Zebrafish Syntenic Relationship to Human/Mouse Genomes Revealed by Radiation Hybrid Mapping

Irene E. Samonte

Biology Department, and Center for Natural Sciences and Environmental Research (CENSER) College of Science, De La Salle University Manila, Philippines

Zebrafish (*Danio rerio*) is an excellent model system for vertebrate developmental analysis and a new model for human disorders. In this study, however, zebrafish was used to determine its syntenic relationship to human/mouse genomes using the zebrafish-hamster radiation hybrid panel. The focus was on genes residing on chromosomes 6 and 17 of human and mouse, respectively, and some other genes of either immunologic or evolutionary importance. Gene sequences of interest and zebrafish expressed sequence tags deposited in the GenBank were used in identifying zebrafish homologs. Polymerase Chain Reaction (PCR) amplification, cloning and subcloning, sequencing, and phylogenetic analysis were done to confirm the homology of the candidate genes in zebrafish. The promising markers were then tested in the 94 zebrafishhamster radiation hybrid panel cell lines and submitted for Logarithm of the Odds (LOD) score analysis to position genes on the zebrafish map. A total of 19 loci were successfully mapped to zebrafish linkage groups 1, 14, 15, 19, and 20. Four of these loci were positioned in linkage group 20, whereas, 3 more loci were added in linkage group 19, thus increasing to 34 loci the number of human genes syntenic to the group. With the sequencing of the zebrafish genome, about 20 more MHC genes were reported linked on the same group.

Key Words: *Danio rerio*, synteny, expressed sequence tags (ESTs), orthologous genes, radiation hybrid (RH) panel

INTRODUCTION

Danio rerio (Hamilton 1822), commonly known as zebrafish, belongs to the family Cyprinidae, order Cypriniformes. They inhabit streams, canals, ditches, ponds and pools, and in slow-moving to stagnant standing water bodies, particularly rice-fields, in India, Pakistan, Bangladesh, Sri Lanka, Thailand, Malay Peninsula, Sumatra, and China (Rahman 1989; Talwar and Jhingran 1991; Menon 1999). Appearance in Colombian waters and elsewhere are presumably due to escape from aquarium facilities. They feed mainly on worms and small crustaceans (Mills and Vevers 1989), and on insect larvae (Shrestha 1990). The five uniformly pigmented horizontal stripes on the side of the body - all extending onto the end of caudal fin rays makes it a popular aquarium species.

The zebrafish has been developed as a powerful model for genetic studies in developmental biology, which has led to the evolution of sophisticated cellular and molecular approaches. Transgenic zebrafish has been developed to explore the dynamic environment of the optically transparent developing zebrafish embryo (Udvadia and Linney 2003). More recently, the rapid progress

^{*}Corresponding author: irenesamonte@yahoo.com

of various zebrafish genomic infrastructure initiatives is facilitating the development of zebrafish models of human disease (Ward and Lieschke 2002). The success in making zebrafish models for human disorders is an indication that genes are shared among these species and these loci behave in similar ways. This prompted a further investigation of the syntenic relationship of zebrafish to human/mouse genomes. Synteny is the presence of genes in the same linkage group in different organisms, without consideration of the gene order (Eppig and Nadeau 1995). Syntenic genes can appear as uninterupted segments of 2 or more genes and/or expressed sequence tags (ESTs) with conserved gene order that likely represents homology segments conserved and intact between organisms, or as singletons of genes and ESTs that are isolated from members of their conserved synteny group.

Different mapping techniques can be utilized for syntenic studies. Michalova et al. (2000) utilized BAC (bacterial artificial chromosome) and PAC (P1 artificial chromosome) for chromosome walking that revealed persistence of syntenic blocks of major histocompatibility complex (MHC) loci from zebrafish to human/mouse. Other panels are also available ranging from haploid and diploid panels to radiation hybrid (RH) panels. In this study, somatic RH panel was used. It can position markers and estimate the physical distances of >1500 markers along a chromosome. In this method, extensive searches for polymorphisms in the genes are not required for it mainly utilizes the difference between the donor and the recipient cell lines (Kwok et al. 1998). The position of the markers is based on their retention in particular numbers of a panel of cell lines. Therefore, the farther apart the 2 markers are, the more likely that radiation will create breaks between them, thereby distributing them on 2 separate chromosomal fragments. The only disadvantage of the RH panel is its inability to nicely resolve the order of closely spaced markers as genetic maps do.

The loci of interest are those within, near, or related in function to the major histocompatibility complex (MHC). MHC derives its name from the observation that tissue transplants mismatched at its loci are rapidly destroyed by the recipient's immune system (Klein 1986). The genes comprising this complex are divided into two categories: the class I and the class II loci, which diverged evolutionarily from each other shortly after the emergence of their common ancestor more than 400 million years ago (Klein and O'hUigin 1993). In humans, the class I and class II loci are found together in a single chromosomal segment, about 3.6 million base pair (bp) in length (Campbell and Trowsdale 1993). A similar situation has been observed for all other mammals (Kasahara et al. 1995), in birds (Kaufman et al. 1999), and in amphibians (Flajnik et al. 1999). Furthermore, in all these taxa, the MHC segment also contains two other groups of loci. One group consists of loci that are structurally unrelated to but functionally involved with the class I loci. The other group is composed of loci that are structurally and functionally unrelated to both the class I and class II loci (Klein and Sato 1998). In contrast to this situation, in the zebrafish and other bony fishes (Sato et al. 2000), the class I loci are not linked to the class II loci (Bingulac-Popovic et al. 1997). Since bony fishes comprise at least ½ of living vertebrates (Eschmeyer 1998), the rearrangement of the MHC loci is indeed interesting and requires further study for better undertanding of its evolution.

This study was performed with the general aim of providing additional information about the organization of the zebrafish MHC, its linked genes (based on Hsa6 and Mmu17 composition), and genes with related function before the zebrafish has been sequenced. The specific objectives are (1) to identify zebrafish orthologues of genes found in human chromosome 6 (Hsa6)/mouse chromosome 17 (Mmu17), and some other genes with either immunologic or evolutionary significance, (2) to map these genes using the 94-cell line radiation hybrid (RH) panel, and (3) to find additional syntenic blocks between zebrafish and human genomes.

MATERIALS AND METHODS

Loci

Zebrafish loci with corresponding orthologues in the human and mouse genomes were investigated. These loci can be categorized as either human chromosome 6 (Hsa6) or non-Hsa6 loci, with 16 and 3 genes sampled, respectively. Of the 16 Hsa6 loci, four (DDR1, PPP1RD10, DOM3Z, and C4) are part of the major histocompatibility complex (MHC), whereas the remaining 12 are genes flanking this important gene cluster. A summary information about these different loci and their corresponding location in human and zebrafish genomes is given in Table 1.

Isolation of Zebrafish Markers and Primer Design

Candidate genes/markers for mapping in zebrafish using the zebrafish radiation hybrid panel were selected using 2 approaches. First, homologous genes on human chromosome 6 (Hsa6) and on mouse chromosome 17 (Mmu17), listed in the human-mouse homology map, which so far, have not yet been mapped in zebrafish, were determined. These sequences were used in Blastn searches for available sequences in other organisms, particularly that of zebrafish and other fishes. In the

	Location	
Loci	Human Chromosome	Zebrafish Linkage group
A. Human chromosome 6 (Hsa6)		
Complement component 4 (C4)	6p21.3	15
Discoidin domain receptor family, member 1 (DDR1)	6p21.3	16
DOM-3 (C. elegans) homolog Z (DOM3Z)	6p21.3	3
E2F transcription factor 3 (E2F3)	6p22	19
Fucosyltransferase 9 (FUT9)	6q16	19
H4 histone family, member H (H4HF)	6p21.3	25
Insulin growth factor 2 receptor (IGF2R)	6q25	20
Myristoylated alanine-rich protein kinase C substrate (MACS)	6q22	17
Plasminogen (PLG)	6q25	20
Protein phosphatase 1, regulatory subunit 10 (PPP1RD10)	6p21.3	19
Ring finger protein (C3HC4 type) 8 (RNF8)	6p21.3	20
REV3 polymerase (REV3L)	6q21	20
Serum glucocorticoid-regulated kinase (SGK)	6q23	23
SFRS protein kinase 1 (SRPK1)	6p21.2-p21.3	8
TATA-box binding protein (TBP)	6q27	13
Thrombospondin 2 (THBS2)	6q27	13
B. Non-Hsa6		
Alpha-2-macroglobulin (A2M/AMG)	12p13.3-p12.3	15
Complement component 3 (C3)	19p13.3-p13.2	1
Secreted protein rich in cysteine, Ostenectin (SPARC)	19p13.3-p13.2	14

Table 1. Human genes isolated in zebrafish and mapped using the zebrafish radiation hybrid panel

second approach, all zebrafish ESTs (\cong 80,000) were retrieved from the GenBank as of December 14, 2000 prior to the release of the partial zebrafish genome sequence in 2001. These ESTs were then used for Blastx searches against the GenBank/EMBL/DDBJ amino acid non-redundant database, to find a match with genes on Hsa6 as listed in the proteome analysis database at the European Bioinformatics Institute (EBI) (http://www.ebi. ac.uk/proteome/). Blastp and Blastn searches were also conducted with the non-zebrafish sequence identified as the highest match in the initial Blastx search. The orthology of each of the promising genes was ascertained by phylogenetic tree building.

To avoid mapping duplication, each of the promising genes was first checked in the zebrafish information network (http://zfin.org/). Nucleotide sequences were aligned and used for primer design. Degenerate primers were made from non-zebrafish sequences and specific primers from zebrafish EST sequences. The amino acid sequences of the homologous genes were aligned using CLUSTALX and CLUSTALW 1.81, and phylogenetic trees were constructed by the Neighbor-Joining method (Saitou and Nei 1987).

Molecular Methods

Genomic DNA was isolated using the QIA amp Tissue Kit (QIAGEN, Hilden, Germany). Contaminating RNA was removed by digestion with RNAse A (30 min at 37°C). PCR amplifications were carried out in 50 µl reaction mixtures using the PTC-200 Programmable Thermal Controller (MJ Research, Biozym, Hess. Oldendorf, Germany). The reaction mixture consisted of 1x reaction buffer containing 1.5 mM MgCl2, 400 mM of each of the 4 deoxynucleoside triphosphates (dNTPs), 0.5 mM of each of the sense and antisense primers, 2 units HotStar Taq polymerase (QIAGEN), and 50-100 ng of genomic DNA template. The PCR program comprised of an initial denaturation step for 15 min at 95°C, followed by 35 cycles, each cycle consisting of denaturation at 94°C for 30 s, annealing for 30 s at temperature specific for each primer pair (see Table 2), and extension at 72°C for 2 min; the reaction was completed by a final extension at 72°C for 10 min. The PCR products were separated by electrophoresis on 2% agarose gels. Purification of the PCR product was done using the QIAquick Gel Extraction Kit (QIAGEN) in preparation for cloning and subcloning. For cloning, the DNA fragments were ligated to the PCR

4-TOPO vector (Invitrogen Life Technologies, Leek, The Netherlands) as recommended by the supplier, and transformed into Escherichia coli TOP10 chemically competent cells (Invitrogen Life Technologies). Plasmid DNA was isolated using the QIAGEN Plasmid Mini Kit. Sequencing was carried out using ThermoSequenase Primer Cycle Sequencing Kit (Amersham Biosciences) with fluorescent primers annealing next to the multiple cloning site of the PCR 4-TOPO vector. The products were separated on a LI-COR 4200 DNA sequencer (MWG Biotech, Ebersberg, Germany). For the list of primers see Table 2.

Linkage Mapping

The linkage of the identified zebrafish markers was tested by PCR on zebrafish radiation hybrid panel composed of 94 cell lines and 2 controls (positive and negative controls) using primers specific to each marker. The recommended RH mapping protocol (http://wwwmap.tuebingen.mpg.de) on PCR, agarose gel electrophoresis, and data analysis was followed. All the markers reported here are also available on the same address. Data were submitted to the webpage http://wwwmap/, maintained by the lab of Robert Geisler in Division 3 - Genetics (Christiane Nüsslein-Volhard) of the Max-Planck-Institut für Entwicklungsbiologie,

Table 2. Specific primers for the different	zebrafish loci and their corresponding PCR product

Locus	Primer designation	Sequence (5' -> 3')	PCR product (bp)
C4	CMU1*	ACATGGAGATCATAATATACAAAG	226
	CML1*	TAGTGCATACTGACATAAAAAG	
DDR1	Ddr1U	CCTGCGCAAGCTTCATTTTCTGT	138
	Ddr1L	AGCGGTCCTTCCAGTCCATCCA	
DOM3Z	Dom3zU	CGACACCGCCCGCCGCCGCTGA	209
	Dom3zL	CTCGTGTATCCGTCTCTGAGGTT	
E2F3	E2fU	GACATCACAGACCACCAATACAG	332
	E2fL	GATCCAATACTCCATCAGAGGAC	
FUT9	Fut9U1	CGCACACATTTCTTCTACGAGTT	262
	Fut9L1	CATGGATGAAGGCGTCTCTTGGT	
	Fut9U2	CATTTCTTCTACGAGTTCGCACA	201
	Fut9L2	CATGGGCGTCTCTTGGTATGAAG	
H4HF	H2f2U	CCGTCGTCTTGCTGCCCGTGGTG	201
	H2f2L	CCGAAGCCGTACAGAGTGCGTCC	
IGF2R	Igfr2U	CAGAGTGGCTGATGGAAGAGATC	301
	Igfr2L	AGCGGTCCTTCCAGTCCATCCA	
MACS	MacsU	TTCGCCTCTCGCTTTCTCTTTTC	284
	MacsL	GCACACCCCCTCATAACATCACA	
PLG	PlgU	CAGTGCAGTGGAGAAAACTACAG	511
	PlgL	GGAATGTTTGTGAGGACTATGGG	
PPP1RD10	Ppp1r10U	CCAGTGGACCCCAGAGAGGTG	470
	Ppp1r10L	CTGAAGACTTTAGCCACTCCCTC	

Table 2 continuation . . .

RNF8	Rnf8U	GTTACATCACAGAGATGGCCTAC	209
	Rnf8L	AGGTCCGTATAGGATCAATTGC	
REV3L	Rev3lU	AAGATCCTCGCCAGCAGAAAAGT	222
	Rev3L	TCTTGAATTAGAAGAGATGGGGT	
SGK	SgkU	CGAGCGCAACGTGTTACTGAAGA	593
	SgkL	ATGCAGGTAACNCAAGCACTGG	
SRPK1	Sfrk1U	GGGGCTCAAACAAATAWTCACCC	268
	Sfrk1L	GCACAAGCACTTTACTGAAGAC	
TBP	TbpU	TGCACAAGCACTTTACTGAAGAC	406
	TbpL	GGGGCTCAAACAAATAATCACCA	
THBS2	Thbs2U	ATCTGCTTCCACATCACCACTTAA	427
	Thhbs2L	GATGATGACAACGATGGTGTTCC	
A2M/AMG	A2MA1s*	AAGGGCTCAGCCTGTGTC	243
	A2M2As*	TTTCATCTTCGACTTCACTG	
	AMG1*	CATCCTTCAATACCTGGAGAGCAG	924
	AMG2*	TGTGTTACCAGAAGGATCGTTTGAG	
C3	CT1*	GAAGCGCCGTGATGAAGCCATTG	169
	CT2*	CATCTGGTTTGCGGTAGGCCT	
	СТ3*	GGTCACCATGACTTTACCTTTCA	264
	CT4*	CGTAAGATCCGTCTTGTTTACGA	
SPARC	Sparc2U	AGGGTGAAGAAGATTTACGAGAA	404
	Sparc2L	AACTGCCAGTGAACGGGGAAGAT	

NOTE : *, primers previouly designed by Samonte et al. (2002); bp, base pair

Tuebingen, Germany for LOD score analyses that position genes on the zebrafish map using the SAMapper program (Stewart 1997).

RESULTS AND DISCUSSION

A total of 19 zebrafish loci was successfully isolated by PCR and subsequently sequenced. As shown in Table 2 and Figure 1, the PCR products from these loci range in size from 122 to 924 bp. To confirm that these genes truly represent the zebrafish orthologues, phylogenetic

analysis was performed on all the loci under study. A general trend was observed in all the loci - that is, zebrafish sequences clearly congregate with other metazoan proteins pertaining to each of the loci. Shown in Figure 2 are 3 of the representative phylogenetic trees (THBS, MACS, and DDR1) constructed. In Fig. 2A, zebrafish THBS2 grouped with other vertebrate THBS2 proteins rather than with the THBS1 and THBS3 gene groups (Figure 2, A). All metazoan MACS proteins, including that of the zebrafish MACS, clustered in 1 group completely separated from the MACS-related proteins (Figure 2, B). The orthology of the zebrafish DDR1 sequence to that of vertebrate DDR1



A. Human chromosome (Hsa) 6 loci

Figure 1. PCR of the different loci using zebrafish genomic DNA (zf:1, 2) and the specific primer pairs in Table 2. The molecular marker (MM) used in phi-X174 HaeIII cut. Control (C) was generated by substituting water instead of zebrafish DNA in the PCR reaction mix. The size of the product is indicated in base pairs (bp)





and not with receptor tyrosine kinase 4 or with the rest of the tyrosinase kinases is clearly shown in Figure 2,C.

All the 19 zebrafish markers are nicely assigned in different linkage groups in the RH map. To better visualize the linkage relationship of these markers, an integrated map of available Hsa6 orthologues prior to the release of the zebrafish genome sequence was prepared (Fig. 3). The orthologues are positioned relative to markers (in lowercase) previously mapped using different methods before the zebrafish has been sequenced [i.e. heat shock (HS), mother of pearl (MOP), chromosome walking using BAC (bacterial artificial chromosomes) and PAC (P1 artificial chromosome), etc.]. Three more markers, namely the 2 copies of FUT9 (zFut1 and zFut2), PPP1R10, and E2F3, mapped in linkage group (LG) 19. LG19 has so far 34 syntenic Hsa6 loci (Sültman et al. 2000; Michalova et al. 2000) and about 20 more MHC loci (not included in the Fig.3) have been added when the genome sequence was analyzed (Sambrook et al. 2005). LG20 had the second most number of Hsa6 orthologues of 13 including the



Figure 3. Zebrafish linkage groups to which human chromosome 6 (Hsa6) orthologues have been mapped. Markers mapped in the present study are underlined and italicized while the rest of the genes in uppercase are from Woods et al., 2000 and Michalova et al., 2000. Markers closely linked to the mapped genes are in small print. Non-Hsa6 loci are marked with asterisk (*). See Table 1 for the full names of the genes

newly described loci IGF2R, PLG, REV3L, and RNF8. The rest are scattered in different linkage groups, some of which as singletons (LGs 4 to 7, 17 and 25) waiting for other genes to be mapped.

The non-Hsa6 markers with either immunologic or evolutionary importance are found in linkage groups 1, 14, and 15. Interestingly, the 3 members of the alpha-2macroglobulin family of proteins (C3, C4, and A2M) are found in different LGs (Figure 3). C3 is found in LG1. The A2M and C4 loci mapped in the same region in LG 15 and are not linked to any of the class I or class II MHC loci.

Overall, the results obtained so far proved the presence of synteny between human/mouse and the zebrafish genomes. One clear evidence is the linkage of a number of Hsa6 orthologues, mostly members of the MHC, on chromosome 19 in zebrafish. Furthermore, the observed non-linkage on 1 zebrafish chromosome of members of the A2M family of genes can be interpreted as supporting the origin of the A2M family of genes by tandem duplications, followed by the dispersal of the copies to different chromosomes. It can also be argued that the association of C4 with the class I/II in tetrapods is accidental and without functional significance (Samonte et al. 2002).

ACKNOWLEDGMENTS

Special thanks to Dr. Werner E. Mayer and Prof. Dr. Jan Klein of the Max Planck Institute for Biology, Department of Immunogenetics, Max Planck Institute, Tuebingen, Germany for their valuable advice, help, and financial support to this undertaking.

REFERENCES

- BINGULAC-POPOVIC J, FIGUEROA F, SATO A, TALBOT S, JOHNSON SL, GATES M, POSTLETHWAIT JH and KLEIN J. 1997. Mapping of the Mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. Immunogenetics 46(2):129-134.
- CAMPBELL RD and TROWSDALE J. 1993. Map of the human MHC, Immunology Today 14:349-352.
- EPPIG JT and NADEAU JH. 1995. Comparative maps: the mammalian jigsaw puzzle. Current Opinions in Genetics and Development 59(6):709-716.
- ESCHMEYER WN. 1998. Catalog of Fishes. Part II. Genera of fishes. San Francisco: Special Publication, California Academy of Science. 3 vols. pp 1-2905.

- FLATJNIK MF, OHTA Y, NAMIKAWA-YAMADA C and NONAKA M. 1999. Insight into the primordial MHC from studies in ectothermic vertebrates. Immunological Reviews 167:59-67.
- KASAHARA M, FLATJNIK MF, ISHIBASHI T and NATORI T. 1995. Evolution of the major histocompatibility complex: a current overview. Transplant Immunology 3(1):1-20.
- KAUFMAN J, MILNE S, GOBEL TWF, WALKER BA, JACOB JP, AUFFRAY C, ZOOROB R and BECK S. 1999. The chicken B locus is a minimal essential major histocompatibility complex. Nature 401(6757: 923-925.
- KLEIN J. 1986. Natural History of the Major Histocompatibility Complex. New York: John Wiley & Sons. pp 1-755.
- KLEIN J and O'HUIGIN C. 1993. Composite origin of major histocompatibility complex genes. Current Opinion in Genetics and Development 3:923.
- KLEIN J and SATO A. 1998. Birth of the major histocompatibility complex. Scandinavian J Immunol 47:199-209.
- KWOK C, KORN RM, DAVIS ME, BURT DW, CRITCHER R, MCCARTHY L, PAW BH, ZON LI, GOODFELLOW PN and SCHMITT K. 1998. Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates. Nucleic Acids Research 26(15):3562-3566.
- MENON AGK. 1999. Check list fresh water fishes of India. p. 234-259. Rec. Zool. Surv. India, Misc. Publ., Occas. Pap. No. 175, 366 p.
- MICHALOVA V, MURRAY BW, SULTMAN H and KLEIN J. 2000. A Contig Map of the Mhc Class I Genomic Region in the Zebrafish Reveals Ancient Synteny. J Immunol 164:5296-5305.
- MILLS D and VEVERS G. 1989. The Tetra encyclopedia of freshwater tropical aquarium fishes. New Jersey: Tetra Press. 208 p.
- RAHMAN AKA. 1989. Freshwater fishes of Bangladesh. Zoological Society of Bangladesh. Dhaka, Bangladesh: Department of Zoology, University of Dhaka. 364 p.
- SAITOU N and NEI M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees, Molecular Biology and Evolution 4(4): 406-425.
- SAMBROOK JG, FIGUERROA F and BECK S. 2005. A genome-wide survey of major histocompatibility complex (MHC) genes and their paralogues in zebrafish. BMC Genomics 6: 152.

- SAMONTE IE, SATO A, MAYER WE, SHINTANI S and KLEIN J. 2002. Linkage Relationships of Genes Coding for Alpha-2-macroglobulin, C3, and C4 in the Zebrafish: Implications for the Evolution of the Complement and Mhc Systems. Scandinavian J Immunol 56(4):344-352.
- SATO A, FIGUEROA F, MURRAY BW, MALAGA-TRILLO E, ZALESKA-RUTCZYNSKA Z, SULTMAN H, TOYOSAWA S, WEDEKIND C, STECK N and KLEIN J. 2000. Non-linkage of the major histocompatibility complex of class I and class II loci in bony fishes. Immunogenetics 51(2):108-116.
- SHRESTHA TK. 1990. Resource ecology of the Himalayan waters. Kathmandu, Nepal: Curriculum Development Centre, Tribhuvan University. 645 p.
- SÜLTMAN H, SATO A, MURRY BW, TAKEZAKI N, GEISLER R, RAUCH GJ and KLEIN J. 2000. Conservation of major histocompatibility complex class III region synteny between zebrafish and human as determined by radiation hybrid mapping. J Immunol 165:6984-93.
- TALWAR PK and JHINGRAN AG 1991. Inland fishes of India and adjacent countries. Volume 1. A.A. Balkema, Rotterdam. p. 41.
- UDVADIA AJ and LINNEY E. 2003. Windows into development: historic, current, and future perspectives on transgenic zebrafish. Developmental Biol 256(1): 1-17.
- WARD AC and LIESCHKE GJ. 2002. The zebrafish as a model system for human disease. Frontier in Biosciences 1(7): 827-33.