



## A Genetic Linkage Map for Zebrafish: Comparative Analysis and Localization of Genes and Expressed Sequences

Michael A. Gates, Lisa Kim, Elizabeth S. Egan, et al.

*Genome Res.* 1999 9: 334-347

Access the most recent version at doi:[10.1101/gr.9.4.334](https://doi.org/10.1101/gr.9.4.334)

---

**References** This article cites 50 articles, 22 of which can be accessed free at:  
<http://genome.cshlp.org/content/9/4/334.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

# A Genetic Linkage Map for Zebrafish: Comparative Analysis and Localization of Genes and Expressed Sequences

Michael A. Gates,<sup>1,3</sup> Lisa Kim,<sup>1</sup> Elizabeth S. Egan,<sup>1</sup> Timothy Cardozo,<sup>1</sup> Howard I. Sirotkin,<sup>1,3</sup> Scott T. Dougan,<sup>1,3</sup> Deval Lashkari,<sup>2</sup> Ruben Abagyan,<sup>1</sup> Alexander F. Schier,<sup>1</sup> and William S. Talbot<sup>1,3,4</sup>

<sup>1</sup>Skirball Institute of Biomolecular Medicine and Department of Cell Biology, New York University Medical Center, New York, New York 10016 USA; <sup>2</sup>Stanford DNA Sequencing and Technology Center, Palo Alto, California 94305 USA

Genetic screens in zebrafish (*Danio rerio*) have isolated mutations in hundreds of genes with essential functions. To facilitate the identification of candidate genes for these mutations, we have genetically mapped 104 genes and expressed sequence tags by scoring single-strand conformational polymorphisms in a panel of haploid siblings. To integrate this map with existing genetic maps, we also scored 275 previously mapped genes, microsatellites, and sequence-tagged sites in the same haploid panel. Systematic phylogenetic analysis defined likely mammalian orthologs of mapped zebrafish genes, and comparison of map positions in zebrafish and mammals identified significant conservation of synteny. This comparative analysis also identified pairs of zebrafish genes that appear to be orthologous to single mammalian genes, suggesting that these genes arose in a genome duplication that occurred in the teleost lineage after the divergence of fish and mammal ancestors. This comparative map analysis will be useful in predicting the locations of zebrafish genes from mammalian gene maps and in understanding the evolution of the vertebrate genome.

A powerful combination of genetics and embryology has established the zebrafish (*Danio rerio*) as an important model organism for the analysis of vertebrate development, physiology, and behavior. Genetic screens have identified mutations in >600 genes with essential functions in the embryo (Driever et al. 1996; Haffter et al. 1996). The transparency and external development of the embryo allow exquisite manipulations, such as dye-labeling, transplantation, and in vivo time-lapse imaging, that illuminate the function of mutated genes at the cellular level (Kimmel 1989; Schier and Talbot 1998). Moreover, molecular analysis of zebrafish mutations has revealed new genes and gene functions (Postlethwait and Talbot 1997; Schier and Talbot 1998). This emphasizes both the potential of the system and the need to develop strategies and infrastructures that will facilitate the cloning of genes defined by zebrafish mutations.

Three approaches have been used to clone mutant loci in zebrafish: insertional mutagenesis, positional cloning, and the candidate approach (Postlethwait and Talbot 1997; Beier 1998). The candidate approach has been used most widely, and >10 loci have been cloned as candidates identified by criteria such as expression pattern and map position (e.g., Schulte-Merker et al.

1994; Talbot et al. 1995; Brand et al. 1996). Zebrafish cDNA sequencing projects will enhance the candidate approach by generating expressed sequence tags (ESTs) that correspond to many new genes. Expression analysis and mapping can then assess these genes as candidates for mutations. As more genes are localized, mapping a mutation becomes an effective method to test many candidates in parallel. Thus construction of gene maps for zebrafish will accelerate the molecular analysis of mutations by providing a large pool of candidates that can be efficiently evaluated with straightforward mapping experiments.

A recent study reporting the map locations of 144 zebrafish genes provided a basis for comparing the genomes of zebrafish and mammals (Postlethwait et al. 1998). Analysis of zebrafish and mammalian gene maps revealed extensive conservation of synteny—genes that are on the same chromosome (syntenic) in zebrafish tend to have syntenic orthologs in mouse and human. This finding raised the possibility that comparative analysis may predict the positions of zebrafish genes from mammalian gene maps. Mapping of additional genes will enhance comparative mapping by (1) identifying the locations and borders of segments of conserved synteny, and (2) determining the extent to which gene order is maintained within a conserved syntenic segment.

We have determined 104 new map positions of zebrafish genes by scoring single-strand conforma-

<sup>3</sup>Present address: Department of Developmental Biology, Stanford University School of Medicine, Stanford, California 94305 USA.

<sup>4</sup>Corresponding author.

E-MAIL [talbot@cmgm.stanford.edu](mailto:talbot@cmgm.stanford.edu); FAX (650) 725-7739.

tional polymorphisms (SSCPs; Brady et al. 1997; Förnler et al. 1998) in a haploid mapping panel. To allow straightforward comparison between our map and those produced with other crosses (e.g., mutant mapping crosses), we also scored 53 previously mapped genes (Postlethwait et al. 1998) and 217 simple sequence length polymorphism (SSLP; Knapik et al. 1998) markers in our mapping panel. These map positions define new candidates for mutations, and phylogenetic analysis of mapped genes defines new regions of conservation between zebrafish and human.

## RESULTS

### Map Construction

To construct a zebrafish linkage map, we scored a total of 390 PCR-based genetic polymorphisms in a mapping panel comprised of 48 individual haploid progeny of a Tü × TL female. These polymorphisms included 105 SSCP from fully sequenced cDNAs and ESTs (of which 104 were mapped, see below) that defined previously unmapped genes (Fig. 1; Table 1). We also developed SSCP to map 10 sequence-tagged sites (STSs) derived from genomic clones in BAC, PAC, and YAC vectors and other sequences (Table 2). To allow comparison between our map and others constructed with different crosses, we analyzed 275 previously mapped markers in our mapping panel. Of these markers, 217 were SSLPs (Knapik et al. 1998), 53 were SSCP linked to previously mapped genes (Table 1), and 5 were STSs associated with cloned random amplified polymorphic DNAs (RAPDs) (Table 2).

By linkage analysis and reference to previous maps (Knapik et al. 1998; Postlethwait et al. 1998), we ordered 389 of the 390 polymorphisms into a map with 25 linkage groups, each representing one zebrafish chromosome (Fig. 2). The map contains 25 gaps whose supporting lod score is <3, which corresponds to a gap of 23 cM or greater for a mapping panel of 48 haploid individuals. All of these large gaps were flanked by previously mapped markers, allowing the positions and approximate sizes of the gaps to be determined from

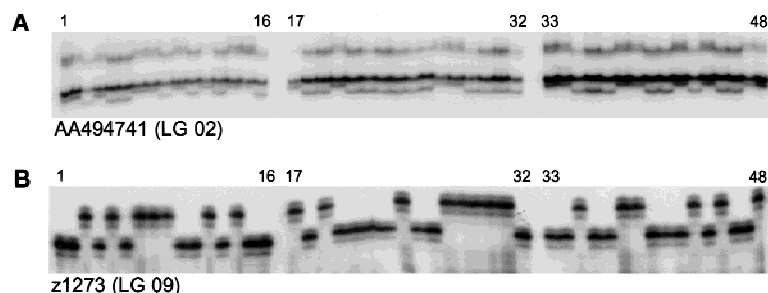
previous maps. The 389 markers occupied a total of 273 unique map positions, or bins (average of 1.42 marker/bin), where markers in each bin are separated from markers in other bins by at least one crossover. The linkage groups in the map encompassed 2894 cM, using the Kosambi mapping function to estimate the number of double-crossovers (2640 cM with no mapping function). This corresponds to 99% coverage of the genome, using 2900 cM as the total length of the female meiotic linkage map (Postlethwait et al. 1994; Johnson et al. 1996). Accordingly, we were able to assign map positions to 114 of the 115 new polymorphisms we developed, as these showed significant linkage ( $\text{lod} > 3.0$ ) to a previously mapped gene or SSLP marker.

### Error Analysis

As one method to assess the error frequency in our data set, we identified double crossovers in short intervals, which occur rarely and are thus likely to represent mis-scored genotypes. Among 18,230 individual genotype assays (an average of 46.9 scorable individuals/locus for the 389 mapped loci), there were only two double crossovers in intervals of 20 cM or less, suggesting the error rate in the data set is <0.1%. There were 20 additional double crossovers in regions >20 cM, an interval large enough to contain bona fide double crossovers. Because each crossover in the data set adds ~2.1 cM to the length of the map, the 22 total double crossovers have only a modest effect on the length of the map, expanding it by ~92 cM.

Another measure of reliability derives from a comparison of map assignments for markers common to our map and previous ones. There is very good overall agreement in the marker order shown in Figure 2 and orders reported previously by Postlethwait et al. (1998) and Knapik et al. (1998). For example, 21 of 25 linkage groups show perfect agreement in the order of 214 SSLPs and 50 genes. For two linkage groups (LG 6 and LG 17), two pairs of markers (*z1138/z3581* and *otx1/pax9*, respectively) are inverted with respect to their published positions. These are pairs of tightly linked markers whose assigned positions could be altered by one or a few genotyping errors in one of the data sets. Alternatively, these discrepancies could reflect differences in the strains used for the various maps. The other discrepancies were markers *z3054* (LG 1) and *wnt4* (LG 11), which were assigned to different positions (LG 10 and LG 15, respectively) in previous maps. We confirmed that *wnt4* mapped to LG 11 in our cross by scoring polymorphisms amplified with a total of three independent primer pairs.

Whereas there was good overall agreement of marker order, there were clear dif-



**Figure 1** Examples of polymorphisms scored in the 48 individual haploid siblings (lanes 1–48) that comprise the mapping panel. (A) SSCP derived from EST AA494741. (B) SSLP marker *z1273* (Knapik et al. 1998).

**Table 1.** Genes and ESTs Mapped by SSCP

Locus	LG	Accon. No	Forward Primer	Reverse Primer	Map Position Reference	Human Ortholog		Mouse Ortholog	
						Name	Location	Name	Location
AA494839	1	AA494839	CTGGATTAATTGTGAGACATGTAT	AGCACCAGAAGCTGGTGGTCA	a				
<i>OCT1</i>	1	AF030519	CCTTTTTCCTTAATCCATCCG	CGCTGGAAAAGAGTCTGGTATT	a				
<i>MEL1aR</i>	1	U31822	ATGACCACGAGTATTATTCCTG	CCCGCTCTGGAGAAATCGCT	a	<i>MTNR1A</i>	04q35.1	<i>Mtnr1a</i>	08
<i>DLX5</i>	1	U67849	GGGACCATATTCTAACAGAGGA	GCTACCCGTTTGTACTTGGAA	a	<i>DLX2</i>	02q32	<i>Dlx2</i>	02-44.0
<i>TCTP</i>	1	AA806157	CATCTGCTGAAATGAGAAGCC	TCTTTTAACTTCCAGAGGCCA	a	<i>TCTP</i>		<i>Trt</i>	
<i>MSXB</i>	1	U16311	TATTTAACGGACCCGTGTCC	ACGGATAACCTCCGTGTACG	a			<i>Msx3</i>	07-68.0
AA494821	1	AA494821	CTCACAGACACACTTGTACAGT	CCATAAAGCCGATCGGGTATAT	b				
AA494741	2	AA494741	GGTTGCAGCTTTAGGGAAATC	TACAGCACCCAGAGACTCC	a				
<i>TFA</i>	2		CGATATGATTGAGAGGACCTACAA	TCATTTTCATCCCTTTTGTGATA	a				
<i>RXRA</i>	2	U29940	GGAAAGCACCAGTCAAATACATG	GCTGGTGCAGTGTACACAGTT	b	<i>RXRG</i>	01q22-q23	<i>Rxrg</i>	01-88.1
<i>CPLA2</i>	2	U10330	GCTATGCAAAAAGCTACACAAA	CTAGTAATTGACITTTTTGGACA	a	<i>FLA2G4</i>	01q25	<i>Pla2g4</i>	
<i>CAH</i>	2	U55177	ATTTGTTCTCGATCCAGCCTT	AGTGCAATTCTGCAGAGGCTA	a	<i>CA2</i>	08q22	<i>Car2</i>	03-10.5
<i>TENW</i>	2	AJ001423	GAAATCCAAACACAAGCCAAA	CACATCCAGATGGAGGACATT	a				
<i>PAX3</i>	2	AF014366	GAACCAAAAATGCAATTAAGCCA	TGAGCCCAAAATAGCAAAGAGA	a	<i>PAX3</i>	02q35	<i>Pax3</i>	01-44.0
<i>RX2</i>	2	AF001908	ATATACAGCCAATGCCACCAC	ACTCCTTGGCCTTCACTCCTTA	a			<i>Rax</i>	18-38.0
AA494789	3	AA494789	GCACGCTGAAAATGGCCTAT	GGTCTTGCCTTAGTCCAGCCCT	a				
<i>EVE1</i>	3	X71845	CCTACAAACCCAAACCTTTGG	AGGTGCAAAAACAGTATTATGAAA	b				
<i>TRA1</i>	3	U54796	CACAGGAAGACGGTTCCTGC	CTGTAGTCCAGAGAGAGAGAA	b				
<i>DLX8</i>	3	U67848	GCCAGAACTCAGATGATGTSAT	CCAAGTTCAGGTTAGGCCACA	a			<i>Dlx7</i>	
AA494848	3	AA494848	AGTGGAAGGCTGCCAACCA	CTACTAGTGCAGGAGTGGCTG	a				
<i>HBAE1</i>	3		CTGCCAAAAGACAAAAGCTGCCGTC	AGCATGGAGCTCACTGAGGTTTCAG	b				
<i>HBA2r</i>	3	AA494750	AATAAGACGGAATGTTCTGCTGG	TGACCTTAAATGCCGACTCCTGAA	a	<i>HBA</i>	16pter-p13.3	<i>Hba-a2</i>	11-16.0
<i>WNT5</i>	4	U51288	GACCAGTTTGTGTGCAAGTAG	CCCCATCCTATTTCGGCCCT	b			<i>Wnt5b</i>	06
<i>B2M</i>	4	L05394	TGGTCTTCTGTGATGTTGCAAG	CCCCACCTAGTAGATCCAA	b	<i>B2M</i>	15q21-q22	<i>B2m</i>	02-69.0
<i>ZASHA</i>	4	U14587	CGGCAAAACACGGGTCCGA	CCAAAACAACATGTTTTATGAAA	b	<i>ASCL1</i>	12q22-q23	<i>Mesh1</i>	
AA495383	5	AA495383	ATTACGGACACAGAGCGTCAT	CAGGAATGTTCAAGCCGATA	a				
<i>WNT11</i>	5	AF067429	GACATCAGATTAGTTCAGTT	ATGACTACACAATGATCAGT	a	<i>WNT11</i>	11q13.5	<i>Wnt11</i>	07mmu48.0
<i>DELTAB</i>	5	AF006488	CTGGAGGAGGGCTTAGATGAG	GTGTGGCTGTGGTCTCTTTT	b				
<i>APOA</i>	5	Y18653	CCGAGGAGTACAAGGAGCAG	CATCACACAGCTTTGGATCT	a	<i>APOA1/4</i>	11q23	<i>Apoa1/4</i>	09-27.0
<i>HER1</i>	5	X97329	CCCTGGTGAGCAAAAAGTGAAG	GGGGAAAACCTAATTTGTTTAAAG	b				
<i>EHH</i>	6	Y08426	GGTGGCCCTAATAACTGGTCCCT	CGCCGGCCGAGAAAAGCCAGA	b				
<i>NIC1</i>	6	U70493	TCAGACCACATGCTCAGTCAG	GGTAACCATGAAAACCGCAGA	b	<i>CHRNA1</i>	02q24-q32	<i>Acta</i>	02-44.0
<i>NETRIN1A</i>	6	AF002717	ACCATTAGAGCTGGACAGAA	ATGTATCTGAAACGTTGACGCC	a			<i>Ntn1</i>	
<i>TYMK</i>	6	AA494759	AGGAAGCCAAAATGAAATAGCC	AGGCGTGTATTAAGATGTCGAAA	a	<i>TYMK</i>		<i>Tmk</i>	
<i>JAK1</i>	6	U82980	CAGCAGATGCTTGACAACTGTGA	CGTAAGAAGGAGAGTACATTT	a	<i>JAK1</i>	01p31.3	<i>Jak1</i>	04-46.3
AA494737	6	AA494737	TGAAACATCCAACATTTTGGCAC	CTACACCGTCTTATAAGCACATT	a				
<i>HER6</i>	6	X97333	GGCCGCTTGGTACAGTCC	GTCAITTTATTGTCTTCCAAAAGA	b	<i>HRY</i>	03q28-q29	<i>Hes1</i>	16-27.0
<i>CONTACT</i>	6	Y12005	AGAAGCACATGGACTCTCGAA	GCTTCTCCTTAGTTCGCCGT	a	<i>GDF5</i>	20q11.2	<i>Gdf5</i>	02-90.0
<i>p110a</i>	6	Y12819	AGGAGCTGTGCAGAGCATTAG	GCCAAGGTTGACTACATCCA	a				
<i>CYCE</i>	7	X83594	CTTCTCATGTATGGAATAGG	CCCACCCTTTTTAGCAGCAATC	b	<i>CCNE</i>	19q13.1	<i>Cone</i>	07-16.0
<i>VNC</i>	7	U41419	CGTTCGCTCAAAATAGTCTTTT	CCACAGAGACAGTCCACCA	b	<i>CDH11</i>	16	<i>Cdh11</i>	08-51.0
<i>CK2A2</i>	7	X99964	AATGAACACAAGGCGCTGTAG	TAATGCAACAGGATGAGGAG	a	<i>CSNK2A2</i>	16q13	<i>Csnk2a2</i>	08-49.5
<i>CAS</i>	7	AA495336	AAATCCCCTAACGTAACAGAG	TAATGTCTCTGTGAAGGGTGC	a	<i>CAS</i>	20q13	<i>Caps1</i>	02-97.0
AA494758	8	AA494758	CCTTGAGACAGATTGTGATTTGG	AAATCTGAAAACAGGAGGAGAA	a				
AA494588	8	AA494588	GGATGTTTTAACAGATAGCCAGC	AAATGTAGCAATCACTCACTTT	a				
<i>NKX2.7</i>	8	U86573	AGCAGCGGAAAGAGCAAGATA	GCTGCTCCAGTTTCCAGGATG	a				
<i>ZFO2</i>	8	L11014	CCCTTCGAGAGGAGAGAGGGC	TATTGTTACGTTCCCTTGAATACA	a	<i>RHO</i>	03q21-q24	<i>Rho</i>	06-51.5
<i>DLX1</i>	9	U67842	CATGTGAACCAGGACTGCTGAA	GGGCAGAGTAGCAGCTTTGAT	a	<i>DLX1</i>	02q32	<i>Dlx1</i>	02-44.0
<i>ENG1</i>	9	X68445	ACCACCACATCCAAAGAAGA	CCAGCAAGGGAAAATACTATGG	b	<i>EN1</i>	02q13-q21	<i>En1</i>	01-64.1
<i>DESMIN</i>	9	U47113	GCAGTGACCCATTACCGCAT	CAGATATGACCCACACAC	b	<i>DES</i>	02q35	<i>Des</i>	01-41.1
<i>ACTR2A</i>	9	AA497191	TCCGAGAAGAGAAAACGACAA	GCTGACCCGAAACAGAAAGATG	b	<i>ACVFR2B</i>	03p22-p21.3		
AA497210	9	AA497210	GCACAAAATGATTTAGCTCGG	GGGTTAGAGAATCCCAATGA	a				
<i>DERM1</i>	9		TCCCATGATGATAGTACAGG	CATTGGATGAAAATCGAAAAA	b				
<i>CX43.4</i>	9	L46801	GAGACTCATAGACTATCATTGCT	GGCCGATGTCAGTATAACTGG	a	<i>CXA6</i>		<i>Gja6</i>	X-68.0
<i>SMAD4</i>	10		TCGAGAACATCTCGCCTACAT	AAAACGAGTTAAAGCGACCGT	a	<i>MADH4</i>	18q21.1	<i>Madh4</i>	18mmu48.0
<i>SMAD2</i>	10		CCCTTTCCTCTCCTGTTTCT	TTCTTAGGGTGTCCGATGAGC	a	<i>MADH2</i>	18q21	<i>Madh2</i>	18mmu48.0
AA494784	10	AA494784	TAGACGAGCAGACAGTGAAGAA	CTCAGGACTGTTCTGCATGAAT	a				
<i>HSC70</i>	10	L77146	CCATCGAGGAGGTCGACTAAG	GGGAAATAGTCCACATTTTATTTC	a	<i>HSPA8</i>	11q23.3-q25	<i>Hsc70</i>	
<i>WNT11r</i>	10		CCGAGGCTACAAACCTTACA	CAGAGAGGTTTTGCCGTTTCT	b	<i>WNT11</i>	11q13.5	<i>Wnt11</i>	07mmu48.0
AA805725	10	AA805725	TGCTTCAAGTTTCTGGCTTT	AGTGAACATCCCGCAATGTG	a				
<i>TCF3</i>	10	AF097410	CAGCCCTACCTCGTCTCAGGG	GCTGTCTCACTCCACAGATTG	a			<i>Tcf3</i>	
<i>MEF2C</i>	10	U86589	TGTCATCCCCATATATTGAA	CATCCCGCTCGTAATTGAGT	a	<i>MEF2C</i>	05q14	<i>Mef2c</i>	13-46.0
AA494754	11	AA494754	CTAAAAGCAGATTAGTTCGCC	CACTTGGGTAACATGGCTGTAT	a				
<i>DANF</i>	11	U65433	TTGGTTCGAGACAGAAAGAGC	TGGCCAAACACACACAATAAA	a				
<i>HER2</i>	11	X97330	GGATTGGAGACCCCTGGTG	CCTCTGCAGCTACACACTCTCA	a			<i>Hes5</i>	04-82.0
<i>WNT4</i>	11	U51267	AGTGTCCGAAAGATGGTGGAGA	CAGGAACAGCCACGAGTTGGA	b			<i>Wnt4</i>	04
<i>RCP</i>	11		CATCATGCAGCTCTTTGGAA	CATGGCAGGCACTTACC	b				
<i>PLASTICIN</i>	11	U89709	GCAGATGAAGAAGCGCTTCAGT	GTGCTGTATCATGACAAAGATA	a	<i>PRPH</i>	12q12-q13	<i>Prph1</i>	15-55.5
<i>DLX7</i>	12	U67845	GAGCTCGAGTCACTGACAAACA	CGTATCAGACTGAAACCTCATAG	a			<i>Dlx7</i>	
<i>RARA2A</i>	12	L03398	CTCCAACTCCGCCAGAGC	GGAAATTAAGGAGAAATACACAGT	b	<i>RARA</i>	17q12	<i>Rara</i>	11-57.0
AA494584	12	AA494584	CCACTGTGACTGATCAGCATCC	TCTCATGTCCGTCAGCAGACAA	a				
<i>TEF</i>	12	U96848	TGGAACAAGAGGAGGAGAGACA	CTCTGCAGCTGCAATTTCCG	a	<i>TEF</i>	22q13	<i>Tef</i>	15-48.7
AA494785	12	AA494785	AGTCGCTGTTCTCTCAAATCA	GCAGTTGCTGTAATGTGAGCTT	a				
AA494779	13	AA494779	GGTGCTACTTACATTCCGAGTA	GCTCTTATTGTACATGGAGGAT	a				
AA494841	13	AA494841	TCCATTGGCAGACAATCAGGTT	CTGGCATTTGGCTAGTCAAAG	a				
<i>MEKK4</i>	13	AA497200	CATTCTGAAAAGCTGAGCAGC	ACACTGGCTTCCATGCTTTTA	a	<i>MEKK4</i>	06	<i>MEKK4</i>	
<i>MELK</i>	13	AA497287	GACCATCATTTGGAGCAGTT	TCATGGACATGTTAGAGACGA	a	<i>KIAA0175</i>		<i>Melk</i>	04-27.0

**Table 1.** (Continued)

Locus	LG	Accn. No	Forward Primer	Reverse Primer	Map Position Reference	Human Ortholog		Mouse Ortholog	
						Name	Location	Name	Location
<i>NKX2.3</i>	13	U66571	GGCATGGTAGTGCCAGTGCAA	ACCGTCAAATGTTCCGGTGAATAT	a	<i>BAPX1</i>	04p16.1	<i>Bapx1</i>	05-23.0
AA494764	14	AA494764	CTTAGTTTATGCGAGTGTCCATCC	TCATGAATCCAGAGTTCACAGTTC	a				
<i>NKX2.5</i>	14	U66572	CATGTGCTTAAATCTCTTGCGAA	CAAACCTTACACTGGCAAAGCCTA	b	<i>CSX</i>	05q34	<i>Nkx2-5</i>	17-13.0
<i>RAN</i>	14	U61395	TTGGGCACATTTTTGTAGGTC	CTGGTTCCTGGTAGGTGATGGA	a	<i>RAN</i>		<i>Rasl2-9</i>	
<i>NGR1</i>	14	AF036149	CAATAGCATAAAGAAGCAACAGG	CAATTTGTAGAGCGGATAAAGCTG	a	<i>NEUROD3</i>	05q23-q31	<i>Neurod3</i>	13-35.0
<i>SMAD5</i>	14		TGCCTTGCAGTAACAGGATT	AGGCAGGACACAAAATGGACT	a	<i>MADH5</i>	5q31	<i>Smad5</i>	13mmu35.0
<i>TEC</i>	14	AF016326	TGCCTCAAGCTACAGCTCAAAA	TTGAGATTGCAGAAAACATGC	a	<i>EMT</i>	05q32	<i>Itk</i>	11-22.0
AI292396	14	AI292396	CATTCAAGTGGTCTTTATTGATG	AGGGACCAACTAGTATGTCCAG	a				
<i>WNT8</i>	14	U10889	CTGTAAATTTCACTGGTCTGC	CGCTAACCTTTACATTTTGCTTG	b			<i>Wnt8a</i>	
<i>FLK1</i>	14	U75995	CCAGAAGTCTGATCGGTCCAT	CGCTAACCCCGAACTAATGAT	b	<i>FLK1</i>	04q12	<i>Kdr</i>	05-42.0
AA494592	14	AA494592	TTTTCTCTTACAGCATGGTG	TTTTTGGAGTGAGGAAATGGT	a				
<i>CAD1</i>	14	X66958	AAGGATTACAACCTATTAATGAATA	TTACACATTGCTGAAATCAAACGT	b	<i>CDX4</i>	Xq13.2	<i>Cdx4</i>	X-42.5
AA494735	14	AA494735	CCTGAGTTTCTGCTTTTGACC	CATGCACGGCTTGGAGGACTTT	a				
AA658753	15	AA658753	CTTCAGAAAGGCCAAAGGAACT	CGATATTGACGAGGACTCCAA	a				
AA605851	15	AA605851	AACCTTCATGTGTGAGCCAAAG	GAGAACCTGTTTCAGATGCGAG	a				
<i>HSP47</i>	15	U31079	GATGAATTTGTAACCTTTAGGCTAT	GCATAGTGTGTTGTAATAATTC	c	<i>CBP1</i>		<i>Cbp1</i>	
<i>BOZDHARMA</i>	15	AB010103	GGGTGTGGTTCAGAAATCGAC	GTACATAAGTTAACTGTAGAACAA	d				
<i>MEL1B</i>	15	U52220	GTGCTCCTGACGCTAGTATGG	AAGAAATGGCCCAACTCACTA	a	<i>MTNR1B</i>	11q21-q22	<i>Mel-1b</i>	
AA605938	15	AA605938	CATGCATCCAGTGAGGAGTTT	ACATAACCTGTGTGCCATC	a				
<i>CHORDIN</i>	15	AF034806	CGGAATGCATAGAGGACTTCA	CAAAACAACAGTCAAAGGCCACA	e				
<i>GAP43</i>	15	L27845	GAATCCAAACCGCAGAGGG	GGCCGGCGCTTTTCTTTTCTTTG	a	<i>GAP43</i>	03	<i>Gap43</i>	16-29.5
<i>CTNNB</i>	16	U41081	CCAAGCTACCGCTCCTTCCA	GGCCGGCTCAGCAACTCTACA	b	<i>CTNNB1</i>	03p22-p21.3	<i>Catnb</i>	09-72.0
AA494840	16	AA494840	CGCCACCATCATTTGAGGCTT	GGCTATAAAGCAACTGCATACA	a				
<i>CK2A1</i>	16	S78875	TGATGGCAGGCATCGCTCTATGA	CAGGTGGACATTTACGTGTTTCTAA	a	<i>CSNK2A1</i>	20p13	<i>Csnk2a1</i>	
<i>MEF2D</i>	16	U66570	GGATATCGAAGACAGACCTGT	GCTCGTCTGTGATTCGCTGAA	a	<i>MEF2D</i>	01q12-q23	<i>Mef2d</i>	03-48.0
AA494836	16	AA494836	CATTTTCAGGGAAACATTTATTGTA	AGTACCTTAAAGGCCACTGTGAT	a				
<i>ZP47POU</i>	16	Y07905	GCACCTGTGAAAAGACGCGAGTTT	GATGCTATGCTACACTGATGTTG	a	<i>POU3F2</i>	06q16	<i>Pou3f2</i>	04-06.3
AA494791	16	AA494791	GTCCCTGCTCTACATGAGGAA	CAGTCTATGTTTCTTCCCGA	a				
AA605743	16	AA605743	TGATTTGGAGATTCTGTGGG	AACAAAGCCAAATGCTTACCTG	a				
<i>SVP46</i>	16	X70300	CCCACCCAGTAATCCATATC	CAGCTTTAATGGCTTCTGGTTG	a				
<i>BMP2a</i>	17	U82233	AAGGGAAACGCTGCTCACTCTC	GACCCGCTCGTACTCGTCCA	a	<i>BMP2</i>	20p12	<i>Bmp2</i>	02-76.1
AA494615	17	AA494615	GCAGACCCCTTAAACACAGTGG	CATTGTGAAGGTTTTGCCTGT	a				
<i>OTX1</i>	17	U14591	CTTTCCAATAAGCCCTCCATTTTC	CCATATTTAGAAAACGAAAACAGC	b	<i>OTX1</i>	02p13	<i>Otx1</i>	11-13.0
<i>PAX9</i>	17	U40931	GCAAATCTCATGTAACAACTGA	GCCTTTTCTCCCACTGCAATT	b	<i>PAX9</i>	14q12-q13	<i>Pax9</i>	12-26.0
<i>NKX2.2</i>	17	X85977	TGACTTGCAATGCAGGGTTGCG	TTGCCAGTGTTTTATCATGGCC	b			<i>HK22</i>	02
AA542570	18	AA542570	ATGAATGGATGAGAGCCACAC	AGCCACCTTACCTCCTGTTT	a				
<i>ICLN</i>	18	Y08484	GTCTCTGTGACGCTGACTGAA	CGCTGAGGAGAACAATTCCTGA	a	<i>CLN51A</i>	11q13.5-q14	<i>Clici</i>	07-50.0
<i>SAPK3</i>	18	Y15075	CTGAATCTAAAGAGACGGCCC	TCTCAGTGTCTACCTGTCTGCAT	a	<i>SAPK3</i>	22q13.3	<i>ERK6</i>	09-31.0
<i>AR1</i>	18	AF004521	CTCGTTTCTCCGTGTGCTCTA	CAATCACAATATCATACATAAGCA	a	<i>CYP19</i>	15q21	<i>Cyp19</i>	09-31.0
<i>LMO2</i>	18		CCACAAACAGACGGAGCCTA	CGCACAAACCGTTCAGAGAT	b				
AA494574	18	AA494574	AGGACTTTACACGTACAATGGC	GACTGAAACGATCCCGATCATG	a				
<i>FKD2</i>	18	AA494814	CTGCTGAAGGATATTTTTAAATGA	GGCTGTTTTGTAAGATCAGCAGA	b				
<i>FLI</i>	18		AAATCTTACCCAAAGACAGACATGACT	CCCCGCCACCAATTTTATGCA	b				
<i>RXR</i>	19	U29942	GAAGCGCTCATCAGCTGACC	GGGAATTCGAAAATAAGATCTC	b	<i>RXR</i>	06p21.3	<i>Rxrb</i>	17-18.49
<i>NTL</i>	19	S67147	CCTCTCAATGTACGATCCA	TCAAAGAGCCCAACAATAACA	b	<i>T</i>	06q27	<i>T</i>	17-04.02
<i>OPRD1</i>	19	AJ001598	AGGCTCAAAGGAGAGGACAG	CCTTGTCTGAATGGTAGGCA	b	<i>OPRD1</i>	01p36.1-p34.3	<i>Oprd1</i>	04-84.8
<i>DYNAMO</i>	19	X99769	CTGCTCTCAGTCTCTCTCCA	TCTCCGGATACAGATGATGCT	b			<i>Gdf6</i>	04
<i>CK2B</i>	19	S76877	CCACATAGAAAATAAAGATCTACT	CCATTTTATATTTTCTACTCTGAC	c	<i>CK2B</i>	06p21.3	<i>Csnk2b</i>	17-19.02
<i>DLX6</i>	19	U67844	CCTTGTATGTTGACATGCTGACT	ACGAAGGAGGTTTGAGTTGCTT	a	<i>DLX6</i>	07q22	<i>dlx6</i>	06-02.0
<i>PDGFRA</i>	20	Z32814	GATGCAAAGAGAGAGAAACAGACA	AATAACAGGCTGGTCTAAAATAG	a	<i>PDGFRA</i>	04q12	<i>Pdgfra</i>	05-42.0
<i>PCMT</i>	20	U37434	TTGAAAGGATGAGTCCGGCAAT	GGCCAGTCCGGTCTTCTGTT	a	<i>PCMT1</i>	06q22.3-q24	<i>Pcmt1</i>	10-07.0
AA658637	21	AA658637	GACAGTCTTTTACCCAAA	ACAGCAGTGGGAAGAGACCTT	a				
<i>SQUINT</i>	21	AF056327	GTTTATCCAAAGCGTTACAACGC	TGCTCATGACCACTTGCCTATT	f				
<i>RRM1</i>	21	U57964	GACCAGAGCCAGTCACTCAAC	ACGGTTCTCCAGGTGAACACAC	a	<i>RRM1</i>	11p15.5	<i>Rrm1</i>	07-69.0
<i>FGFR4</i>	21	U23839	GCACATCTTGTGAATGTGAGAAAC	GAGAGAGATTTATATTTATCAATGC	b	<i>FGFR4</i>	05q35.1-qter	<i>Fgfr4</i>	13-33.0
<i>MSXD</i>	21	X65062	TTAGGGTACAGCATGTATAC	GGTGCACATTAACCTCATCAATA	b				
AA494752	21	AA494752	CGCTGTACTCTGCCCCGG	AGGGACAGAACTGGAGTATGAT	a				
AA494804	22	AA494804	AGCCATACAATGGAAATAAGACAT	CCCATTTCCTGGGAGGAGCTT	a				
<i>VASA</i>	22	AB005147	GGTAGAGTTCGAGGTGGTTTC	TCCAACCTGTCTCAGTTGTTG	a			<i>Ddx4</i>	13-67.0
AA494763	22	AA494763	GTTTGGACCAAGTTACATAGCAC	AAATCTGCATCTGATCTGCTGG	a				
AA494769	22	AA494769	CATGAAAAAATCTGAACTGGAGG	GTTTGGCTTTAGGGTGGCAGATT	a				
<i>NADL1.1</i>	23	X89204	GGGATCCACGGCGGGCAGCA	CATGTACATTTCCGCATCAGT	b	<i>L1CAM</i>	Xq28	<i>L1cam</i>	X-29.51
<i>CHAF</i>	23	X81128	GCTTTGTGCTGCTACTTTGTCA	GACATCTCAAAATACCTTCAATC	a	<i>RAF1</i>	03p25	<i>Raf1</i>	08-52.5
AA494822	23	AA494822	TGGTCTCACATCAACAATTTCT	TGACGAATTCCTGTGAAATGAA	a				
<i>PRIM1</i>	23	AA494761	GTTTCTGTTTAAATAAGCTGGTG	AGAGTGTATCTCAGAAGGACTTT	a	<i>PRIM1</i>	12q13	<i>Prim1</i>	10-77.0
<i>HER4</i>	23	X97332	CGCAGTCACTCTGGAGACC	GATCTGTTAAGAACCTTCTATTCA	b			<i>HES5</i>	04-82.0
AA494818	23	AA494818	GTATTCGACTCTTCCGAGCC	TGAAAGTTTTAAACGTTTTGCTTT	a				
AA494819	23	AA494819	GTGGAAATGTTATCTTTACTGTCT	GCCAATGGAGATGTTTAAAGTGT	a				
<i>TARAMA</i>	23	X94119	GGATGTTGAGGATCCATCCTGTGAT	CTGCAAGTACAATGCTCCGGCCA	b	<i>ACVR1B</i>	12q13	<i>Acvrlk4</i>	
<i>POUC</i>	23	X88432	CAACCGTGGACAGACTCTGA	ATGTCCACAGATTCCTCTGG	b	<i>MPOU</i>		<i>Emb</i>	15-57.0
AA494567	23	AA494567	AAGCAGTTTGTACAAGCTGGAG	CGGATTAACACACTACCACGAG	a				
<i>CMYC</i>	24	L11710	GGTGGGAGTTCACATTAATGA	ATGTCCACAGTACCTAATACAGGT	a	<i>MYC</i>	08q24.12-q24.13	<i>Myc</i>	15-32.0
AA658764	24	AA658764	AGCGATATGACATTTGTCAGG	AACGTACAGCTTCCATTGTG	a				
AA494590	25	AA494590	TCAATAAAAGTCTCTGTACAGGG	ATTTTATGCGAAACCTTAAATGTAC	a				
<i>RAG1</i>	25	U13983	GACTGAATTTCAAAAATCTTTCAG	AAGATACCTTGGAAAGTGTAGAGC	a	<i>RAG1</i>	11p13	<i>Rag1</i>	02-56.0
<i>PAX6</i>	25	X61389	CTGAAAACACCTGCGAGCTGAC	GTTGCTGTGTTGTTGATCTCTC	a	<i>PAX6</i>	11p13	<i>Pax6</i>	02-58.0
<i>WT1</i>	25	X85734	TAAACAGCTAATCTGAACATAACTG	TGCTGCACAGCAAAACAGACAT	b	<i>WT1</i>	11p13	<i>Wt1</i>	02-58.0
AI292377	25	AI292377	AATGTGGGTGAAAATCTTGAAT	AGGATGGCTTTTCAATCAGTT	a				
<i>MYOD</i>	25	Z38945	TGCAACATTTCAAAAACAAATGAAA	GGTCAATTTTCCAGTGGATATAACA	a	<i>MYOD1</i>	11p15.4	<i>Myod1</i>	07-23.5
<i>ISL2</i>	25	X88805	GAGACGTGAGAGCCACC	GCTACACAAGGATGGTCTTGAAT	b				
<i>WHN</i>	Unk	Y11537	GCTTTGAAGAACAGCAAAACG	AAGACATCCCTTCTGGAGGA	a	<i>WHN</i>		<i>Hfh11</i>	11-45.0

Map position references: (a) This paper; (b) Postlethwait et al. (1998); (c) Förnzler et al. (1998); (d) Fekany et al. (1999); (e) Fisher et al. (1997); (f) Feldman et al. (1998). Primers for loci with no accession number were described by Postlethwait et al. (1998), except for *smad2*, *smad4*, and *smad5*, which were derived from our unpublished data (H.I. Sirotkin; A. F. Schier; and W.S. Talbot). The *WHN* gene was not significantly linked to other markers in the data set, and its position is currently unknown (UNK). Human and mouse map positions were obtained from the OMIM (<http://www3.ncbi.nlm.nih.gov/OMIM/>) and MGD (<http://www.informatics.jax.org/maptools.html>) databases, respectively.

**Table 2.** Mapped STSs

Locus	LG	Forward Primer	Reverse Primer	Method	Map Position Reference
12R.360	2	ACAGGTGCGTCCAATAGCTCC	ACAGGTGCGTGATCAAATGTT	Size poly.	b
BAC.69B22T7	4	GGAGTTTTAAACAAGACATAACCA	GTTGCGGGTTCATTCCGCTTT	SSCP	a
BAC.79D15T7	4	TGGTAGATCACACTGTTGGG	CGAGAAGAGACCCGTAGAAC	SSCP	a
YAC.70B7T7	7	ATGGGGCTGGTGAAGTTGTAG	TGGATTTGAAAAAGGGCTTG	SSCP	a
STS.082	7	TAAGCGTTTCAAGGGAGTGT	ATGAGCTTTCACACGTACGCA	SSCP	a
20K.850	10	GGAGGTGAGCGCGAGTATT	ATACACAAGAACACGGCTCAG	Tsp509 I poly.	c,d
20K.875	10	CTAAAAGGTTTGGGAACCCCTG	GGAGGCAAGCCCGATTATTTA	Presence/absence	c,d
15T.760	12	CTGTCTGGAGAAAGTCTTATTTG	GGATGCCACTGGTACTAATTGATA	Nla III poly.	b
18AF.1190	12	CCCTCTGCACAGAACTGAAACCTC	CCGTTTCTGTGAAGACAGGAAG	Mnl I poly.	b
PAC.84H13SP6	12	CTGTGACGTTTGTATGTTGAAA	TGTTCCACCCTGCCTGCG	SSCP	a
STS.C10a	13	CGTACTCCACACCACACTCCAT	AGCGCCTTCTAATGGATTGTGTC	SSCP	a
YAC.6B10T7	13	GAGCGGGTGTTCACACAAAT	CACAAATTGGCAAGGTGCTA	SSCP	a
BAC.2L7SP6	19	GAGCAGAACCCTATGGACGT	AATGTGGCGTAAGGCGCGAC	SSCP	a
BAC.173J7T7	19	CAGGCCTGTTCTGGAATGT	GCATTAAGGAAAACGGCAA	SSCP	a
STS.C10b	24	CGTACTCCACACCACACTCCAT	AGCGCCTTCTAATGGATTGTGTC	SSCP	a

Map position references: (a) This paper; (b) Talbot et al. (1998); (c) Schier et al. (1997); (d) Zhang et al. (1998).

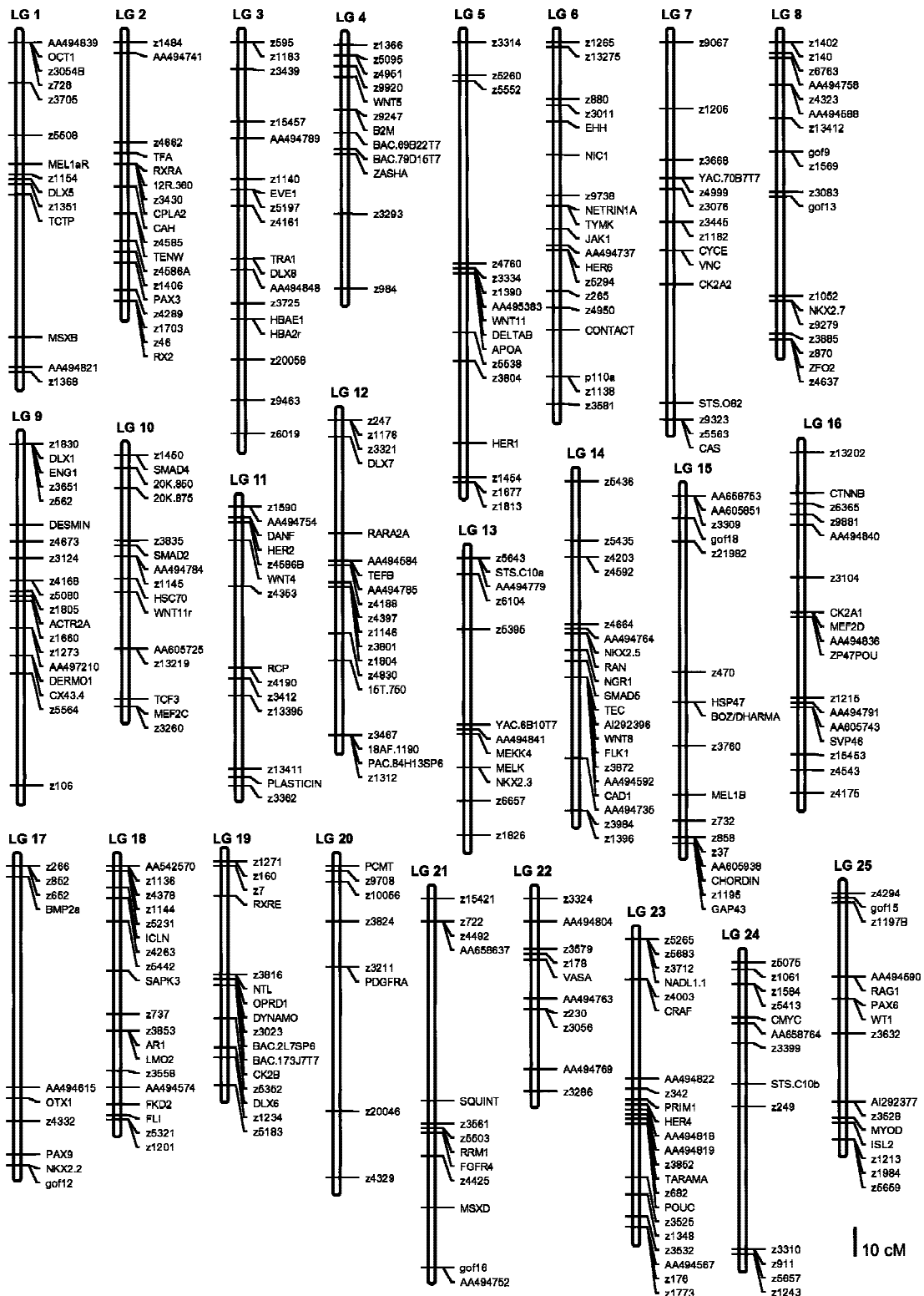
ferences in distances between our map and others, particularly the microsatellite map (Knapik et al. 1998). This can be seen in total length (2894 vs. 2350 cM) and most strikingly in particular regions. For example, two markers on LG 20, z3211 and z20046, are separated by 39 cM in our map but only 10 cM in the microsatellite map (Knapik et al. 1998). Differences of this magnitude reflect more than statistical variation, because the standard error for markers separated by 10 cM in our map is 4.2 cM and the 95% confidence interval is 2.9–21.3 cM. Sex-based differences in meiotic recombination could be a factor, as the microsatellite map was sex-averaged (Knapik et al. 1998) whereas our haploid map monitors only female meiosis. Whatever the cause of these differences, it is clear that one must be cautious in using marker distances in one cross to infer distances in others, although marker orders are readily comparable between maps.

### Comparative Analysis

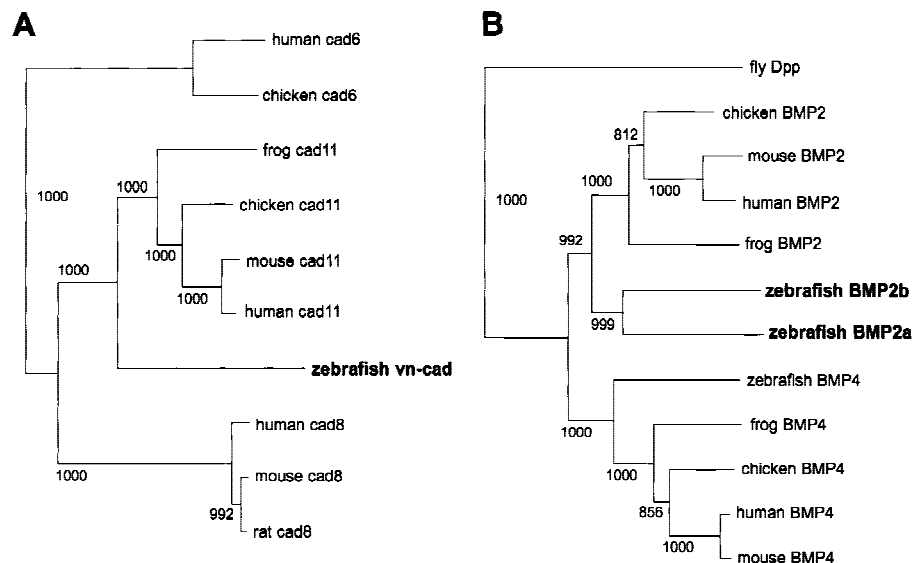
Previous work identified regions of synteny conserved between zebrafish and mammals (Postlethwait and Talbot 1997; Postlethwait et al. 1998). To investigate conservation of synteny in light of the additional mapped genes, we examined locations of zebrafish genes and their counterparts in mouse and human. Using the BLAST family of search programs (Altschul et al. 1997), we identified mammalian sequences significantly similar to zebrafish genes for which sequence and map location were available from this and previous studies (see Methods for description of search criteria). To determine which homologous sequences are likely to represent mammalian orthologs of zebrafish genes, we constructed phylogenetic trees using the CLUSTALX program (Thompson et al. 1997). Because they are descended from the same gene in the last com-

mon ancestor of two species, orthologs share a terminal branch of a phylogenetic tree. For example, in the tree shown in Figure 3A, human cadherin 11 is more closely related to zebrafish ventral neural cadherin (*vn-cad*; Franklin and Sargent 1996) than to any other known human sequence. Thus, human *cadherin 11* is the likely ortholog of zebrafish *vn-cad*. Phylogenetic analysis identified 134 human and 152 mouse genes that are probable orthologs of mapped zebrafish genes (Tables 1 and 3). There were 18 cases, such as that shown in Figure 3B, in which two zebrafish genes appeared to be orthologous to a single mammalian gene (Table 4; see also Stock et al. 1996; Pfeiffer et al. 1998; Sefton et al. 1998; references corresponding to GenBank accession nos. in Table 4). Conversely, we noted three cases in which two mammalian genes appeared to be orthologous to a single zebrafish gene. Some of these cases may be resolved with the discovery of additional genes in zebrafish or mammals, but many are likely to represent true cases of extra genes, as we consider further in the Discussion.

By comparing the map locations of orthologous genes, we identified synteny conserved between zebrafish and mammals (Table 5; Fig. 4). Each point in the Oxford grid shown in Figure 4A represents a pair of orthologous genes plotted by map location in zebrafish and human. Bins occupied by more than one point indicate cases in which orthologous genes are syntenic in both species. Of 124 genes mapped in both zebrafish and human, 79 genes are located in a bin containing two or more points, defining a total of 28 conserved syntenic groups. The observed clustering is greater than that predicted by a random Poisson distribution (Fig. 4A; see Methods for calculation). This confirms and extends the observation (Postlethwait et al. 1998) that syntenic genes in zebrafish tend to have orthologs



**Figure 2** Genetic linkage map of the zebrafish genome. Shown are the positions of 389 markers genotyped on a single mapping panel consisting of 48 haploid zebrafish embryos. The map includes 157 gene and EST markers, 15 STS markers, and 217 SSLP markers. The SSLPs, which are designated as "z" and "gof" markers, are described in Knapik et al. (1998) and Goff et al. (1992), respectively. ESTs with mammalian orthologs are named by their human counterparts. ESTs with no clear orthology are named by their GenBank accession nos. Primer sequences and GenBank accession nos. for mapped genes, ESTs, and STSs are shown in Tables 1 and 2. Linkage group nomenclature and orientation follows Postlethwait et al. (1998).



**Figure 3** Phylogenetic analysis of zebrafish protein sequences for (A) vn-cadherin and (B) BMP2a and BMP2b. CLUSTALX was used to display the relationships of zebrafish gene products and their homologs in other organisms, as described in the text. Numbers at nodes represent bootstrap support for 1000 replications. NCBI protein sequence ID numbers for the proteins shown are human cad6 (1705545), chicken cad6 (2134302), frog cad11 (3377485), chicken cad11 (3511021), mouse cad11 (1705549), human cad11 (1377894), zebrafish vn-cad (1345125), human cad8 (1705547), mouse cad8 (3023433), rat cad8 (2804294), fly Dpp (118409), chicken BMP2 (2501173), mouse BMP2 (1345611), human BMP2 (115068), frog BMP2 (115070), zebrafish BMP2b (2804175), zebrafish BMP2a (2149148), zebrafish BMP4 (2804177), frog BMP4 (399122, chicken BMP4 (2501175), human BMP4 (115073), and mouse BMP4 (461633).

that are on the same chromosome in human. Figure 4B shows a similar analysis of synteny conserved between zebrafish and mouse. This zebrafish–mouse comparison identified 28 conserved syntenic groups involving 73 of 135 genes examined. Figure 4C shows a human–mouse Oxford grid compiled from the same set of genes used for the comparisons with zebrafish in Figures 4A and 4B. Seventy-seven of the 103 genes mapped in both human and mouse fall into 21 conserved syntenic groups.

## DISCUSSION

We have localized 104 zebrafish genes and ESTs within a framework map derived from 275 previously mapped genes, STSs, and SSLPs. This work increases the total number of mapped zebrafish genes to >250, representing a majority of zebrafish genes for which full-length cDNA sequences are available. These genes can now be efficiently tested as candidates for mutations. Furthermore, the localization of these genes advances the comparison of the zebrafish genome with those of other vertebrates.

### Mapping the Correspondence Between Zebrafish and Human Genomes

By comparing the map locations of zebrafish genes and

their mammalian counterparts, Postlethwait et al. (1998) discovered that syntenic genes in zebrafish tend to be syntenic in mammals. We have further analyzed the conservation between zebrafish and mammalian genomes, considering the map locations reported here and in previous work. With the increase in gene number, we have identified eight new syntenic groups conserved between zebrafish and human and added from one to three genes to each of six previously recognized groups. The human–mouse comparison (Fig. 4C) compiled with the same genes used for the zebrafish analysis suggests that conservation between human and mouse extends over larger intervals than between human and zebrafish: The average human–mouse segment contains more genes (3.7) than the average human–zebrafish segment (2.8). Additional

mapping experiments are needed to better evaluate parameters important for comparative mapping, such as the size of the average segment and the extent to which gene order is maintained within conserved segments.

Despite the current uncertainties surrounding these parameters, the recent molecular analysis of the *you-too* (*yot*) locus illustrates the utility of comparative mapping (Karlstrom et al. 1999). Genetic mapping localized the *yot* locus to LG 9, which is homologous to human chromosome 2 (Fig. 4A; Postlethwait and Talbot 1997; Postlethwait et al. 1998). This, together with phenotypic analysis of *yot* mutants, suggested the human chromosome 2 gene *gli2* as a candidate for *yot*. As predicted by the zebrafish–human comparison, the zebrafish ortholog of *gli2* mapped to LG 9 and was found to be disrupted in *yot* mutants (Karlstrom et al. 1999). This example suggests that comparative mapping will be useful in identifying candidate genes for other zebrafish mutations.

### Genome Duplications and Vertebrate Evolution

The important role of genome duplication in the evolution of the vertebrates was recognized by Ohno (1970), who postulated that divergence of duplicated genes resulting from tetraploidization events drives the

**Table 3.** Orthologs of Other Mapped Zebrafish Genes

Locus	Accn. No.	LG	Human Ortholog		Mouse Ortholog	
			Name	Location	Name	Location
<i>DELTA</i>	AF030031	1	<i>Dll1</i>		<i>Dll1</i>	17-08.2
<i>OTX3</i>	D26174	1	<i>OTX1</i>	02p13	<i>Otx1</i>	11-13.0
<i>ENG3</i>	X68447	2	<i>EN2</i>	07q36	<i>En2</i>	05-15.0
<i>PTC1</i>	X98883	2			<i>Ptch2</i>	04
<i>TWHH</i>	U30710	2	<i>SHH</i>	07q36	<i>Shh</i>	05-16.0
<i>CDC27</i>	H56843	3	<i>CDC27</i>	17q12-q23.2	<i>Cdc27</i>	11-63.0
<i>ELRC</i>	U17601	3	<i>HUC</i>		<i>HuC</i>	09-05.0
<i>HBAA1</i>	U50382	3	<i>HBA</i>	16pter-p13.3	<i>Hba-a2</i>	11-16.0
<i>HOXB5</i>	X68248	3	<i>HOXB5</i>	17q21-q22	<i>HoxB5</i>	11-56.0
<i>RARA2B</i>	L03399	3	<i>RARA</i>	17q12	<i>Rara</i>	11-57.0
<i>ISL1</i>	D21135	5	<i>ISL1</i>	05q		
<i>LIM3</i>	U34590	5			<i>Lhx3</i>	02-16.0
<i>LIM6</i>	AF001299	5	<i>LHX1</i>	11p13-p12	<i>Lhx1</i>	11-48.0
<i>SOX19</i>	X79821	5	<i>SRY</i>	Yp11.3	<i>Tdy</i>	Y-03.0
<i>SVP44</i>	X70299	5	<i>TFCOUP1</i>	05q14	<i>Erbaf3</i>	13-45.0
<i>TNC</i>	X89203	5	<i>HXB</i>	09q33	<i>Tnc</i>	04-32.2
<i>BRN1.2</i>	U43657	6	<i>POU3F1</i>	01p34.1	<i>Pou3f1/Scip</i>	04-56.1
<i>TIE1</i>	AF053633	6	<i>TIE</i>	01p34-p33	<i>Tie1</i>	04-50.0
<i>CYCD1</i>	X87581	7	<i>CCND1</i>	11q13	<i>Ccnd1</i>	07-72.3
<i>ENG2</i>	X68446	7	<i>EN2</i>	07q36	<i>En2</i>	05-15.0
<i>FGF3</i>	Z48714	7	<i>FGF3</i>	11q13	<i>Fgf3</i>	07-72.4
<i>SHH</i>	Z35669	7	<i>SHH</i>	07q36	<i>Shh</i>	05-16.0
<i>TBX6</i>	U80951	7			<i>Tbx6</i>	07-61.0
<i>COL2A1</i>	U23822	8	<i>COL2A1</i>	12q13.11-q13.2	<i>Col2a1</i>	15-54.5
<i>MHC2DAA</i>	U08871	8	<i>MHCII</i>	06p21.3	<i>H2</i>	17-18.56
<i>MHC2DAB</i>	U08870	8	<i>MHCII</i>	06p21.3	<i>H2</i>	17-18.56
<i>MHC2DDB</i>	U08872	8	<i>MHCII</i>	06p21.3	<i>H2</i>	17-18.56
<i>ACTBB</i>	X76051	9	<i>INHBB</i>	02cen-q13	<i>Inhbb</i>	01-64.1
<i>BRN1.1</i>	U43656	9	<i>POU3F3</i>	03p14.2	<i>Pou3f3</i>	01-22.0
<i>DLX2</i>	U03875	9	<i>DLX2</i>	02q32	<i>Dlx2</i>	02-44.0
<i>EVX2</i>	X99290	9	<i>EVX2</i>	02q31-q32	<i>Evx2</i>	02-45.0
<i>HOXD4</i>	Y14548	9	<i>HOXD4</i>	02q31-q32	<i>Hoxd4</i>	02-45.0
<i>WNT10A</i>	U02544	9			<i>Wnt10a</i>	
<i>GATA1</i>	U18311	11	<i>GATA1</i>	Xp11.23	<i>Gata1</i>	X-02.1
<i>GATA2</i>	U18312	11	<i>GATA2</i>	03	<i>Gata2</i>	06-38.5
<i>SNA1</i>	S68799	11			<i>Sna</i>	02-97.0
<i>DLX3</i>	X65060	12	<i>DLX3</i>	17q21.3-q22	<i>Dlx3</i>	
<i>KRX20</i>	X70322	12	<i>EGR2</i>	10q21.1-q22.1	<i>Egr2</i>	10-35.0
<i>MSXC</i>	X65061	13			<i>Msx3</i>	07-68.0
<i>PAX2</i>	X63961	13	<i>PAX2</i>	10q24.3-q25.1	<i>Pax2</i>	19-43.0
<i>RET1</i>	X94363	13	<i>RET</i>	10q11.2	<i>Ret</i>	06-53.2
<i>EGR1</i>	U12895	14	<i>EGR1</i>	05q31.1	<i>Egr1</i>	18-16.0
<i>MSXE</i>	U50563	14	<i>MSX1</i>	04p16.1	<i>Msx1</i>	05-21.0
<i>LIM1</i>	L37802	15	<i>LHX1</i>	11p13-p12	<i>Lhx1</i>	11-48.0
<i>RADAR</i>	914116	16			<i>Gdf6</i>	04
<i>AXIAL</i>	Z22762	17	<i>HNF3B</i>	20p11	<i>Hnf3b</i>	02-84.0
<i>BMP4</i>	U90122	17	<i>BMP4</i>	14q22-q23	<i>Bmp4</i>	14-14.0
<i>GSC</i>	L03395	17	<i>GSC</i>	14q32.1-q32.1	<i>Gsc</i>	12-52.0
<i>OTX2</i>	D26173	17	<i>OTX2</i>	14q21-q22	<i>Otx2</i>	14-20.5
<i>SNAP25B</i>	AF091594	17	<i>SNAP25</i>	20p11.2	<i>Snap25</i>	02-78.2
<i>WNT2</i>	U51266	18	<i>WNT2</i>	07q31	<i>Wnt2</i>	06-06.2
<i>MHC1UAA</i>	Z46776	19	<i>MHCI</i>	06p21.3	<i>H2</i>	17-23.0
<i>BMP2B</i>	D49971	20	<i>BMP2A</i>	20p12	<i>Bmp2</i>	02-76.1
<i>SNAP25A</i>	AF091593	20	<i>SNAP25</i>	20p11.2	<i>Snap25</i>	02-78.2
<i>BF</i>	U34662	21	<i>BF/C2</i>	06p21.3	<i>Bf/C2</i>	17-18.85
<i>EPB72</i>	Y13473	21	<i>EPB72</i>	09q33-q34	<i>Epb7.2</i>	02-24.0
<i>POU2</i>	D28548	21	<i>POU5F1</i>	06p21.3	<i>Pou5f1</i>	17-19.23
<i>SCL</i>	AF038873	22	<i>TAL1</i>	01p32	<i>Tal1</i>	04-49.5
<i>HOXC5</i>	X68324	23	<i>HOXC3D</i>	12q13	<i>Hoxc5</i>	15-57.4
<i>RARG</i>	S74156	23	<i>RARG</i>	12q13	<i>Rarg</i>	15-57.4
<i>SNA2</i>	U24225	23			<i>Sna</i>	02-97.0
<i>VAL</i>	AF006641	23			<i>Krml</i>	02-91.0
<i>WNT1</i>	X58880	23	<i>WNT1</i>	12q12-q13	<i>Wnt1</i>	15-56.8
<i>WNT10BR</i>	263561	23	<i>WNT10B</i>	12q13	<i>Wnt10b</i>	15-56.8
<i>FKD3</i>	AF052246	25	<i>FKD5</i>		<i>Mf3</i>	09-41.0

Mammalian orthologs of previously mapped zebrafish genes were identified by the systematic phylogenetic analysis described in the text. The map positions of these genes were reported by Postlethwait et al. (1998). Human and mouse map positions were obtained from the OMIM (<http://www3.ncbi.nlm.nih.gov/Omim/>) and MGD (<http://www.informatics.jax.org/maptools.html>) databases, respectively.

emergence of new gene functions. At least two genome duplications occurred early in the evolution of vertebrates (for review, see Holland and Garcia-Fernández 1996; Sidow 1996). Analysis of chordate genomes reveals a fourfold increase in the number of genes and members of gene families in mammals relative to the cephalochordate amphioxus. For example, tetrapods have four paralogous Hox complexes whereas amphioxus has only one (excluding the *ParaHox* complex, which is thought to have arisen very early in the evolution of animals; Brooke et al. 1998). These paralogous genes derived from genome duplications (rather than multiple independent tandem duplications), because mapping studies show that paralogous genes often reside in duplicated chromosomal segments (Morizot 1990; Lundin 1993; Amores et al. 1998; Postlethwait et al. 1998).

Examination of teleost genomes has provided insight into the timing of these duplications. Mammalian genes and chromosomal segments have identifiable orthologs in zebrafish and other teleosts, indicating that these genes were formed by duplications predating the divergence of ray- and lobe-finned fishes, the lineages leading to extant teleosts and tetrapods, respectively (Morizot 1990; Holland and Garcia-Fernández 1996; Sidow 1996; Postlethwait et al. 1998). Unexpectedly, these studies have also revealed that teleost gene families often contain more members than the corresponding families in tetrapods (for review, see Wittbrodt et al. 1998). For example, zebrafish have extra genes in the *hox*, *dlx*, *msx*, *engrailed*, *hedgehog*, *pax*, *otx*, and *bmp* gene families (Ekker et al. 1992, 1995, 1997; Mori et al. 1994; Stock et al. 1996; Martinez-Barbera et al. 1997; Amores et al. 1998; Pfeffer et al. 1998; Force et al. 1999).

Mapping experiments provide insight into the origins of these extra genes in teleosts. Because extra zebrafish genes are dispersed throughout the genome, rather than clustered with related genes, it seems that zebrafish gene families were not expanded by widespread tandem duplication (Postlethwait et al. 1998; this paper). Instead, there are some cases in which syntenic zebrafish genes have extra paralogs that are also syntenic. For ex-

**Table 4.** Mammalian Genes with Two Apparent Zebrafish Orthologs

Mammalian gene	Zebrafish Co-orthologs	
<i>BMP2</i>	<i>BMP2A</i> (U82233)	<i>BMP2B</i> (AF072456)
<i>DLL1</i>	<i>DELTA</i> (AF030031)	<i>DELTA</i> (Y11760)
<i>DLX2</i>	<i>DLX5</i> (U67843)	<i>DLX2</i> (U03875)
<i>EN2</i>	<i>ENG3</i> (X68447)	<i>ENG2</i> (X68446)
<i>FKD5</i>	<i>FKD5</i> (AF052248)	<i>FKD3</i> (AF052246)
<i>GDF6</i>	<i>RADAR</i> (914116)	<i>DYNAMO</i> (X99769)
<i>Hes5</i> (mouse)	<i>HER2</i> (X97330)	<i>HER4</i> (X97332)
<i>Islet2</i> (rat)	<i>ISLET2</i> (X88805)	<i>ISLET3</i> (D38454)
<i>L1CAM</i>	<i>NADL1.1</i> (X89204)	<i>NADL1.2</i> (X89205)
<i>LHX1</i>	<i>LIM6</i> (AF001299)	<i>LIM1</i> (L37802)
<i>Msx3</i> (mouse)	<i>MSXB</i> (U16311)	<i>MSXC</i> (X65061)
<i>Ntn1</i> (mouse)	<i>NETRIN1A</i> (AF002717)	<i>NETRIN1</i> (AF017420)
<i>OTX1</i>	<i>OTX3</i> (D26174)	<i>OTX1</i> (U14591)
<i>PAX2</i>	<i>PAX2.1</i> (X63961)	<i>PAX2.2</i> (AF072547)
<i>RARA</i>	<i>RARA2B</i> (L03399)	<i>RARA2A</i> (L03398)
<i>SHH</i>	<i>TWHH</i> (U30710)	<i>SHH</i> (Z35669)
<i>SNA</i>	<i>SNA1</i> (S68799)	<i>SNA2</i> (U24225)
<i>SNAP25</i>	<i>SNAP25B</i> (AF091594)	<i>SNAP25A</i> (AF091593)

Phylogenetic analysis identified 18 mammalian genes that clustered with pairs of zebrafish genes (see example in Fig. 3B). In all of these cases, this clustering was supported by a bootstrap value of at least 800 of 1000. Other mammalian genes with two possible zebrafish orthologs whose bootstrap values were <800 of 1000 were identified but not listed here. The mammalian genes shown are from human unless otherwise noted. GenBank accession nos. are in parentheses.

ample, *bmp2a* and *snap25b* are located on LG 17, and the corresponding extra genes, *bmp2b* and *snap25a*, are both located on LG 20. This supports the view (Postlethwait et al. 1998; Wittbrodt et al. 1998) that these extra genes resulted from a genome duplication. Future mapping studies will address this issue further, as more zebrafish genes in families with expanded membership are localized.

If extra genes in zebrafish are the products of a genome duplication, when did this occur? One possibility is that three genome duplications occurred prior to the divergence of ray- and lobe-finned fishes. According to this model, zebrafish genes outnumber their mammalian counterparts because duplicate genes were lost at a higher frequency in the tetrapod lineage. Alternatively, an extra genome duplication occurred in the fish lineage after it diverged from the tetrapod lineage. Phylogenetic analysis can distinguish between these possibilities, as shown in Figure 5. The products of a duplication after divergence would be equally related to a single element (i.e., a gene or chromosomal segment) in the nonduplicated lineage (Fig. 5A). In contrast, if a shared genome duplication was followed by differential loss of paralogs, then one of the extra paralogs in the high-retention lineage would be more closely related to the surviving paralog in the other lineage (Fig. 5B). Our phylogenetic analysis of expanded zebrafish gene families identified 11 clear cases of mammalian genes equally related to two zebrafish genes (the relationship schematized in Fig. 5A) and only 4 cases in which one of the extra zebrafish genes

appeared to be more closely related to the mammalian gene. Thus our comparative analysis of the zebrafish gene map supports the view based on detailed analysis of zebrafish *hox* clusters (Amores et al. 1998) that the lineage leading to zebrafish underwent a genome duplication after the divergence of ray- and lobe-finned fishes.

It is likely that this extra genome duplication is not a recent evolutionary occurrence, and it may predate the radiation of teleosts. The genomes of other teleosts, including pufferfish (*Fugu rubripes*) and medaka (*Oryzias latipes*), also have extra genes (for review, see Wittbrodt et al. 1998), suggesting that the duplication preceded the last common ancestor of these species. In addition, the fact that zebrafish, medaka, and pufferfish are diploid species argues against a recent genome duplication, which would result in tetraploidy or pseudotetraploidy, as seen, for example, in the African clawed frog (*Xenopus laevis*) and rainbow trout (*Onchorynchus mykiss*), respectively (Bisbee et al. 1997; Wittbrodt et al. 1998; Young et al. 1998). Finally, duplicate genes can be expressed in dissimilar patterns, which suggests that the duplication producing these genes was followed by enough time for them to diverge and perhaps acquire different functions. For example, the zebrafish *pax2.1* and *pax2.2* genes are both apparently orthologous to mammalian *pax2*. The *pax2.1* gene is expressed in the presumptive cerebellum earlier than *pax2.2* and only *pax2.1* is detected in the developing pronephros (Pfeifer et al. 1998). Furthermore, mutational analysis

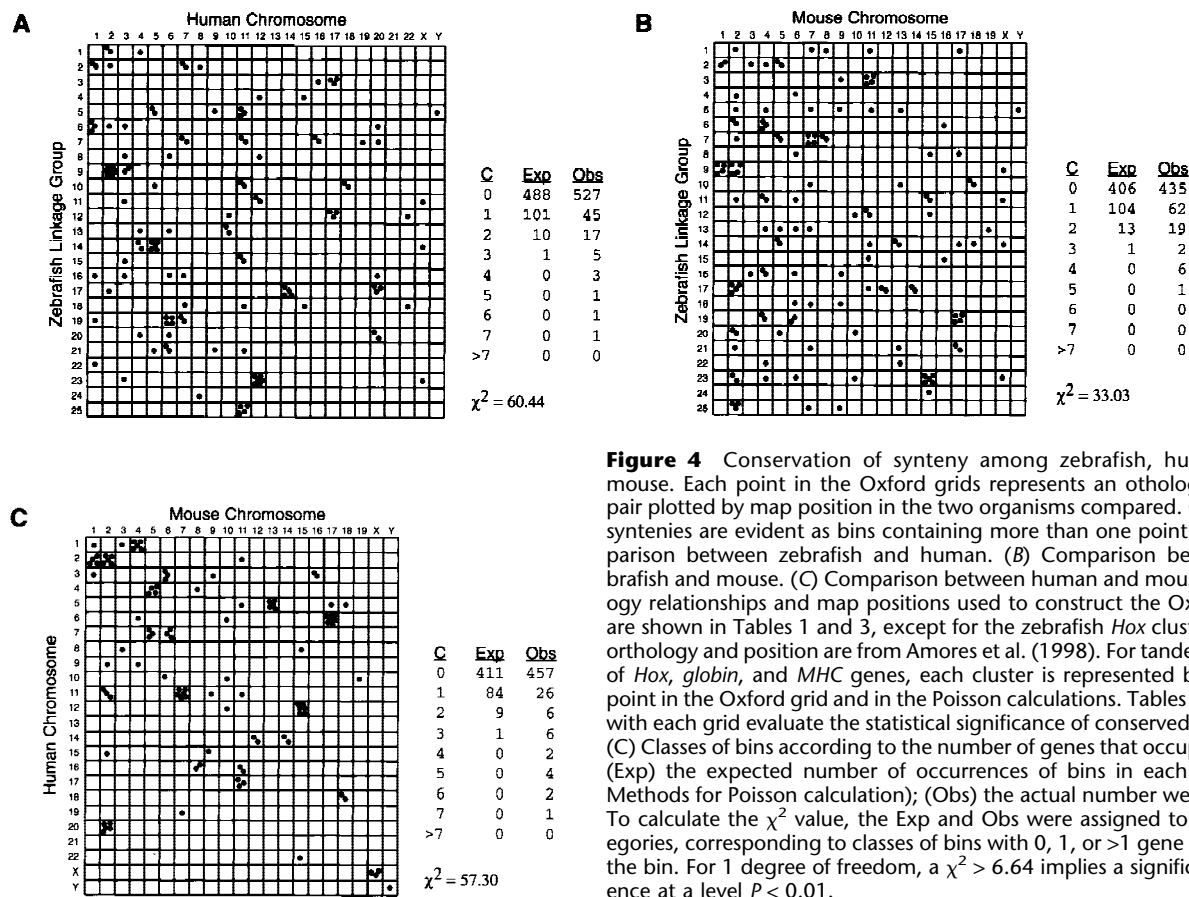
**Table 5.** Summary of Zebrafish–Human Conserved Syntenies

Zebrafish		Human		Zebrafish		Human	
LG	gene	gene	location	LG	gene	gene	location
1	<i>OTX3</i>	<i>OTX1</i>	02p13	12	<i>HOXBB</i>	<i>HOXB</i>	17q21–q22
	<i>DLX5</i>	<i>DLX2</i>	02q32		<i>RARA2A</i>	<i>RARA</i>	17q12
2	<i>RXRA</i>	<i>RXRG</i>	01q22–q23	13	<i>DLX3</i>	<i>DLX3</i>	17q21.3–q22
	<i>CPLA2</i>	<i>PLA2G4</i>	01q25		<i>RET1</i>	<i>RET</i>	10q11.2
	<i>TWHH</i>	<i>SHH</i>	07q36		<i>PAX2</i>	<i>PAX2</i>	10q24.3–q25.1
3	<i>ENG3</i>	<i>EN2</i>	07q36	14	<i>MSXE</i>	<i>MSX1</i>	04p16.1
	<i>RARA2B</i>	<i>RARA</i>	17q12		<i>FLK1</i>	<i>FLK1</i>	04q12
	<i>CDC27</i>	<i>CDC27</i>	17q12–q23.2		<i>NGR1</i>	<i>NEUROD3</i>	05q23–q31
5	<i>HOXBA</i>	<i>HOXB</i>	17q21–q22	15	<i>SMAD5</i>	<i>MADH5</i>	05q3
	<i>ISL1</i>	<i>ISL1</i>	05q		<i>EGR1</i>	<i>EGR1</i>	05q31.1
	<i>SVP44</i>	<i>TFCOUP1</i>	05q14		<i>TEC</i>	<i>EMT</i>	05q32
6	<i>LIM6</i>	<i>LHX1</i>	11p13–p12	17	<i>NKX2.5</i>	<i>CSX</i>	05q34
	<i>WNT11</i>	<i>WNT11</i>	11q13.5		<i>LIM1</i>	<i>LHX1</i>	11p13–p12
	<i>APOA</i>	<i>APOA1/4</i>	11q23		<i>MEL1B</i>	<i>MTNR1B</i>	11q21–q22
7	<i>JAK1</i>	<i>JAK1</i>	01p31.3	19	<i>PAX9</i>	<i>PAX9</i>	14q12–q13
	<i>TIE1</i>	<i>TIE</i>	01p34–p33		<i>OTX2</i>	<i>OTX2</i>	14q21–q22
	<i>BRN1.2</i>	<i>POU3F1</i>	01p34.1		<i>BMP4</i>	<i>BMP4</i>	14q22–q23
9	<i>ENG2</i>	<i>EN2</i>	07q36	20	<i>GSC</i>	<i>GSC</i>	14q32.1–q32.1
	<i>SHH</i>	<i>SHH</i>	07q36		<i>AXIAL</i>	<i>HNF3B</i>	20p11
	<i>CYCD1</i>	<i>CCND1</i>	11q13		<i>SNAP25B</i>	<i>SNAP25</i>	20p11.2
10	<i>FGF3</i>	<i>FGF3</i>	11q13	21	<i>BMP2a</i>	<i>BMP2</i>	20p12
	<i>VNC</i>	<i>CDH11</i>	16		<i>RXRE</i>	<i>RXRβ</i>	06p21.3
	<i>CK2A2</i>	<i>CSNK2A2</i>	16q13		<i>CK2B</i>	<i>CK2B</i>	06p21.3
11	<i>ACTBB</i>	<i>INHBB</i>	02cen–q13	23	<i>MHC1UAA</i>	<i>MHCI</i>	06p21.3
	<i>ENG1</i>	<i>EN1</i>	02q13–q21		<i>NTL</i>	<i>T</i>	06q27
	<i>EVX2</i>	<i>EVX2</i>	02q31–q32		<i>DLX6</i>	<i>DLX6</i>	07q22
12	<i>HOXDA</i>	<i>HOXD</i>	02q31–q2	25	<i>HOXAA</i>	<i>HOXA</i>	07p15–p14.2
	<i>DLX1</i>	<i>DLX1</i>	02q32		<i>SNAP25A</i>	<i>SNAP25</i>	20p11.2
	<i>DLX2</i>	<i>DLX2</i>	02q32		<i>BMP2B</i>	<i>BMP2</i>	20p12
13	<i>DESMIN</i>	<i>DES</i>	02q35	25	<i>POU2</i>	<i>POU5F1</i>	06p21.3
	<i>BRN1.1</i>	<i>POU3F3</i>	03p14.2		<i>BF</i>	<i>BF/C2</i>	06p21.3
	<i>ACTR2A</i>	<i>ACVR2B</i>	03p22–p21.3		<i>WNT1</i>	<i>WNT1</i>	12q12q–q13
14	<i>WNT11r</i>	<i>WNT11</i>	11q13.5	25	<i>PRIM1</i>	<i>PRIM1</i>	12q13
	<i>HSC70</i>	<i>HSPA8</i>	11q23.3–q25		<i>TARAMA</i>	<i>ACVR1B</i>	12q13
	<i>SMAD2</i>	<i>MADH2</i>	18q21		<i>HOXCA</i>	<i>HOXC</i>	12q13
15	<i>SMAD4</i>	<i>MADH4</i>	18q21.1	25	<i>RARG</i>	<i>RARG</i>	12q13
	<i>PLASTICIN</i>	<i>PRPH</i>	12q12–q13		<i>WNT10BR</i>	<i>WNT10B</i>	12q13
	<i>HOXCB</i>	<i>HOXC</i>	12q13		<i>WT1</i>	<i>WT1</i>	11p13
16				25	<i>PAX6</i>	<i>PAX6</i>	11p13
					<i>RAG1</i>	<i>RAG1</i>	11p13
					<i>MYOD</i>	<i>MYOD1</i>	11p15.4

This summary of genes that define conserved syntenies was compiled from the analysis shown in Tables 1 and 3 and, in the case of the Hox clusters, Amores et al. (1998).

shows that these expression differences reflect the distinct functions of these duplicate genes. Inactivation of *pax2.1* leads to the loss of the cerebellum, indicating that the later expression of *pax2.2* is insufficient for normal development of the CNS (Brand et al. 1996).

This and other recent examples (Kishimoto et al. 1997; Feldman et al. 1998; Schauerer et al. 1998) suggest that in many cases duplicate genes closely related in sequence have acquired distinct functions by virtue of their divergent expression patterns.



**Figure 4** Conservation of synteny among zebrafish, human, and mouse. Each point in the Oxford grids represents an orthologous gene pair plotted by map position in the two organisms compared. Conserved syntenies are evident as bins containing more than one point. (A) Comparison between zebrafish and human. (B) Comparison between zebrafish and mouse. (C) Comparison between human and mouse. Orthology relationships and map positions used to construct the Oxford grids are shown in Tables 1 and 3, except for the zebrafish *Hox* clusters whose orthology and position are from Amores et al. (1998). For tandem clusters of *Hox*, *globin*, and *MHC* genes, each cluster is represented by a single point in the Oxford grid and in the Poisson calculations. Tables associated with each grid evaluate the statistical significance of conserved syntenies. (C) Classes of bins according to the number of genes that occupy the bin; (Exp) the expected number of occurrences of bins in each class (see Methods for Poisson calculation); (Obs) the actual number we observed. To calculate the  $\chi^2$  value, the Exp and Obs were assigned to three categories, corresponding to classes of bins with 0, 1, or >1 gene occupying the bin. For 1 degree of freedom, a  $\chi^2 > 6.64$  implies a significant difference at a level  $P < 0.01$ .

## METHODS

### DNA Preparation and PCR

The mapping panel was prepared from 48 haploid individuals obtained from a single female derived from a cross between a Tü strain male and a TL strain female. The Tü and TL strains were described by Haffter et al. (1996). To extract genomic DNA, each embryo was placed in 50  $\mu$ l DNA preparation buffer [10 mM Tris-HCl (pH 8.3), 1.0 mM EDTA, 12.5 mM KCl, 0.3% Tween 20, 0.3% NP-40] in a microtiter plate well, heated to 98°C for 10 min, and incubated at 55°C overnight with proteinase K (1 mg/ml). The proteinase K was then inactivated by incubation at 98°C for 10 min. For a single PCR assay, DNA from 1/4000 of this DNA preparation was used as a template.

PCR amplification was performed in 12.5- $\mu$ l reaction mixtures containing 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.001% gelatin, 0.1 mg/ml BSA, 100  $\mu$ M each dNTP, 1  $\mu$ Ci of [ $\alpha$ -<sup>32</sup>P]dATP, 0.25 unit of *Taq* polymerase, and 100 nM each primer. Thermocycling was done under standard conditions consisting of an initial denaturation at 94°C for 2 min, followed by 45 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, and a final incubation at 72°C for 5 min.

### Gel Electrophoresis

SSLP PCR reactions were diluted 2:1 with loading buffer (80% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol), denatured at 95°C for 5 min, and electrophoresed on a 6% denaturing gel (19:1 acrylamide/bis). SSLP gels were run at

60 W at room temperature for 2.5 to 3 hr, transferred to Whatman filter paper, and exposed to film at -80°C with an intensifying screen.

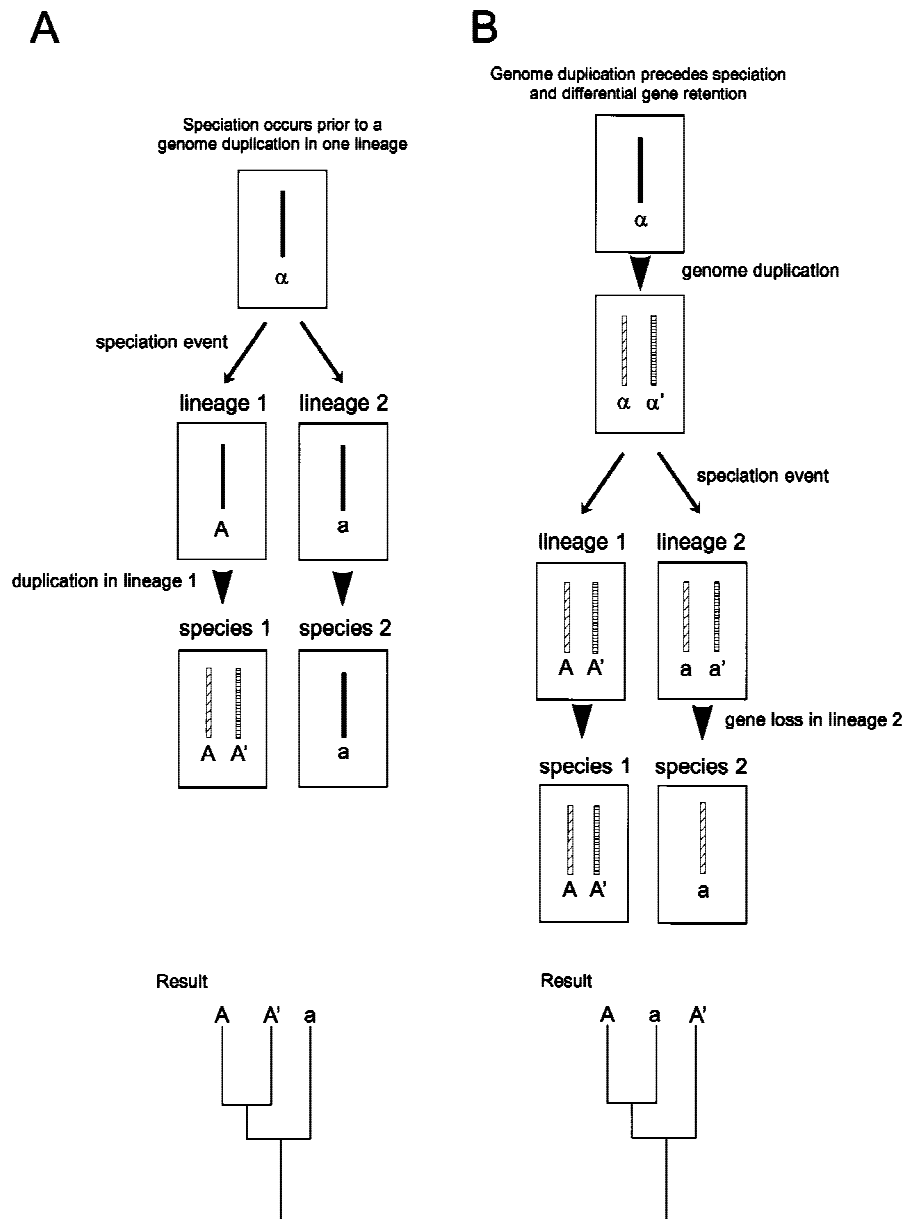
SSCP PCR reactions were diluted 4:1 with SSCP buffer (80% formamide, 10 mM NaOH, 0.1% bromophenol blue, 0.1% xylene cyanol) denatured at 98°C for 5 min, placed on ice for 5 min, and then immediately loaded on a 4.5% (39:1 acrylamide/bis) nondenaturing gel. SSCP gels were run at 40W at 4°C for 3 hr, transferred to Whatman filter paper, dried, and exposed to film at -80°C with an intensifying screen.

### Primer Design

The *D. rerio* subset of DNA sequence from the nonredundant (NR) database at the National Center for Biotechnology Information (NCBI, Bethesda, MD) was downloaded using Nentrez network tools. TIGR Assembler (Sutton et al. 1995) software was used to cluster the genes and ESTs in this set into contiguous sequence fragments (contigs). Primers used to develop SSCPs were designed from these contigs using Primer 3.0.6 (Rozen and Skaletsky 1997) and synthesized at the Stanford DNA Sequencing and Technology Center as described previously (Lashkari et al. 1997). Primers for SSLP markers were obtained from Research Genetics (Huntsville, AL).

### Phylogenetic Analysis

The NCBI NR database was used for searches. Each query sequence designated for ortholog assignment was aligned in a



**Figure 5** Two models explaining the origin of extra paralogs in one of two related species. (A) If a genome duplication occurs in lineage 1 after its divergence from lineage 2, then the resulting paralogous genomic elements (genes or chromosomes) A and A' of an extant species from lineage 1 are more closely related to each other than they are to their ortholog a in an extant species from lineage 2. This relationship is evident in the resulting phylogenetic tree, diagrammed at the bottom. (B) If a genome duplication produces paralogous genomic elements  $\alpha$  and  $\alpha'$ , then a subsequent speciation event will produce lineage 1 with paralogs A and A' and lineage 2 with paralogs a and a'. A and a are orthologs, as they are both directly descended from  $\alpha$ , and A' and a', descended from  $\alpha'$ , are also orthologs. Lineage-specific loss of paralogs (e.g. a' in lineage 2) results in species 2 having fewer paralogs than species 1. The derivation of A in species 1 and a in species 2 from their common ancestor  $\alpha$  is evident in the resulting phylogenetic tree, diagrammed at the bottom.

pairwise manner (Needleman and Wunsch 1970; zero end gap penalties, gonnet matrix, gap opening penalty 2.4, gap extension penalty 0.15) with every sequence in the database and the alignments ranked according to their ZEGA similarity

scores (ZEGA probability; Abagyan and Batalov 1997). An all-against-all protein domain search of the SCOP database (Murzin et al. 1995) indicated that a ZEGA probability of 0.00001 would not eliminate homologs even in weak identity cases (not shown). This cutoff, or a total of 50 sequences for those cases in which many homologs existed in the database, was used to group the top scoring sequences from the database search. This group of sequences was then multiply aligned, and neighbor-joining trees incorporating evolutionary distances were extracted from the final alignment using the CLUSTALX program (Thompson et al. 1997). Trees were tested by analysis of 1000 bootstrap replicates. The resulting trees were examined to identify orthologs.

#### Data Collection and Analysis

Data were collected using Map Manager software (Manly 1993; <http://mcbio.med.buffalo.edu/mapmgr.html>), and analysis was conducted with Map Manager and Map Maker software (Lander et al. 1987). All loci, including published markers, were initially analyzed and ordered independently of published map positions. Each locus was initially placed at the position that maximized its lod score as reported by the "Links" command of Map Manager. Local order was then determined by manually placing the marker at the location that minimized the number of double recombinants. Data were then exported to Map Maker to confirm map order using the "ripple" and "try" features. Data analysis ordered 389 of the 390 loci into 50 multilocus groups that were supported by lod scores  $>3.0$ . The 50 groups were then assigned to the standard 25 linkage groups based upon information from previously published map locations. The 50 separate groups were positioned by choosing the order and orientation that maximized lod score and minimized the number of recombinants between the end loci of respective groups. Map graphics were created with Map

Maker using the Kosambi mapping function.

For Figure 4, the expected number (Exp) of occurrences of bins of a given class were calculated according to a Poisson distribution:  $\text{Exp} = n\{(e^{-\lambda})(\lambda^C)/C!\}$ , where C = class according

to number of genes occupying a bin, and  $n$  = number of bins. The parameter  $\lambda$  was estimated as  $\lambda = k/n$ , where  $k$  = total number of data points. For zebrafish–human comparison,  $n = 600$  and  $k = 124$ ; for zebrafish–mouse comparison,  $n = 525$  and  $k = 135$ ; for human–mouse comparison,  $n = 504$  and  $k = 103$ . The Exp values shown in Figure 4 were rounded to the nearest integer. These calculations make the simplifying assumption that all chromosomes in a given species are of equal length.

## Error Analysis

Several steps were taken to monitor the quality of the data, particularly those generated for SSCP markers. The combination of the complex banding patterns possible in SSCP coupled with the occasional ability of some primer pairs to amplify multiple products led us to adopt strict criteria when considering a locus for analysis. Polymorphisms were scored only for the predominant banding pattern on the gel, and we ensured that there were no inconsistencies between the multiple bands that were assumed to represent a given SSCP allele. Finally, putative polymorphisms for which >6% of the individuals (3/48) were assigned different alleles in independent scorings were not further considered for mapping. Discrepant data points were left as unscored. The error rate of mapping is often correlated to the number of double crossovers (particularly in small regions) in the data set. In the first complete evaluation of the data set, there were 27 double recombinants, 7 of which occurred in regions of <20 cM. The genotype assays for all 27 points were repeated and independently rescored. We found that 5 of the 7 small-interval double recombinants were mis-scored in the original assays and confirmed the original genotypes of the other 22 double recombinants. Thus the final error analysis left two small-interval double recombinants present in the dataset.

## ACKNOWLEDGMENTS

We thank the members of the Talbot and Schier laboratories for helpful discussions; R. Burdine, B. Feldman, E. Heckscher, D. Kingsley, and J. Postlethwait for critical comments on the manuscript; and Ron Davis for encouragement and support. We acknowledge fellowship support from the National Institutes of Health (NIH) (H.I.S) and American Cancer Society (S.T.D.). W.S.T. is a Pew Scholar in Biomedical Science. This work was supported by NIH grants R01 RR12349 (W.S.T.) and R21 HG01704 (A.F.S.).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

Abagyan, R. and S. Batalov. 1997. Do aligned sequences share the same fold? *J. Mol. Biol.* **273**: 355–368.

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

Amores, A., A. Force, Y.L. Yan, L. Joly, C. Amemiya, A. Fritz, R.K. Ho, J. Langeland, V. Prince, Y.L. Wang et al. 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**: 1711–1714.

Beier, D.R. 1998. Zebrafish: Genomics on the fast track. *Genome Res.* **8**: 9–17.

Bisbee, C.A., M.A. Baker, A.C. Wilson, I. Haji-Azimi, and M. Fischberg. 1997. Albumin phylogeny for clawed frogs (*Xenopus*). *Science* **195**: 785–787.

Brady, K.P., L.B. Rowe, H. Her, T.J. Stevens, J. Eppig, D.J. Sussman, J. Sikela, and D.R. Beier. 1997. Genetic mapping of 262 loci derived from expressed sequences in a murine interspecific cross using single-strand conformational polymorphism analysis. *Genome Res.* **7**: 1085–1093.

Brand, M., C.-P. Heisenberg, Y.-J. Jiang, D. Beuchle, K. Lun, M. Furutani-Seiki, M. Granato, P. Haffter, M. Hammerschmidt, D. Kane et al. 1996. Mutations in zebrafish genes affecting the formation of the boundary between midbrain and hindbrain. *Development* **123**: 179–190.

Brooke, N.M., J. Garcia-Fernandez, P.W. Holland. 1998. The *ParaHox* gene cluster is an evolutionary sister of the *Hox* gene cluster. *Nature* **392**: 920–922.

Driever, W., L. Solnica-Krezel, A.F. Schier, S.C.F. Neuhauss, J. Malicki, D.L. Stemple, D.Y.R. Stainier, F. Zwartkruis, S. Abdelilah, Z. Rangini et al. 1996. A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**: 37–46.

Ekker, M., J. Wegner, M.A. Akimenko, and M. Westerfield. 1992. Coordinate embryonic expression of three zebrafish engrailed genes. *Development* **116**: 1001–1010.

Ekker, S.C., A.R. Ungar, P. Greenstein, D.P. von Kessler, J.A. Porter, R.T. Moon, and P.A. Beachy. 1995. Patterning activities of vertebrate *hedgehog* proteins in the developing eye and brain. *Curr. Biol.* **5**: 944–955.

Ekker, M., M.A. Akimenko, M.L. Allende, R. Smith, G. Drouin, R.M. Langille, E.S. Weinberg, and M. Westerfield. 1997. Relationships among *msx* gene structure and function in zebrafish and other vertebrates. *Mol. Biol. Evol.* **14**: 1008–1022.

Fekany, K., Y. Yamanaka, T. Leung, H.I. Sirotkin, J. Topczewski, M.A. Gates, M. Hibi, A. Renucci, D. Stemple, A. Radbill et al. 1999. The zebrafish *bozozok* locus encodes Dharma, a homeodomain protein essential for induction of the gastrula organizer and dorsoanterior embryonic structures. *Development* **126**: 1427–1438.

Feldman, B., M.A. Gates, E.S. Egan, S.T. Dougan, G. Rennebeck, H.I. Sirotkin, A.F. Schier, and W.S. Talbot. 1998. Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* **395**: 181–185.

Fisher, S., S.L. Amacher, and M.E. Halpern. 1997. Loss of *cerebrum* function ventralizes the zebrafish embryo. *Development* **124**: 1301–1311.

Förnizler, D., H. Her, E.W. Knapik, M. Clark, H. Lehrach, J.H. Postlethwait, L.I. Zon, and D. R. Beier. 1998. Gene mapping in zebrafish using single-strand conformation polymorphism analysis. *Genomics* **51**: 216–222.

Force, A., M. Lynch, F.B. Pickett, A. Amores, Y.-L. Yan, and J.H. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* (in press).

Franklin J.I. and T.D. Sargent. 1996. Ventral neural cadherin, a novel cadherin expressed in a subset of neural tissues in the zebrafish embryo. *Dev. Dyn.* **206**: 121–130.

Goff, D.J., K. Galvin, H. Katz, M. Westerfield, E.S. Lander, and C.J. Tabin. 1992. Identification of polymorphic simple sequence repeats in the genome of the zebrafish. *Genomics* **14**: 200–202.

Haffter, P., M. Granato, M. Brand, M.C. Mullins, M. Hammerschmidt, D.A. Kane, J. Odenthal, F.J.M. van Eeden, Y.-J. Jiang, C.-P. Heisenberg et al. 1996. The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**: 1–36.

Holland, P.W.H. and J. Garcia-Fernández. 1996. *Hox* genes and chordate evolution. *Dev. Biol.* **173**: 382–395.

Johnson, S.L., M.A. Gates, M. Johnson, W.S. Talbot, S. Horne, K. Baik, S. Rude, J.R. Wong, and J.H. Postlethwait. 1996. Centromere-linkage analysis and consolidation of the zebrafish genetic map. *Genetics* **142**: 1277–1288.

Karlstrom, R.O., W.S. Talbot, and A.F. Schier. 1999. Comparative

- synteny cloning of zebrafish *you-too*: Mutations in the Hedgehog target *gli2* affect ventral forebrain patterning. *Genes & Dev.* **13**: 388–393.
- Kimmel, C.B. 1989. Genetics and early development of zebrafish. *Trends Genet.* **5**: 283–288.
- Kishimoto, Y., K.H. Lee, L. Zon, M. Hammerschmidt, and S. Schulte-Merker. 1997. The molecular nature of zebrafish *swirl*: BMP2 function is essential during early dorsoventral patterning. *Development* **124**: 4457–4466.
- Knapik, E.W., A. Goodman, M. Ekkt, M. Chevrette, J. Delgado, S. Neuhauss, N. Shimoda, W. Driever, M.C. Fishman, and H.J. Jacob. 1998. A microsatellite genetic linkage map for zebrafish. *Nat. Genet.* **18**: 338–343.
- Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, and L. Newburg. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.
- Lashkari, D.A., J.H. McCusker, and R.W. Davis. 1997. Whole genome analysis: Experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR. *Proc. Natl. Acad. Sci.* **94**: 8945–8947.
- 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.
- Manly, K.F. 1993. A Macintosh program for storage and analysis of experimental genetic mapping data. *Mamm. Genome* **4**: 303–313.
- Martinez-Barbera, J.P., H. Toresson, S. Da Rocha, and S. Krauss. 1997. Cloning and expression of three members of the zebrafish Bmp family: Bmp2a, Bmp2b and Bmp4. *Gene* **198**: 53–59.
- Mori, H., Y. Miyazaki, T. Morita, H. Nitta, and M. Mishina. 1994. Different spatio-temporal expressions of three *otx* homeoprotein transcripts during zebrafish embryogenesis. *Brain Res. Mol. Brain Res.* **27**: 221–231.
- Morizot, D.C. 1990. Use of fish gene maps to predict ancestral vertebrate genome organization. In *Isozymes: Structure, function, and use in biology and medicine*, pp. 207–234. Wiley/Liss, New York, NY.
- Murzin, A.G., S.E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**: 536–540.
- Needleman, S.B. and C.D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**: 443–453.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag, Heidelberg, Germany.
- Pfeffer, P.L., T. Gerster, K. Lun, M. Brand, and M. Busslinger. 1998. Characterization of three novel members of the zebrafish Pax2/5/8 family: Dependency of Pax5 and Pax8 expression on the Pax2.1 (noi) function. *Development* **125**: 3063–3074.
- Postlethwait, J.H., S. Johnson, C.N. Midson, W.S. Talbot, M. Gates, E.W. Ballinger, D. Africa, R. Andrews, T. Carl, J.S. Eisen et al. 1994. A genetic map for the zebrafish. *Science* **264**: 699–703.
- Postlethwait, J.H. and W.S. Talbot. 1997. Zebrafish genomics: From mutants to genes. *Trends Genet.* **13**: 183–190.
- Postlethwait, J.H., Y.-L. Yan, M.A. Gates, S. Horne, A. Amores, A. Brownlie, A. Donovan, E.S. Egan, A. Force, Z. Gong et al. 1998. Vertebrate genome evolution and the zebrafish gene map. *Nat. Genet.* **18**: 345–349.
- Rozen, S. and H.J. Skaletsky. 1997. Primer 3.0 (<http://www.genome.wi.mit.edu>). Whitehead Institute, Cambridge, MA.
- Schauerte, H. E., F. van Eeden, C. Fricke, J. Odenthal, U. Strahle, and P. Hafter. 1998. *Sonic hedgehog* is not required for the induction of medial floor plate cells in the zebrafish. *Development* **125**: 2983–2993.
- Schier, A.F. and W.S. Talbot. 1998. The zebrafish organizer. *Curr. Opin. Gen. Dev.* **8**: 464–471.
- Schier, A.F., S.C.F. Neuhauss, K.A. Helde, W.S. Talbot, and W. Driever. 1997. The *one-eyed pinhead* gene functions in mesoderm and endoderm formation in zebrafish and interacts with *no tail*. *Development* **124**: 327–342.
- Schulte-Merker, S., F.J. van Eeden, M.E. Halpern, C.B. Kimmel, and C. Nüsslein-Volhard. 1994. *no tail (ntl)* is the zebrafish homologue of the mouse T (*Brachyury*) gene. *Development* **120**: 1009–1015.
- Sefton, M., S. Sanchez, and M.A. Nieto. 1998. Conserved and divergent roles for members of the Snail family of transcription factors in the chick and mouse embryo. *Development* **125**: 3111–3121.
- Sidow, A. 1996. Gen(om)e duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev.* **6**: 715–722.
- Stock, D.W., D.L. Ellies, Z. Zhao, M. Ekker, F.H. Ruddle, and K.M. Weiss. 1996. The evolution of the vertebrate Dlx gene family. *Proc. Natl. Acad. Sci.* **93**: 10858–10863.
- Sutton, G., O. White, M. Adams, and A. Kerlavage. 1995. TIGR Assembler: A new tool for assembling large shotgun sequencing projects. *Genome Sci. & Tech.* **1**: 9–19.
- Talbot, W.S., B. Trevarrow, M.E. Halpern, A.E. Melby, G. Farr, J.H. Postlethwait, T. Jowett, C.B. Kimmel, and D. Kimelman. 1995. A homeobox gene essential for zebrafish notochord development. *Nature* **378**: 150–157.
- Talbot, W.S., E.S. Egan, M.A. Gates, C. Walker, B. Ullmann, S.C.F. Neuhauss, C.B. Kimmel, and J.H. Postlethwait. 1998. Genetic analysis of chromosomal rearrangements in the *cyclops* region of the zebrafish genome. *Genetics* **148**: 373–380.
- Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.
- Wittbrodt, J., A. Meyer, and M. Schartl. 1998. More genes in fish? *BioEssays* **20**: 511–515.
- Young, W.P., P.A. Wheeler, V.H. Coryell, P. Keim, and G.H. Thorgaard. 1998. A detailed linkage map of rainbow trout produced using doubled haploids. *Genetics* **148**: 839–850.
- Zhang, J., W.S. Talbot, and A.F. Schier. 1998. Positional cloning identifies zebrafish *one-eyed pinhead* as a permissive EGF-related ligand required during gastrulation. *Cell* **92**: 241–251.

Received December 18, 1998; accepted in revised form February 9, 1999.