A Comparative Map of the Zebrafish Genome

Ian G. Woods,[1] Peter D. Kelly,[1] Felicia Chu,[1] Phuong Ngo-Hazelett,[2] Yi-Lin Yan,[2] Hui Huang,[1] John H. Postlethwait,[2] and William S. Talbot[1,3]

[1]Department of Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA;
[2]Institute of Neuroscience, University of Oregon, Eugene, Oregon 97405, USA

Zebrafish mutations define the functions of hundreds of essential genes in the vertebrate genome. To accelerate the molecular analysis of zebrafish mutations and to facilitate comparisons among the genomes of zebrafish and other vertebrates, we used a homozygous diploid meiotic mapping panel to localize polymorphisms in 691 previously unmapped genes and expressed sequence tags (ESTs). Together with earlier efforts, this work raises the total number of markers scored in the mapping panel to 2119, including 1503 genes and ESTs and 616 previously characterized simple-sequence length polymorphisms. Sequence analysis of zebrafish genes mapped in this study and in prior work identified putative human orthologs for 804 zebrafish genes and ESTs. Map comparisons revealed 139 new conserved syntenies, in which two or more genes are on the same chromosome in zebrafish and human. Although some conserved syntenies are quite large, there were changes in gene order within conserved groups, apparently reflecting the relatively frequent occurrence of inversions and other intrachromosomal rearrangements since the divergence of teleost and tetrapod ancestors. Comparative mapping also shows that there is not a one-to-one correspondence between zebrafish and human chromosomes. Mapping of duplicate gene pairs identified segments of 20 linkage groups that may have arisen during a genome duplication that occurred early in the evolution of teleosts after the divergence of teleost and mammalian ancestors. This comparative map will accelerate the molecular analysis of zebrafish mutations and enhance the understanding of the evolution of the vertebrate genome.

A challenge raised by the rapid progress of genome sequencing projects is to define the functions of the 40,000–70,000 genes that constitute the vertebrate genome (Dunham et al. 1999; Hattori et al. 2000). Genetic screens in zebrafish (*Danio rerio*) have identified mutations that define the functions of hundreds of essential genes—a number that is increasing rapidly as new screens are conducted to explore specific mutant phenotypes (Driever et al. 1996; Haffter et al. 1996). Because vertebrates share certain fundamental similarities, the identification of zebrafish mutations can provide important insights about genes with functions that are conserved in other vertebrates, including humans. Indeed, numerous zebrafish mutant phenotypes resemble human disease conditions and, in several cases, it is clear that the similar abnormalities result from inactivation of orthologous genes in the two species (Brownlie et al. 1998; Wang et al. 1998; Childs et al. 2000). Powerful cellular methods, such as cell labeling, transplantation, and microinjection, enable a detailed understanding of zebrafish mutant phenotypes (e.g., Melby et al. 1996; Moens et al. 1996). Thus, phenotypic analysis in zebrafish can provide functional information that is difficult to obtain in other species.

Recent efforts have developed maps and other genomic infrastructure to accelerate the molecular analysis of zebrafish mutations (Talbot and Hopkins 2000). More than 2000 simple-sequence length polymorphism (SSLP) markers have been meiotically mapped, allowing the rapid localization of mutations and providing a large pool of possible entry points for positional cloning projects (Knapik et al. 1998; Shimoda et al. 1999). Meiotic and radiation hybrid (RH) mapping projects have localized >1100 genes and expressed sequence tags (ESTs) that can be rapidly tested as candidates for mutations (Postlethwait et al. 1998; Gates et al. 1999; Geisler et al. 1999; Hukriede et al. 1999; Kelly et al. 2000). More than 25 mutated genes have been cloned by the candidate gene approach, emphasizing the utility of zebrafish gene maps for the molecular analysis of mutations (see Talbot and Hopkins 2000).

Gene maps are uniquely valuable for comparative studies, which have identified groups of genes that are syntenic (on a single chromosome) in zebrafish and human (for review, see Postlethwait et al. 1999; Meyer and Schartl 1999). Previous comparative studies have identified 28 groups of two or more genes that are syntenic in zebrafish and human, suggesting that these genes were syntenic in the last common ancestor and that this relationship has been preserved since the divergence of the lineages leading to zebrafish and humans (Postlethwait et al. 1998; Gates et al. 1999). Analysis of these conserved segments has facilitated the selection of candidate genes for zebrafish mutations (Karlstrom et al. 1999; Schmid et al. 2000; Miller

et al. 2000). Despite the utility of current comparative maps, much additional work is required to discover the complete set of conserved syntenies and to learn the extent to which gene order has been preserved within syntenic groups. Additional comparative studies are also required to address the hypothesis that a genome duplication occurred in the lineage leading to modern teleosts after the split of teleost and tetrapod ancestors (Amores et al. 1998; Postlethwait et al. 1998; Gates et al. 1999; Meyer and Schartl 1999).

Here we report the genetic mapping of 691 previously unlocalized genes and ESTs and a comparative analysis identifying the human counterparts of 804 mapped zebrafish genes and ESTs. The comparative analysis has identified 139 new syntenies conserved between zebrafish and human, raising the total to 167. We identified conserved syntenies on all zebrafish linkage groups, substantially expanding the reach of comparative maps. Several zebrafish linkage groups have conserved syntenies with multiple human chromosomes, an observation that we consider in detail in the accompanying paper (Postlethwait et al. 2000). In addition, the mapping of duplicate genes has defined the locations of 13 pairs of chromosomal segments that may have arisen in a genome duplication that occurred early in the evolution of teleosts. This comparative map will accelerate the molecular analysis of zebrafish mutations and enhance the understanding of the evolution of the vertebrate genome.

RESULTS

Genetic Mapping of Genes and ESTs

To meiotically map zebrafish genes and expressed sequences, we scored 754 single-stranded conformational polymorphisms that correspond to genes and ESTs in the heat shock (HS) panel, a group of 42 homozygous diploid individuals generated by heat shock treatment of F_2 individuals from a cross of the divergent strains C32 and SJD (Kelly et al. 2000). Of these polymorphisms, 691 represent previously unmapped genes and ESTs and 63 have been localized in previous genetic or radiation hybrid maps (Postlethwait et al. 1998; Gates et al. 1999; Geisler et al. 1999; Hukriede et al. 1999). Locus names, accession numbers, and primer sequences for the 1503 genes and ESTs mapped in the HS panel in this paper and in previous work (Kelly et al. 2000) are shown in Table 1 (available as supplementary material at http://www.genome.org). Of these genes and ESTs, 853 represent UniGene clusters (Table 1) and are therefore likely to correspond to different genes (http://www.ncbi.nlm.nih.gov/UniGene/Dr.Home.html). Sequence comparisons did not reveal significant overlap among the 650 ESTs that are not in UniGene clusters, which suggests that most of these also represent unique genes. This work raises the total number of markers scored in the HS panel to 2119, including 1503 genes and ESTs and 616 previously mapped SSLP markers (Shimoda et al. 1999). The dataset contains a total of 84,674 genotypes, an average of 40 individuals scored for each marker.

Linkage analysis assigned all of the genes and ESTs positions supported by LOD scores of ≥3 (Fig. 1). The 2119 markers that have been scored in the HS panel occupied 743 unique map positions. In total, the map spanned 3004 cM and the average distance between groups of markers was 4 cM (3004 cM/743 unique map positions).

Comparative Analysis

Comparative genome mapping relies on the identification of orthologous genes—loci in two different species descended from a single locus in the last common ancestor of the two species. Orthologs are best identified by their branching patterns on phylogenetic trees but, because many of the mapped ESTs have relatively short coding regions, this is not a practicable methodology for ESTs. Therefore, we have employed criteria used by the HomoloGene database (http://www.ncbi.nlm.nih.gov/HomoloGene/), that is, putative orthologs are UniGene (http://www.ncbi.nlm.nih.gov/UniGene/) clusters with strong matches in reciprocal BLAST searches between zebrafish and mammals (see Methods). Using these criteria, we identified 804 putative orthologs between human and zebrafish and 388 putative orthologs between mouse and zebrafish in our dataset and in previous comparative maps (Postlethwait et al. 1998; Gates et al. 1999). To assess the reliability of assigning putative orthologs with EST sequences in this manner, we compared ortholog assignments derived from ESTs and completely sequenced cDNAs representing the same genes. In the analysis of 95 ESTs representing 43 different genes, there was no case in which an ortholog assignment derived from an EST differed from the assignment derived from the corresponding completely sequenced cDNA. Thus it seems that reciprocal BLAST searches with ESTs can provide useful information about possible orthologs despite the uncertainty inherent in comparative analyses conducted with these fragmentary sequences.

A convenient display of conserved syntenies is the Oxford grid, which arrays putative orthologs from two species according to chromosome for each species (Edwards 1991). An important feature of comparative genetic maps is the number of conserved syntenies, which are reflected directly in the number of boxes in the Oxford grid that have multiple entries. This display (Fig. 2) shows that, despite scatter, the distribution is distinctly nonrandom ($\chi^2 = 158$; $P < 0.001$), with clusters appearing, for example, at human chromosome
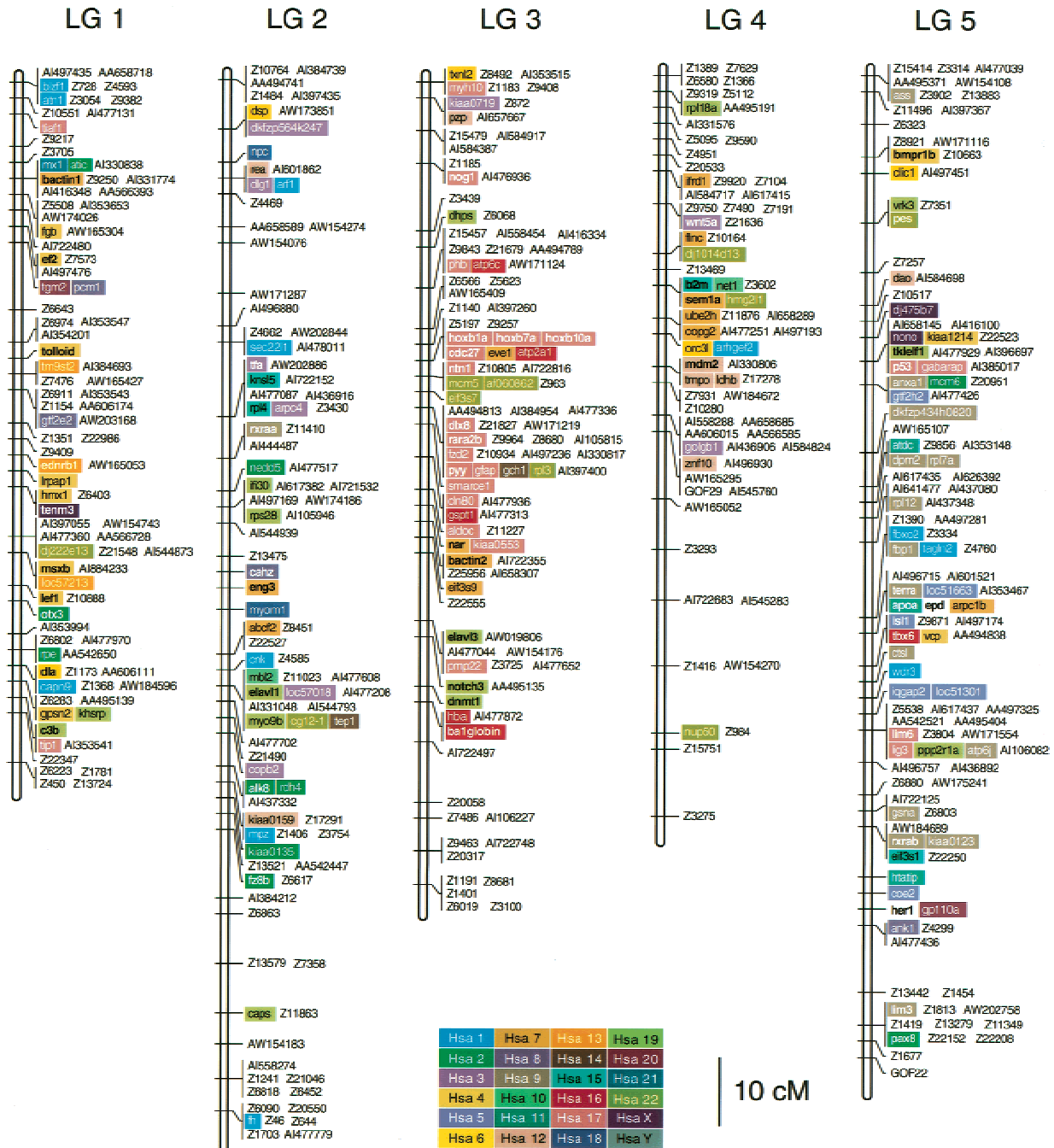
Figure 1 (continues)

(Hsa) 2, 6, and 9, Hsa 6/LG 19 and 20, Hsa 9/LG 5, Hsa 12/LG 23, Hsa 14/LG 17 and 20, and Hsa 17/LG 3, 12, and 15. Zebrafish and human shared 167 conserved syntenies involving two or more putatively orthologous gene pairs in the dataset (Fig. 2; Table 2, available as supplementary material at http://www.genome-.org). Although some of these clusters, especially those that involve only two orthologous gene pairs, might reflect incorrect ortholog assignments or ancestrally

nonsyntenic genes that independently joined the same chromosomes in fish and mammalian lineages, most of these gene groups were likely to have been syntenic in the last common ancestor of zebrafish and human. The analysis showed that there were 136 loci not in conserved syntenies between zebrafish and human among the 804 putatively orthologous gene pairs, so that 83.1% of the orthologs are in conserved syntenies. To put this in perspective, 90.4% of the 375 ortholo-
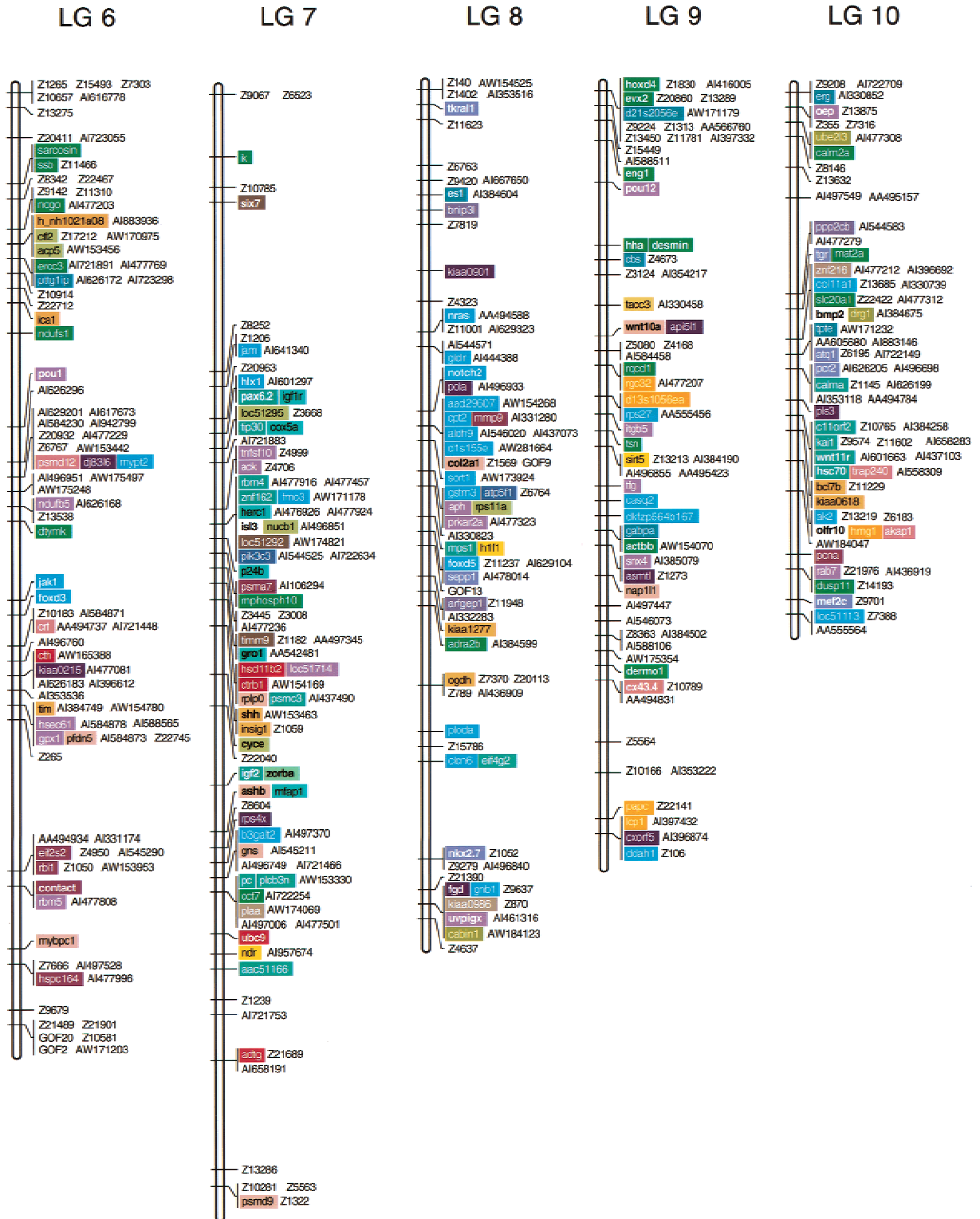
LG 6   LG 7   LG 8   LG 9   LG 10

Figure 1   (continues)

gous pairs between mouse and human in our data-set were in conserved syntenies. We conclude that there is extensive conservation of syntenies between zebrafish and mammals, but that mouse and hu-man, which diverged ~112 million years ago, have greater conservation than zebrafish and human, which
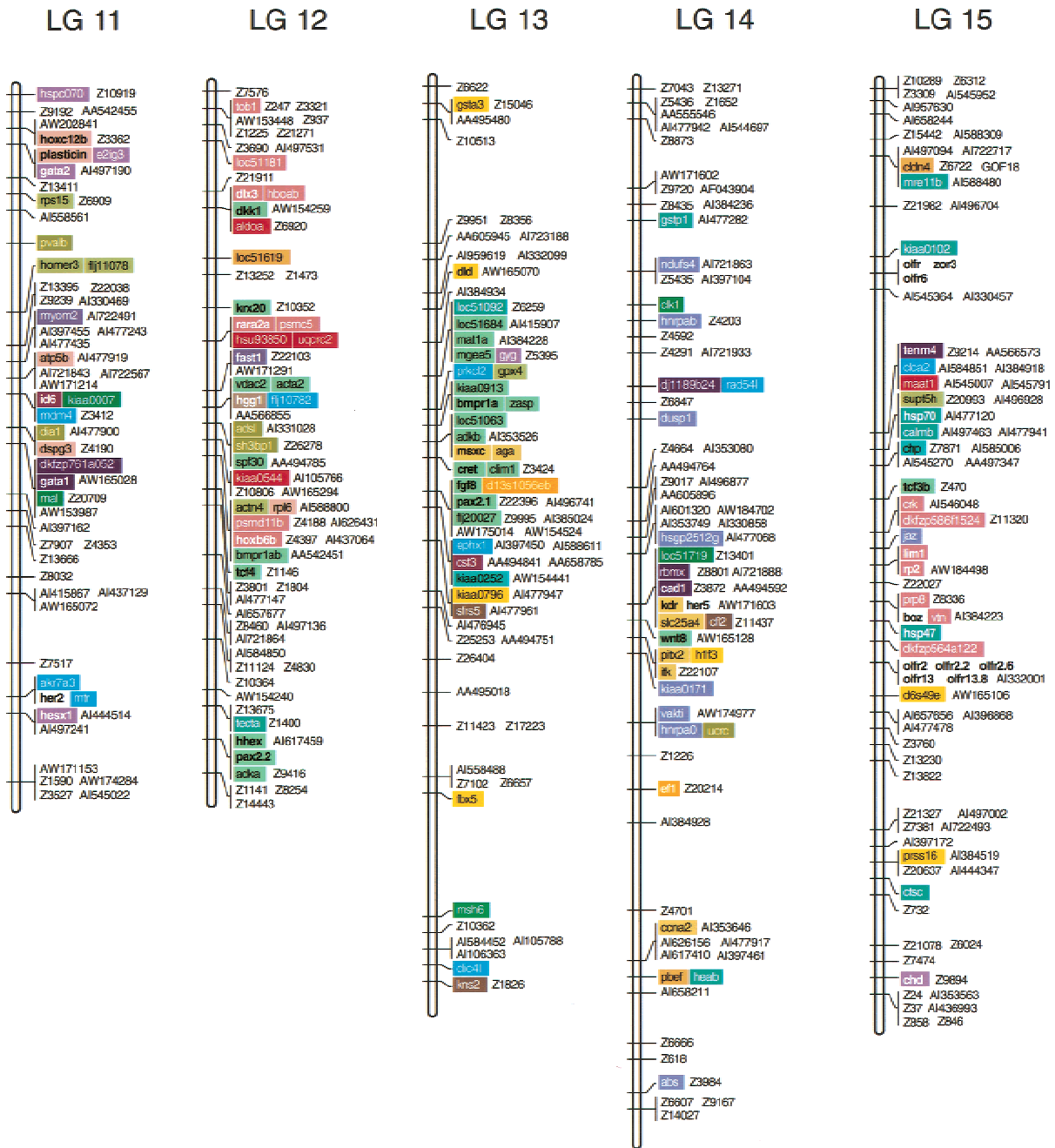
Figure 1 (continues)

diverged ~450 million years ago (Kumar and Hedges 1998).

To investigate the extent to which gene order has been maintained in regions of conserved synteny, we examined the positions of genes in two of the large syntenic groups that were identified by the comparative analysis (Fig. 3). Putative human orthologs of 12 zebrafish genes and ESTs on LG 17 are distributed over the length of Hsa 14 (Fig. 3A), which suggests that a large number of the genes on these chromosomes were syntenic in the last common ancestor of zebrafish and human. There were changes in gene order, however, and significant intrachromosomal rearrangements have apparently occurred in the fish and/or mammalian lineages since their divergence. A similar conclusion derives from the analysis of LG 8 and Hsa 1 (Fig. 3B). The putative human orthologs of 11 genes and ESTs on LG 8 are spread over the short arm of Hsa 1, which joined the long arm of Hsa 1 after the divergence of primates and carnivores (Murphy et al. 2000).
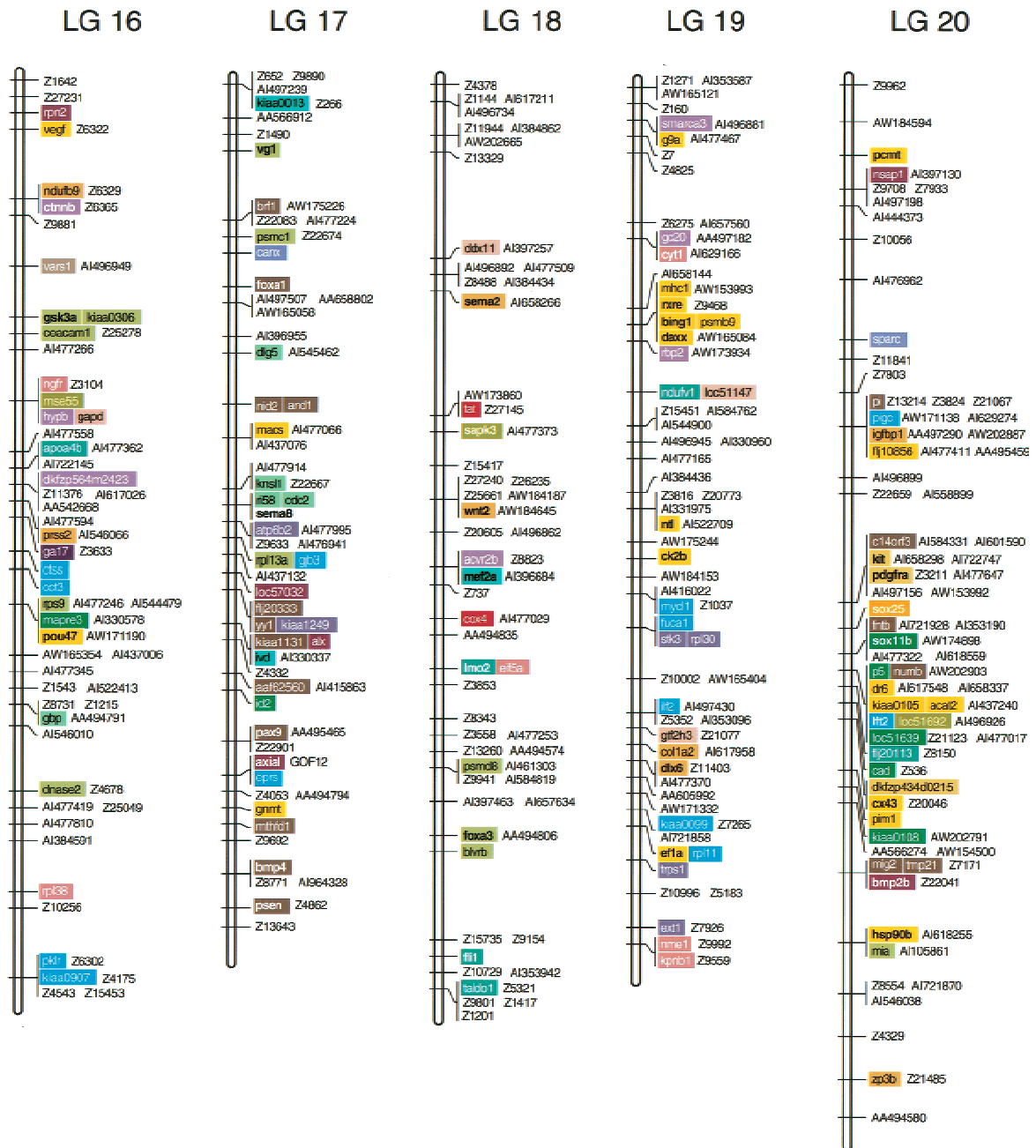
Figure 1 (continues)

This indicates that the genes on this chromosome arm comprise an ancient, conserved synteny but it is clear that intrachromosomal rearrangements have altered gene order within this region.

To identify chromosomal segments that might have resulted from the teleost genome duplication, we analyzed the positions of pairs of apparent duplicate zebrafish genes (Fig. 4). Putative duplicates were identified in the comparative analysis described above as cases in which two zebrafish genes appeared to be or-

thologous to a single gene in mammals. The set of putative duplicate genes identified by our analysis and in previous work (Amores et al. 1998; Postlethwait et al. 1998; Gates et al. 1999; Geisler et al. 1999) contained map positions for 59 pairs of putative duplicate genes (Table 3, available as supplementary material at http://www.genome.org). A graph of the positions of these genes revealed possible duplicate chromosomal segments as boxes containing multiple duplicate gene pairs (Fig. 4). The points were clearly clustered in a
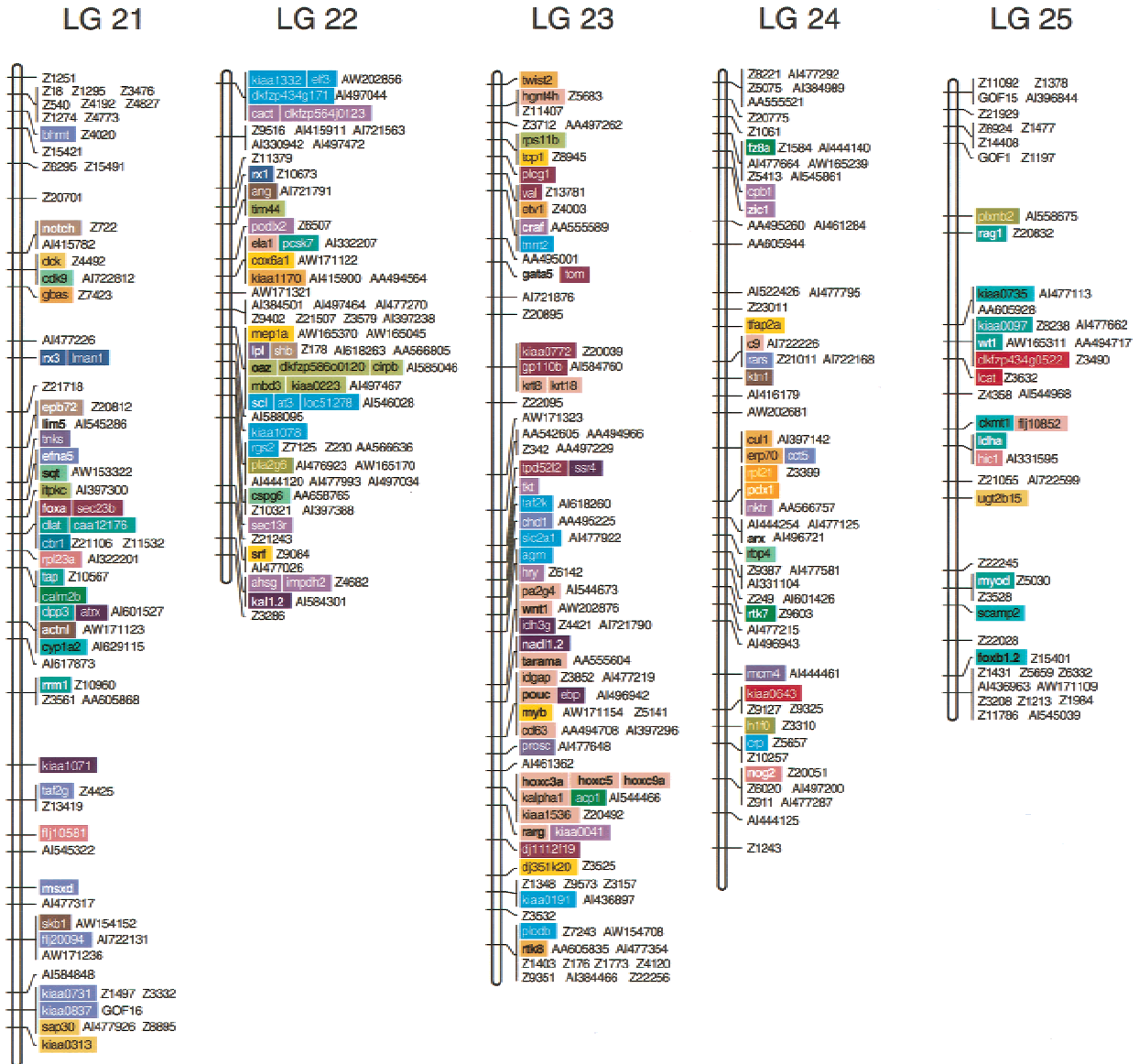
Figure 1 Genetic linkage map of the zebrafish genome. The positions of 2119 polymorphic markers mapped in the heat shock (HS) panel are shown. Genes and expressed sequence tags (ESTs) with identified human orthologs are color coded according to the chromosomal position of the human gene. Gene names are shown in bold type. According to zebrafish convention (http://zfin.org/zf_info/nomen.html), ESTs with putative human orthologs are named according to their human counterparts. Other ESTs are named by their GenBank accession numbers. Simple-sequence length polymorphism names begin with "Z" or "GOF", following the nomenclature of Shimoda et al. (1999). This map reports revised positions for three genes, *tbx6*, *isl3*, and *six7*, that were assigned to different locations in previous maps (Postlethwait et al. 1998; Geisler et al. 1999; Kelly et al. 2000). In addition, we have revised two previous assignments (Kelly et al. 2000) for ESTs, and we have removed from the dataset eight ESTs that were originally mapped in the HS panel because of discrepancies detected in repeat genotype assays. Sequence comparisons of the ESTs mapped in the HS panel by Kelly et al. (2000) revealed 14 cases in which two different ESTs that were likely to correspond to different regions of the same gene were mapped to the identical positions. We have removed one member of each of these putative duplicate pairs from the dataset.

nonrandom manner ($\chi^2 = 34.6$; $P < 0.001$) and there were 13 segments that contained two or more duplicate gene pairs, including LG 5–LG 21, LG 7–LG 25, LG 11–LG 23, LG 16–LG 19, and LG 3–LG 12.

DISCUSSION

We have genetically mapped 691 previously unlocalized genes and ESTs by scoring polymorphisms in the HS meiotic mapping panel. This raises the total number of markers mapped in the HS panel to 2119, of which 1503 are genes and ESTs. Sequence comparisons suggest that most of these ESTs define unique genes. Together with previous SSLP and gene mapping efforts, this work brings the total number of genetically
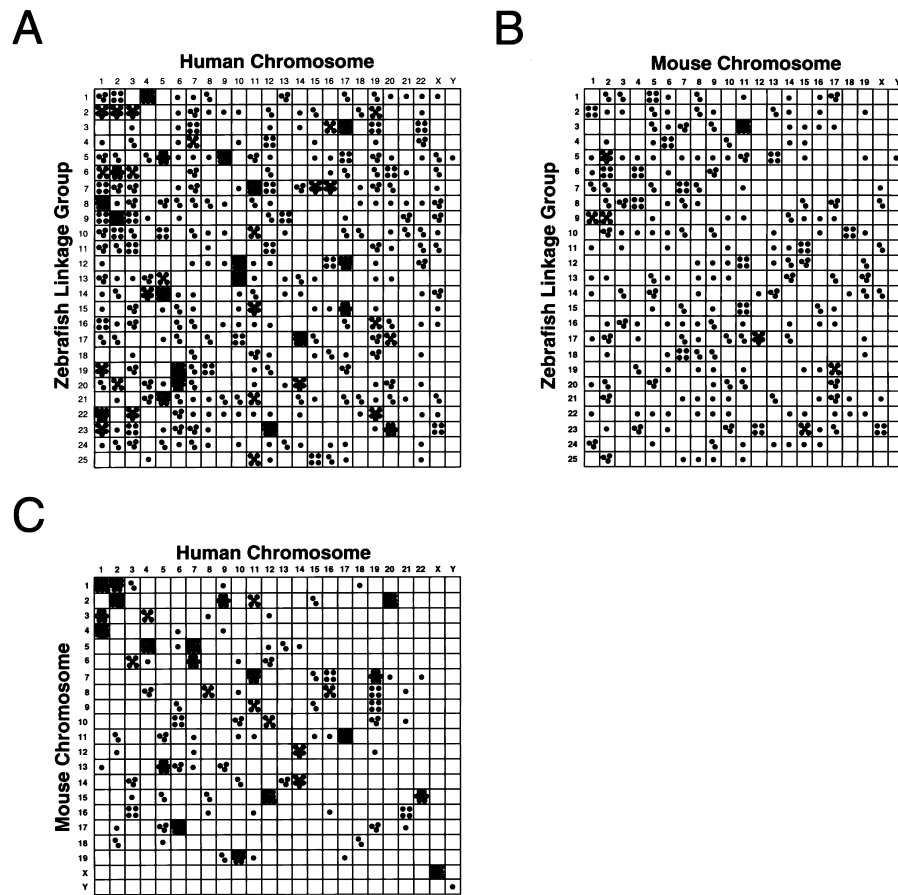
A

B

C

Figure 2 Oxford grids showing conservation of synteny among zebrafish, human, and mouse. Each dot represents an orthologous gene pair plotted by position in the two species compared. Boxes that contain more than one dot represent conserved syntenies. (*A*) Zebrafish–human comparison; (*B*) Zebrafish–mouse comparison; (*C*) Human–mouse comparison. Human and mouse orthologs of zebrafish genes mapped in the heat shock panel are listed in Table 1. Orthologs of genes mapped in previous work (Postlethwait et al. 1998; Gates et al. 1999) are also included in these grids. The genes in the zebrafish–human grid are summarized in Table 2. All genes in a *Hox* complex are represented by a single dot. All three grids clearly show a nonrandom distribution; χ^2 values, calculated according to Gates et al. (1999) are: $\chi^2 = 158$, zebrafish–human; $\chi^2 = 26.0$, zebrafish–mouse; and $\chi^2 = 220$, human–mouse. For 1 degree of freedom, a $\chi^2 > 10.83$ indicates a significant difference between a random distribution and the observed distribution at a significance level $P < 0.001$. The map positions of zebrafish genes and expressed sequence tags summarized in the graphs were derived from this study and from previous work (Amores et al. 1998; Postlethwait et al. 1998; Gates et al. 1999; Kelly et al. 2000).

can be used for a variety of mapping projects (Shimoda et al. 1999). In particular, mutations are typically mapped with respect to SSLPs, and SSLPs also form the framework of both available RH maps (Geisler et al. 1999; Hukriede et al. 1999). SSLPs and other markers shared between different maps can be used to identify corresponding regions in different maps and, therefore, to compare the positions of markers that have been localized in one but not both panels. Currently the average interval between common markers in the HS and T51 RH (Geisler et al. 1999) maps is <10 cM, so that one can determine if two markers mapped in the different panels occupy the same region of a linkage group.

Conserved Synteny

By examining the map locations of zebrafish genes and their human counterparts, previous comparative analyses identified 28 groups of two or more genes that were syntenic in both species (Amores et al. 1998; Postlethwait et al. 1998; Gates et al. 1999). We have extended this comparison by analyzing the genes and ESTs mapped in this paper and in previous work. With the increase in the number of mapped, putatively orthologous gene pairs to 804, the comparative analysis identified 139 new conserved syntenies that involved two or more orthologous gene pairs, which raises the total to 167. Identification of these conserved syntenies expands comparative approaches to a large part of the zebrafish genome. This will increase the likelihood that comparative analysis can suggest candidate genes for zebrafish mutations, an approach that has already proved useful in the molecular identification of the *you too*, *snailhouse*, and *sucker* mutations (Karlstrom et al. 1999; Miller et al. 2000; Schmid et al. 2000). Our analysis also identified 136 putatively orthologous gene pairs that were not members of conserved syntenies. A few of these may reflect incorrect ortholog assignments and map errors, but most currently unpaired
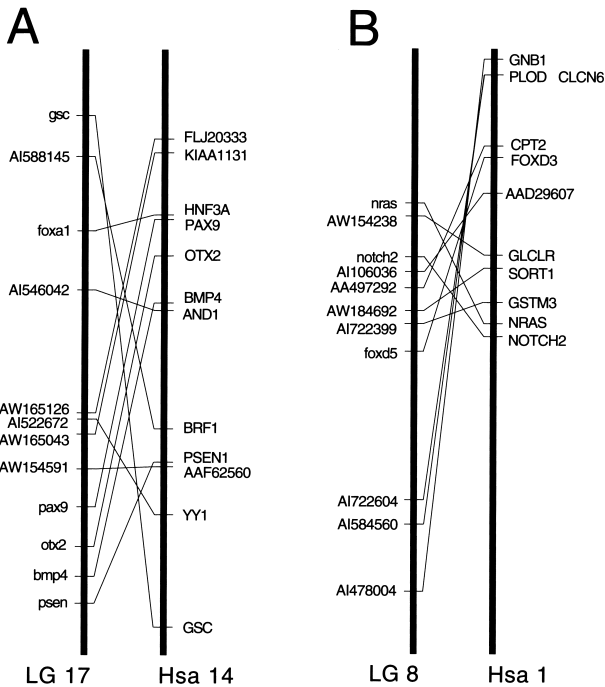
mapped polymorphisms to >3500 so that the average interval between markers is ~0.9 cM or ~500 kb (using 3000 cM as the length of the female meiotic map and assuming 1.7×10^9 bp in the haploid genome). The polymorphisms we have defined will be useful in testing candidate genes for mutations and in identifying polymorphic markers near mutations for positional cloning projects.

Zebrafish mapping efforts have employed a variety of meiotic mapping crosses and two RH panels (for review, see Talbot and Hopkins 2000). To facilitate comparisons among these maps, we have scored 800 previously mapped markers in the HS panel. Most of these (616) are SSLPs, which are robust markers that

Figure 3 Maps showing gene order with conserved syntenic groups. (*A*) Maps showing 12 genes and expressed sequence tags (ESTs) on zebrafish linkage group LG 17 and the positions of their human counterparts on chromosome (Hsa) 14. (*B*) Maps showing 11 genes and ESTs on zebrafish LG 8 and the positions of their human counterparts on the short arm of Hsa 1. In both panels, the entire chromosome (thick line) is shown, but the scale is different in *A* and *B*. Names of zebrafish genes are shown; ESTs are listed by GenBank accession number.

markers will probably join conserved syntenies as more loci are added to the comparative map. Thus a minimum estimate of the complete set of conserved syntenies is ~300 and the large number of presently unpaired markers suggests that the true number may be significantly higher.

Comparative analysis identified nine conserved syntenies containing ≥10 genes spread over the length of the chromosome (Figs. 2, 3), which indicates that groups of genes about the size of extant human chromosomes were in place in the last common ancestor of zebrafish and humans >450 million years ago. Zebrafish also has many chromosome segments that are orthologous to smaller portions of human chromosomes, which shows that, despite some very large conserved regions, translocations have also disrupted syntenies in the two lineages. Analysis of conserved syntenies also shows that there is not a one-to-one correspondence between zebrafish linkage groups and mammalian chromosomes. Some mammalian chromosomes share syntenies with more than one zebrafish linkage group and, as we consider in detail in the accompanying paper (Postlethwait et al. 2000), several zebrafish linkage groups share conserved syntenies

with more than one human chromosome. Although our results show some large regions of conserved syntenies, the orders of loci within the chromosome segments are often quite rearranged (Fig. 3). This suggests that chromosomal inversions have been fixed in fish and human lineages more often than translocations. As more genes are added to the comparative map, preservation of gene order in shorter segments may become apparent. Accordingly, genomic DNA sequencing in the pufferfish *Fugu rubripes* has shown that order has often been preserved in small groups of contiguous genes that span a megabase or so in human (Elgar et al. 1999).

Teleost Genome Duplication

Many studies have shown that gene families in zebrafish tend to have expanded membership as compared with mammals (Force et al. 1999; Postlethwait et al. 1999). Other teleosts also appear to have expanded gene families (Wittbrodt et al. 1998; Meyer and Schartl 1999) and in a few cases it is clear that medaka and pufferfish have orthologs of individual members of ze-
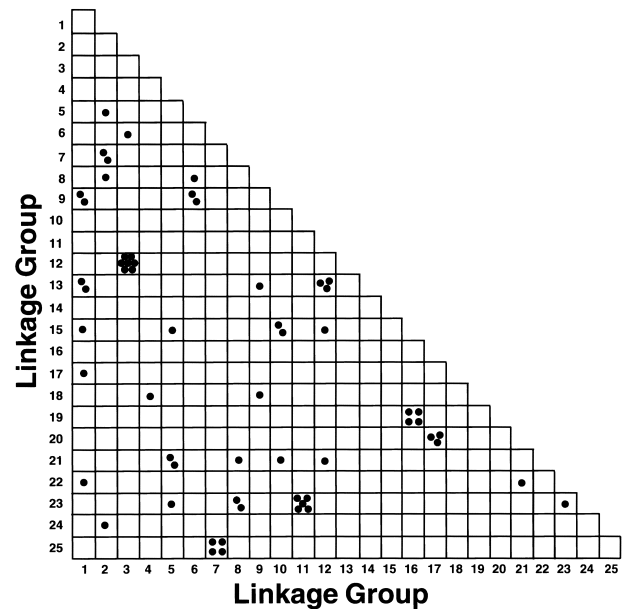

Figure 4 Positions of duplicate chromosomal segments in the zebrafish genome. Each dot in the grid represents a pair of putative duplicate zebrafish genes plotted according to the map positions of the two genes. Boxes that contain multiple points represent putative duplicate chromosomal segments. Names, accession numbers, and positions of putative duplicate genes mapped in the heat shock (HS) panel and in previous work are summarized in Table 3. The distribution of points in the grid is distinctly nonrandom. The $\chi^2 = 34.6$, calculated according to Gates et al. (1999). For 1 degree of freedom, a $\chi^2 > 10.83$ indicates a significant difference between a random distribution and the observed distribution at a significance level $P < 0.001$. The map positions of zebrafish genes and ESTs summarized in the graph were derived from this study and from previous work (Amores et al. 1998; Postlethwait et al. 1998; Gates et al. 1999; Geisler et al. 1999; Kelly et al. 2000).

brafish duplicate gene pairs (Naruse et al. 2000; Smith et al. 2000). Thus the current evidence supports the view that the duplicate genes arose early in the evolution of teleosts, >100 million years ago, before the divergence of lineages leading to medaka (*Oryzias latipes*), pufferfish, and zebrafish.

In principle, this gene family expansion could be caused by extra tandem duplication in the fish lineage, extensive loss of preexisting duplicates in the mammalian lineage, or extra duplication of chromosomal segments, chromosomes, or the entire genome in the fish lineage. Previous phylogenetic studies have argued against the idea that the expanded families result from retention in the fish lineage of a large number of duplicates that were present in the last common ancestor of zebrafish and human. Comparisons of *hox* clusters and other loci show that in a majority of cases both members of zebrafish duplicate gene pairs are equally related to their mammalian orthologs, which is consistent with an origin of these duplicates after the split of fish and mammalian ancestors (Amores et al. 1998; Gates et al. 1999; Meyer and Schartl 1999, and references therein). In some cases, however, zebrafish may have retained ancestral duplicates. In accord with previous mapping studies (Postlethwait et al. 1998; Gates et al. 1999), we find that apparent duplicates are not clustered together as the tandem duplication model would predict. Of 59 pairs of putative duplicates, one pair of duplicates was distantly located on the same linkage group (*nadl1.1* and *nadl1.2* on LG23) and there were 58 pairs in which the duplicates were located on different linkage groups. These results show that duplicate genes did not generally arise by tandem duplication. Instead, the results support the possibility that duplicate gene pairs arose by chromosome duplication (Amores et al. 1998; Postlethwait et al. 1998). The analysis identified 13 groups of 2–7 syntenic genes with duplicates that were also syntenic on a different chromosome. For example, *foxb1.1*, *hlx1*, *islet3*, and *pax6.2* are all located on LG 7, and their putative duplicates, *foxb1.2*, *hlx3*, *islet2*, and *pax6.1*, are all located on LG 25. As with comparisons of individual genes between zebrafish and mammals, the presence of duplicates of some chromosomal segments in zebrafish implies that there is not a single zebrafish counterpart for every group of syntenic mammalian genes. Thus, sequence comparisons and mapping together suggest that most zebrafish duplicate gene pairs arose from duplication of chromosomes or chromosomal segments in the fish lineage after the split of teleost and mammalian ancestors.

Analysis of the 59 putative duplicate gene pairs indicates that portions of 20 of the 25 linkage groups contain putative duplicate segments. Although our current sample size is relatively small, these results are consistent with the suggestion that these duplicate seg-

ments resulted from a duplication of the whole genome (Amores et al. 1998; Postlethwait et al. 1998; Meyer and Schartl 1999). An alternative to the genome duplication hypothesis is that chromosomal segments that represent a fraction of the genome were duplicated independently. Additional work, including sequence analysis of a large set of duplicate genes and mapping studies in other species, is needed to distinguish between these hypotheses. We favor, however, the genome duplication hypothesis because of the presumed deleterious effect of gene dosage imbalances caused by duplication of chromosomes or large chromosomal segments and also because the documentation of relatively recent genome duplications in salmonids and some cyprinids (Allendorf and Thorgaard 1984; Larhammar and Risinger 1994; Young et al. 1998) provides precedent for genome duplications in fish. It is important to note that duplicates have not yet been identified for most genes in our current dataset and that detailed comparative analysis of duplicate segments suggests that the fraction of genes with duplicates could be as low as 20% (Postlethwait et al. 2000). Thus the genome duplication hypothesis predicts that ≤80% of duplicate genes were lost after the duplication. In a number of cases the expression patterns of duplicate genes have diverged and it has been suggested that retained duplicates have persisted because they have acquired distinct functions either because of reciprocal deleterious deletion of essential gene subfunctions or the evolution of novel, beneficial, positively selected subfunctions (Force et al. 1999). Analysis of the expression patterns and functions of the genes in duplicate chromosomal segments identified here and in other studies will provide an important test of this hypothesis.

METHODS

Primer Design and Linkage Analysis

Sequences of *D. rerio* genes and ESTs (M. Clark and S. Johnson, Washington University Zebrafish Genome Resources Project; http://zfish.wustl.edu) were obtained from the NCBI nonredundant sequence database (nr), and primers were designed as described (Kelly et al. 2000). UniGene clusters containing mapped genes and ESTs were assigned from UniGene build 10 (http://www.ncbi.nlm.nih.gov/UniGene/Dr.Home.html).

Primer synthesis, polymorphism detection, and linkage analysis with MapManager software (Manly 1993) were performed as previously described (Kelly et al. 2000). Because double crossovers do not often occur in short intervals, many such double crossovers reflect incorrect genotype assignments. Markers with double crossovers in intervals of <20 cM were excluded from the dataset unless the double recombinant genotype was confirmed in a second assay. The final dataset contained 11 markers with double crossovers in intervals of <20 cM. Nine of these 11 markers are in one position adjacent to the centromere of LG 7. Because the centromere may block recombinational interference, these markers are likely to be genuine double crossovers. The complete geno-

type data set is available online (http://zebrafish.stanford.edu).

Sequence Comparisons

Zebrafish genes and ESTs were assigned putative human orthologs by BLASTX searches (Altschul et al. 1997) with the accession numbers of mapped zebrafish genes and ESTs against the NCBI human nonredundant protein sequence database (http://www.ncbi.nlm.nih.gov/blast/blast.cgi). For EST clones that have been sequenced on both ends, the sequences of both 5′ and 3′ ESTs were used for BLASTX searches. If the results of these searches had expect scores (E values) of ≤-5, the putative orthologs were further tested with reciprocal searches against the zebrafish subset of nonredundant sequences (nr) and dbEST databases. A human ortholog was confirmed if the original zebrafish gene or EST (or a gene or EST that showed highly significant overlap with the original sequence) was in the top five matches of the reciprocal search by TBLASTN. Map positions for these orthologs were found using the OMIM (http://www.ncbi.nlm.nih.gov/Omim), LocusLink (http://www.ncbi.nlm.nih.gov/LocusLink), and GeneMap'99 (http://www.ncbi.nlm.nih.gov/genemap99) databases. Mouse orthologs were identified using the Homolo-Gene database (http://www.ncbi.nlm.nih.gov/HomoloGene) and their map locations were found using Locuslink and the Mouse Genome Database (http://www.informatics.jax.org). Orthologs and their map positions are listed in Table 1 (available as supplementary material at http://www.genome.org).

ACKNOWLEDGMENTS

REFERENCES

Allendorf, F.W. and Thorgaard, G.H. 1984. Tetraploidy and the evolution of salmonid fishes. In *Evolutionary Genetics of Fishes*, (ed. BJ. Turner), pp. 1–46. New York, Plenum Press.

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.

Brownlie, A., Donovan, A., Pratt, S.J., Paw, B.H., Oates, A.C., Brugnara, C., Witkowska, H.E., Sassa, S., and Zon, L.I. 1998. Positional cloning of the zebrafish *sauternes* gene: A model for congenital sideroblastic anaemia. *Nat. Genet.* **20:** 244–250.

Childs, S., Weinstein, B.M., Mohideen, M.-A.P.K., Donohue, S., Bonkovsky, H., and Fishman, M.C. 2000. Zebrafish *dracula* encodes ferrochelatase and its mutation provides a model for erythropoietic protoporphyria. *Curr. Biol.* **10:** 1001–1004.

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

Dunham, I., Hunt, A.R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Smink, L.J., Ainscough, R., Almeida, J.P., Babbage, A., et al. 1999. The DNA sequence of human chromosome 22. *Nature* **402:** 489–495.

Edwards, J.H. 1991. The Oxford grid. *Ann. Hum. Genet.* **55:** 17–31.

Elgar, G., Clark, M.S., Meek, S., Smith, S., Warner, S., Edwards, Y.J., Bouchireb, N., Cottage, A., Yeo, G.S., Umrania, Y., et al. 1999. Generation and analysis of 25 Mb of genomic DNA from the pufferfish *Fugu rubripes* by sequence scanning. *Genome Res.* **9:** 960–971.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.-L., and Postlethwait, J.H. 1999. Preservation of duplicate genes by complementary, degenerate mutations. *Genetics* **151:** 1531–1545.

Gates, M.A., Kim, L., Egan, E.S., Cardozo, T., Sirotkin, H.I., Dougan, S.T., Laskari, D., Abagyan, R., Schier, A.F., and Talbot, W.S. 1999. A genetic linkage map for zebrafish: Comparative analysis of genes and expressed sequences. *Genome Res.* **9:** 334–347.

Geisler, R., Rauch, G.-J., Baier, H., van Bebber, F., Bross, L., Dekens, M.P.S., Finger, K., Fricke, C., Gates, M.A., Geiger, H., et al. 1999. A radiation hybrid map of the zebrafish genome. *Nat. Genet.* **23:** 86–89.

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

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.-S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.-K., et al. 2000. The DNA sequence of human chromosome 21. *Nature* **405:** 311–319.

Hukriede, N.A., Joly, L., Tsang, M., Miles, J., Tellis, P., Epstein, J.A., Barbazuk, W.B., Li, F.N., Paw, B., Postlethwait, J.H., et al. 1999. Radiation hybrid mapping of the zebrafish genome. *Proc. Natl. Acad. Sci.* **96:** 9745–9750.

Karlstrom, R.O., Talbot, W.S., and Schier, A.F. 1999. Comparative synteny cloning of zebrafish *you-too*: Mutations in the Hedgehog target *gli2* affect ventral forebrain patterning. *Genes & Dev.* **13:** 388–393.

Kelly, P.D., Chu, F., Woods, I.G., Ngo-Hazelett, P., Cardozo, T., Huang, H., Kimm, F., Liao, L., Yan, Y.-L., Zhou, Y., et al. 2000. Genetic linkage mapping of zebrafish genes and ESTs. *Genome Res.* **10:** 558–567.

Knapik, E.W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M.C., and Jacob, H.J. 1998. A microsatellite genetic linkage map for zebrafish. *Nat. Genet.* **18:** 338–343.

Kumar, S. and Hedges, S.B. 1998. A molecular timescale for vertebrate evolution. *Nature* **392:** 917–920.

Larhammar, D. and Risinger, C. 1994. Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpio*. *Mol. Phylogenet. Evol.* **3:** 59–68.

Manly, K.F. 1993. A Macintosh program for storage and analysis of experimental genetic mapping data. *Mamm. Genome* **4:** 303–313.

Melby, A. E., Warga, R.M., and Kimmel, C.B. 1996. Specification of cell fates at the dorsal margin of the zebrafish gastrula. *Development* **122:** 2225–2237.

Meyer, A. and Schartl, M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. in Cell Biol.* **11:** 699–704.

Miller, C.T., Schilling, T.F., Lee, K.-H., Parker, J., and Kimmel, C.B. 2000. *sucker* encodes a zebrafish Endothelin-1 required for

ventral pharyngeal arch development. *Development*
127: 3815–3828.

Moens, C.B., Yan, Y.-L., Appel, B., Force, A.G., and Kimmel, C.B.
1996. *valentino*: A zebrafish gene required for normal hindbrain
segmentation. *Development* **122:** 3981–3990.

Murphy, W.J., Sun, S., Yuhki, N., Hirschmann, D., Menotti-
Raymond, M., and O'Brien, S.J. 2000. A radiation hybrid map of
the cat genome: implications for comparative mapping. *Genome
Res.* **10:** 691–702.

Naruse, K., Fukamachi, S. Mitani, H., Kondo, M., Matsuoka, T.,
Kondo, S., Hanamura, N., Morita, Y., Hasegawa, K., Nishigaki, R.,
et al. 2000. A detailed linkage map of medaka, *Oryzias latipes*:
comparative genomics and genome evolution. *Genetics*
154: 1773– 1784.

Postlethwait, J.H., Amores, A., Force, A., and Yan, Y.-L. 1999. The
zebrafish genome. *Methods Cell. Biol.* **60:** 149–163.

Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.-L., Kelly,
P.D., Chu, F., Huang, H., Hill, A., and Talbot, W.S. 2000.
Zebrafish comparative genomics and the origins of vertebrate
chromosomes. *Genome Res.*, **10:** 1890–1902.

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.

Schmid, B., Furthauer, M., Connors, S.A., Trout, J., Thisse, B., Thisse,
C., and Mullins, M.C. 2000. Equivalent genetic roles for
bmp7/snailhouse and *bmp2b/swirl* in dorsoventral pattern
formation. *Development* **127:** 957–967.

Shimoda, N., Knapik, E.W., Ziniti, J., Sim, C., Yamada, E., Kaplan, S.,
Jackson, D., de Sauvage, F., Jacob, H., and Fishman, M.C. 1999.
Zebrafish genetic map with 2000 microsatellite markers.
Genomics **58:** 219–232.

Smith, S., Metcalfe, J.A., and Elgar, G. 2000. Identification and
analysis of two snail genes in the pufferfish (*Fugu rubripes)* and
mapping of human *SNA* to 20q. *Gene* **247:** 119–128.

Talbot, W.S. and Hopkins., N. 2000. Zebrafish mutations and
functional analysis of the vertebrate genome. *Genes & Dev.*
14: 755–762.

Wang, H., Long, Q., Marty, S.D., Sassa, S., and Lin, S. 1998. A
zebrafish model for hepatoerythropoietic porphyria. *Nat. Genet.*
20: 239–243.

Wittbrodt, J., Meyer, A., and Schartl, M. 1998. More genes in fish?
BioEssays **20:** 511–515.

Young, W.P., Wheeler, P.A., Coryell, V.H., Keim, P., and Thorgaard,
G.H. 1998. A detailed linkage map of rainbow trout produced
using doubled haploids. *Genetics* **148:** 839– 850.