects/D\\_rerio/](http://www.sanger.ac.uk/Projects/D_rerio/)) for contigs and clones containing protocadherin exons additional to those we discovered previously (Noonan et al. 2004). We identified one finished clone (BX005294) that contained 31 complete *DrPcdh2 $\gamma$*  exons. This clone maps to zebrafish supercontig 14327, as does *DrPcdh2 $\alpha$* . Draft sequence from this contig indicates that *DrPcdh2 $\gamma$*  is located directly 3' of the *DrPcdh2 $\alpha$*  constant region, with no intervening *Pcdh $\beta$*  cluster.

### Phylogenetic analysis of coelacanth, human, and zebrafish protocadherins

To identify paralog subgroups in coelacanth, we initially made two CLUSTALW alignments of coelacanth and human protocadherin extracellular domain protein sequences, the first including human and coelacanth *Pcdh $\alpha$* , and the second, *Pcdh $\beta$*  and  $\gamma$  proteins (<http://www.ebi.ac.uk/clustalw/>). Because extracellular domains in all human, coelacanth, and zebrafish protocadherin paralogs are of nearly identical length, the resulting alignments have very few gaps. We then used these alignments to estimate maximum likelihood phylogenies in SEMPHY v1.0 using the amino acid substitution matrix of D.T. Jones, W.K. Taylor, and J.M. Thornton under default parameters (Jones et al. 1992; Friedman et al. 2002; Figs. 2, 3). We then made individual CLUSTALW alignments of human, zebrafish, and coelacanth protocadherin protein extracellular domain sequences for each of the coelacanth paralog subgroups identified in the coelacanth–human comparison. We generated ML trees from these alignments using SEMPHY, as above. The resulting trees show the nearest zebrafish

and human relatives for each coelacanth Pcdh paralog subgroup (Fig. 5). We rendered and edited trees in TreeEdit (<http://evolve.zoo.ox.ac.uk/>). To test for evidence of gene conversion in coelacanth protocadherin cluster genes, we identified protocadherin ectodomains in each protein with HMMER2.2 as described (Durbin et al. 1998; Noonan et al. 2004; <http://hmmer.wustl.edu/>). We built nucleotide alignments of each ectodomain with RevTrans, a Python application that aligns coding sequences based on the protein alignment (Wernersson and Pedersen 2003; <http://www.cbs.dtu.dk/services/RevTrans/>). We estimated ML gene trees in SEMPY using the Kimura 2-parameter model of nucleotide substitution with a transition-transversion ratio of 2. We estimated synonymous and nonsynonymous substitution rates for each tree by using CODEML (Yang 1997; <http://abacus.gene.ucl.ac.uk/software/paml.html>).

### Motif searches

We used MEME (Bailey and Elkan 1994; <http://meme.sdsc.edu/meme/website/intro.html>) to identify coelacanth protocadherin cluster promoter motifs as described (Noonan et al. 2004). To identify variable exon splice donor sites, we used MEME to search 150 bp of coding sequence directly upstream of each coelacanth and zebrafish variable exon stop codon. We compared the predicted coelacanth and zebrafish splice site motifs with the known consensus human variable exon splice donor site (Wu and Maniatis 1999) and made logograms for all three with WebLogo (Crooks et al. 2004; <http://weblogo.berkeley.edu/logo.cgi>).

### Global multiple-sequence alignments

We used multi-LAGAN (Brudno et al. 2003; <http://lagan.stanford.edu/>) with translated anchoring to align human, mouse, coelacanth, and zebrafish protocadherin cluster sequences. The multi-LAGAN web server outputs a VISTA plot of the multiple-sequence alignment on the basis of user-specified percent identity and window-size values. No human-coelacanth or human-zebrafish intergenic conservation was evident at our settings (75%, 75-bp window size). Human and mouse protocadherin cluster sequences, however, show extensive conservation at these values (Supplemental Fig. S1).

### Acknowledgments

We thank the Stanford Human Genome Center Sequencing Group for generating the primary coelacanth sequence data for this study. Dr. Arend Sidow provided insightful comments on the potential value of a coelacanth genome sequence to comparative studies. We thank the members of the Myers lab for discussions and support.

### References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., et al. 1998. Zebrafish *hox* clusters and vertebrate genome evolution. *Science* **282**: 1711–1714.
- Bailey, T. and Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the second international conference on intelligent systems for molecular biology*, pp. 28–36, AAAI Press, Menlo Park, CA.
- Brinkmann, H., Venkatesh, B., Brenner, S., and Meyer, A. 2004. Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. *Proc. Natl. Acad. Sci.* **101**: 4900–4905.
- Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Green, E.D., Sidow, A., Batzoglou, S., and the NISC Comparative Sequencing Program. 2003. LAGAN and multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**: 721–731.
- Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. 2004. WebLogo: A sequence logo generator. *Genome Res.* **14**: 1188–1190.
- Danke, J., Miyake, T., Powers, T., Schein, J., Shin, H., Bosdet, I., Erdmann, M., Caldwell, R., and Amemiya, C.T. 2004. Genome resource for the Indonesian coelacanth, *Latimeria menadoensis*. *J. Exp. Zool.* **301A**: 228–234.
- Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. 1998. *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge University Press, Cambridge, UK.
- Forey, P.L. 1998. *History of the coelacanth fishes*. Chapman and Hall, New York.
- Friedman, N., Ninio, M., Pe'er, I., and Pupko, T. 2002. A structural EM algorithm for phylogenetic inference. *J. Comput. Biol.* **9**: 331–353.
- Gorr, T., Kleinschmidt, T., and Fricke, H. 1991. Close tetrapod relationships of the coelacanth *Latimeria* indicated by haemoglobin sequences. *Nature* **351**: 394–397.
- Heemstra, P.C. and Greenwood, P.H. 1992. New observations on the visceral anatomy of the late-term fetuses of the living coelacanth fish and the oophagy controversy. *Proc. R. Soc. Lond. B.* **249**: 49–55.
- Holder, M.T., Erdmann, M.V., Wilcox, T.P., Caldwell, R.L., and Hillis, D.M. 1999. Two living species of coelacanths? *Proc. Natl. Acad. Sci.* **96**: 12616–12620.
- Jones, D.T., Taylor, W.R., and Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**: 275–282.
- Koh, E.G.L., Lam, K., Christoffels, A., Erdmann, M.V., Brenner, S., and Venkatesh, B. 2003. *Hox* gene clusters in the Indonesian coelacanth, *Latimeria menadoensis*. *Proc. Natl. Acad. Sci.* **100**: 1084–1088.
- Kohmura, N., Senzaki, K., Hamada, S., Kai, N., Yasuda, R., Watanabe, M., Ishii, H., Yasuda, M., Mishina, M., and Yagi, T. 1998. Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron* **20**: 1137–1151.
- Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., and Dubchak, I. 2000. VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**: 1046–1047.
- McPherson, J.D., Marra, M., Hiller, L., Waterson, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K., et al. 2001. A physical map of the human genome. *Nature* **409**: 934–941.
- Noonan, J.P., Grimwood, J., Schmutz, J., Dickson, M., and Myers, R.M. 2004. Gene conversion and the evolution of protocadherin gene cluster diversity. *Genome Res.* **14**: 354–366.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag, Berlin, Germany.
- Postlethwait, J.H., Yan, Y.L., Gates, M.A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E.S., Force, A., Gong, Z., et al. 1998. Vertebrate genome evolution and the zebrafish gene map. *Nat. Genet.* **18**: 345–349.
- Smith, J.L.B. 1939. A surviving fish of the order *Actinistia*. *Trans. Roy. Soc. S. Afr.* **27**: 47–50.
- Takezaki, N., Figueroa, F., Zaleska-Rutczynska, Z., Takahata, N., and Klein, J. 2004. The phylogenetic relationship of tetrapod, coelacanth and lungfish revealed by the sequences of 44 nuclear genes. *Mol. Biol. Evol.* **21**: 1512–1524.
- Tasic, B., Nabholz, C.E., Baldwin, K.K., Kim, Y., Rueckert, E.H., Ribich, S.A., Cramer, P., Wu, Q., Axel, R., and Maniatis, T. 2002. Promoter choice determines splice site selection in protocadherin  $\alpha$  and  $\gamma$  pre-mRNA splicing. *Mol. Cell* **10**: 21–33.
- Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, V. 2003. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* **13**: 382–390.
- Wang, X., Weiner, J.A., Levi, S., Craig, A.M., Bradley, A., and Sanes, J.R. 2002. Gamma protocadherins are required for survival of spinal interneurons. *Neuron* **36**: 843–854.
- Wernersson, R. and Pederson, A.G. 2003. RevTrans—Constructing alignments of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.* **31**: 3537–3539.
- Wu, Q. and Maniatis, T. 1999. A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* **97**: 779–790.
- Wu, Q., Zhang, T., Cheng, J.F., Kim, Y., Grimwood, J., Schmutz, J., Dickson, M., Noonan, J.P., Zhang, M.Q., Myers, R.M., et al. 2001. Comparative DNA sequence analysis of mouse and human protocadherin clusters. *Genome Res.* **11**: 389–404.
- Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**: 555–556.

Zardoya, R. and Meyer, A. 1996. Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene. *Proc. Natl. Acad. Sci.* **93**: 5449–5454.  
———. 1997. The complete DNA sequence of the mitochondrial genome of a “living fossil,” the coelacanth (*Latimeria chalumnae*). *Genetics* **146**: 995–1010.

### Web site references

<http://lagan.stanford.edu/>; LAGAN.  
<http://hmmmer.wustl.edu/>; HMMER.  
<http://abacus.gene.ucl.ac.uk/software/paml.html>; PAML.  
<http://www.cbs.dtu.dk/services/RevTrans/>; RevTrans.

<http://weblogo.berkeley.edu/logo.cgi>; WebLogo.  
<http://meme.sdsc.edu/meme/website/intro.html>; MEME.  
[http://www.sanger.ac.uk/Projects/D\\_rerio/](http://www.sanger.ac.uk/Projects/D_rerio/); Zebrafish whole genome and shotgun sequences.  
<http://www.ncbi.nlm.nih.gov/>; GenBank and OrfFinder.  
<http://www.ebi.ac.uk/clustalw/>; Clustalw.  
<http://www-shgc.stanford.edu/>; Stanford Human Genome Center.  
<http://www-shgc.stanford.edu/myerslab/>; Myers Lab.  
<http://evolve.zoo.ox.ac.uk/>; TreeEdit.  
<http://genome.jgi-psf.org/xenopus/>; *Xenopus tropicalis* genome assembly v.1.0.

Received July 13, 2004; accepted in revised form September 17, 2004.