Immunogenetics (2002) 54:367–380
DOI 10.1007/s00251-002-0478-0

ORIGINAL PAPER

Jennifer G. Sambrook · Roslin Russell
Yagnesh Umrania · Yvonne J. K. Edwards
R. Duncan Campbell · Greg Elgar · Melody S. Clark

Fugu orthologues of human major histocompatibility complex genes: a genome survey

Received: 10 April 2002 / Accepted: 23 May 2002 / Published online: 16 July 2002

Abstract The major histocompatibility complex (MHC) region in fish has been subjected to piecemeal analysis centering on the in-depth characterization of single genes. The emphasis has been on those genes proven to be involved in the immune response such as the class I and class II antigen presenting genes and the complement genes. The *Fugu* genome data presents the opportunity to examine the short-range linkage of potentially all the human MHC orthologues and examine conserved synteny with the human and, to a more limited extent, zebrafish genomes. Analysis confirms the existence of a limited MHC locus in *Fugu* comprising the MHC class Ia genes and associated class II region genes involved in class I antigen presentation. Identification of additional human MHC orthologues indicates the completely dispersed nature of this region in fish, with a maximum of six MHC genes maintained within close proximity in any one contig. The majority of the other genes are present in the genome data as either singletons or pairs. Comparison with zebrafish substantiates previously observed linkages between class III region orthologues and hints at an ancient conserved class III region.

Keywords MHC · *Fugu* · Evolution · Synteny · Fish

Introduction

The human major histocompatibility complex (MHC) region is one of the most gene dense regions of the human genome, comprising 220 genes in 4.6 Mb of DNA on chromosome (Chr) 6p21.3 (The MHC Sequencing Consortium 1999). It is a region intimately associated with immune function, with approximately 40% of the gene products having some immunological role and contains genes essential to both the adaptive and innate immune systems. This is a highly complex region of medical importance, and analysis using a whole range of techniques will be necessary to understand the dynamics of the genes within it. One of these techniques is comparative genomics, examining orthologous regions in a number of different organisms. Only by studying the evolutionary processes can it be determined whether gene clustering in such an "immune-dense" region has functional significance.

Traditionally, the human MHC has been sub-divided into three main regions, class II, class III and class I (centromeric to telomeric location, respectively) according to gene content, specifically the class I and class II antigen presenting molecules. The central class III region is somewhat different from its neighbors in that it contains a mix of genes that are structurally and functionally unrelated (Klein and Sato 1998), many of which have no involvement in immune function. However, it does contain a number of immune-related genes such as the tumour necrosis factor family members, the heat shock proteins, and the complement components associated with innate immunity. Recent analyses of surrounding gene content and isochore boundaries have further extended the MHC locus with the designation of an extended class II region (preceding the class II region) (Herberg et al. 1998; Stephens et al. 1999) and an extended class I region (telomeric to the class I region) (Totaro et al. 1996).

In addition to the complete sequence data for the human MHC region, gene maps and considerable tracts of sequence data exist for the MHC of other mammals (Beck et al. 2001; Renard et al. 2001; Walter and Gunther 1998). Within these species, the MHC constitutes a single chromosomal segment; the overall complement of genes, with the exception of those encoding the class I and class II antigen presenting molecules, is very similar although rearrangements do occur (Chardon et al. 1999; McShane et al. 2001). As the spectrum is broadened to include more diverse species such as birds and amphibia (Kaufman et al. 1999; Namikawa et al. 1995;

J.G. Sambrook · R. Russell · Y. Umrania · Y.J.K. Edwards
R.D. Campbell · G. Elgar · M.S. Clark (✉)
Fugu Genomics, HGMP Resource Centre,
Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SB, UK
e-mail: mclark@hgmp.mrc.ac.uk
Tel.: +44-1223-494562, Fax: +44-1223-494512

Nonaka et al. 1997a, b; Ohta et al. 1999), these still maintain the MHC as a single chromosomal block, but the order and content is more radically altered. The prime example is the minimal MHC of the chicken which comprises only 19 genes (Kaufman et al. 1999).

This situation changes when considering marine vertebrates: the teleost fish and elasmobranchs, which diverged from the main tetrapod lineage over 450 million years ago. Sharks, the oldest vertebrate ancestors possessing an adaptive immune system, maintain the linkage between the class I and class II genes (Ohta et al. 2000, 2002) indicating that the close linkage between these two classes of genes has been maintained for at least 460 million years in representatives of most vertebrate taxa. However, this linkage has been lost in the teleosts, with the orthologues to the human MHC genes dispersed across a number of different chromosomes, demonstrating a highly derived arrangement.

Within the teleosts, a core MHC region is still identifiable, comprising a mix of class Ia antigen presenting genes linked to genes found within the human class II region. Comprehensive studies of this region in zebrafish (Michalova et al. 2000; Murray et al. 1999; Sültmann et al. 2000a), *Fugu* (Clark et al. 2000, 2001) and most recently medaka (Matsuo et al. 2002) reveal a conserved core of genes comprising the class Ia antigens, genes traditionally found within the human classical class II region, but which are involved in class I antigen presentation (*PSMB8*, *PSMB9*, *PSMB9-L*, *PSMB10*, *ABCB3B*, *TAPBP*) and a variety of genes which currently have no known involvement in immunity [*RXRB*, *COL11A2*, *FABGL* (*KE6*), *BRD2* (*RING3*), *DAXX*, *ZNF297* (*BING1*), *KNSL2* and *FLOT1*]. The order of these is highly conserved between the three fish species, which represent widely separated orders of teleosts, indicating the potential teleost-wide nature of this arrangement.

The class II antigen genes are not linked to this core piscine MHC region (Sato et al. 2000a), with evidence from zebrafish suggesting that they are split between at least three loci and at least two different chromosomes (Bingulac-Popovic et al. 1997; Graser et al. 1998). Research in fish has centered on the class I and II genes due to their immediate association with disease resistance and susceptibility. The class III region genes, in comparison, have been relatively neglected. A survey of class III region genes in zebrafish has indicated that these genes are also dispersed, but that there are two core regions (*KE4*, *SACM2L*, *RPS18*) and (*PPT2*, *DSP1*, *BAT2*, *CSNK2B*, *SK12 W*, *GTF2H4*, *G9A*, *PBX2*, *FLOT1*) which are linked to the piscine core MHC (Sültmann et al. 2000b).

Fugu has been promoted as a model vertebrate with a minimal genome of only 400 Mb, one-eighth the size of human. It has recently been the subject of a whole shotgun genome sequencing program (The Fugu Genome Consortium 2002) and draft sequence data is publicly available. This is in the form of unordered contigs (scaffolds) ranging in size from a few kb to 657 kb. The

Table 1 List of genes from the human MHC class I region used in this study with details of the *Fugu* scaffolds identified

Gene name	Alternate names	SPTR	Scaffold	Scaffold size (bp)	Number of genes on scaffold	Human map data of genes on *Fugu* scaffolds
Telomere						
HLA		Q9GJE4	S002818	34344	3	1p35.2
HLA		Q9GJE4	S005787	11732	3	6p21.3
HCGIX-4		Q93065	NF			
RAN	*TC4*, *RANBP1*	P43487	S001096	77603	7	9p13.3; 11q; 12q24.11; 22
GNL1	*HSR1*	P36915	S000627	109693	8	2q37.1; 3q26.33; 12p13
ABCF1	*ABC50*, *TSAP*	O14897	S003320	27893	2	6p21.3
PPP1R10	*FB19*	O00405	S005653	11725	2	6p21.3
DDX16	*DPB2*, *KIAA0577*	O60231	S000639	108398	4	1q23.3; 6p21.3; 19q13.3; 21q22.13
KIAA0170		Q14676	NF			
TUBB		Q13885	S000755	99369	5	6p21.3
FLOT1		FSS	S001569	58816	6	6p21.3; 17
FLOT1 (dup)		FSS	S000755	99369	5	6p21.3
IER3	*IEX-1*, *PRG1*	P46695	NF			
DDR1	*CAK*, *RTK6*, *TRKE*	Q08345	S000755	99369	5	6p21.3
GTF2H4	*P52*, *TFIIH*	Q92759	S000146	198132	17	1; 5p15.3, 6p; 7p21.1, 10p13; 19q13.11
C6orf15	*STG*	Q9Y6W6	S000200	176920	12	1; 2p23.2; 6p21.1; 14q22.1
CDSN	Corneodesmosin, S	O43509	NF			
C6orf16	*SEEK1*	Q9Y6W5	NF			
C6orf17	*SPR1*	Q9Y6W4	NF			
C6orf18	*HCR*	Q9Y6W2	NF			
TCF19	*SC1*	Q13176	S009360	4287	1	none: singleton
POU5F1	*OTF3*	P31359	NF			
USP8P	*KIAA0055*-homologue	P40818	S002593	35969	2	18q
DHFRP2	*DHFRP*	Q9UMI0	NF			
MICA	*PERB11*.1	Q96QC4	NF			
MICB	*PERB11*.2	Q96QC5	NF			
Class III						

Table 2 List of genes from the human MHC class III region used in this study with details of the *Fugu* scaffolds identified

Gene name	Alternate names	SPTR	Scaffold	Scaffold size (bp)	Number of genes on scaffold	Human map data of genes on *Fugu* scaffolds
Class I						
BAT1	*D6S81E*	Q13838	S002455	38865	3	17q24
ATP6G2	*NG38*	O75348	NF			
NFKBIL1	*IKBL*	Q14625	S000499	121955	12	2p; 4p; 6p21.3; 8p21; 11q12; 12q; 17p13; 22q12
LTA	*LT*, *TNFB*, *TNFSF1*	P01374	NF			
TNF	cachectin, *DIF*, *TNFA*, *TNFSF2*	P01375	S000795	96159	5	6p21.3; 6p25.3; 8q21; 10p11
LTB	*p33*, *TNFC*, *TNFSF3*	Q06643	NF			
LST1	*B144*, *D6S49E*	Q13669	NF			
1C7	*D6S2570E*, *NKP30*	O14930/1/2	NF			
AIF1	*G1*, *IBA1*, *IRT-1*	P55008	S000061	265888	12	1p33; 2q32; 6p21.3; 7p
BAT2	*G2*, *D6S51*, *D6S51E*	P48634	S003195	28805	2	6p21.3
BAT3	*G3*, *D6S52E*	P46379	S006155	10373	2	7p21.1
G3A	*APOM*, *NG20*	O95445	NF			
G4	*D6S53E*, *NG34*	O95873	NF			
BAT4	*G5*, *D6S54E*	O95872	S000150	194105	8	4q22; 7; 16p13; 17q
CSNK2B	*CSK2B*, *G5A*, phosvitin	P13862	S000061	265888	12	1p33; 2q32; 6p21.3; 7p
CSNK2B (dup)			S003195	28805	2	6p21.3
G5B	*C6orf19*	Q9UKT0	NF			
G5C	*C6orf20*, *NG33*	O95871	NF			
BAT5	*D6S82E*, *NG26*	O95870	S000755	99366	5	6p21.3
G6F	*C6orf21*, *NG32*	O95869	NF			
G6E	*C6orf22*	Q9UMP8	NF			
G6D	*C6orf23*, *NG25*	O95868	NF			
G6C	*C6orf24*, *NG24*	O95867	NF			
G6B	*C6orf25*, *NG31*	O95866	NF			
DDAH2	*DDAH*, *G6a*, *NG30*	O95865	S001433	62791	7	6p21.3; 7p14; 9q34.1; 12q13; 19q13.4
CLIC1	*G6*, *NCC27*	O00299	S001052	78951	3	6p; 21q22.1
CLIC1 (dup)			S008599	5024	1	none: singleton
MSH5	*G7*, *MutS*	O43196	S001065	77894	7	2q12; 7q21.1; 10p14; 12p13
NG23	*C6orf26*	Q9Y335	NF			
G7C	*C6orf27*, *NG37*	Q96QC8	S000537	118515	7	4q; 6p21; 11q; 17q25.3
G7C (dup)			S000951	84286	8	12q24; 17
G7C (dup)			S001291	69002	8	10q22.1; 17q
VARS2	*G7A*, *Val-TRS*	P26640	S000553	118466	13	1; 4q21.2; 5q35; 6p21.3; 6q12; 18q21
VARS2 (dup)			S000755	99366	5	6p21.3
LSM2	*G7B*, snRNP	Q9Y333	S001541	60480	4	2q; 3q27
HSPA1L	hum70 t, *HSP70-HOM*	P34931	S001038	82082	6	6p21; 1p
HSPA1A/	*HSP70–1*/*HSP70–2*	P08107	S002561	35253	6	16p; 17q
HSPA1B			S006421	10015	3	16p13.3; 21q21.1
			S007341	8129	2	6p21.3
			S011760	2003	1	none: singleton
G8	*D6S57*	Q9UBA6	NF			
NEU1	*G9*, *NEU*, *SIAL1*	Q99519	S005829	13098	1	none: singleton
NG22	*CTL4*, *C6orf29*, *FLJ14491*	Q9Y332	S005753	14665	2	8q23.3
NG36/G9a	*BAT8*	Q96KQ7	S002594	35358	5	6p21.3
G10	*D6S59E*, *NG35*	Q9Y330	S002594	35358	5	6p21.3
C2	*CO2*	P06681	NF			
BF	*CFAB*, *GBG*, *PBF2*	P00751	S000298	147440	21	1q23.1; 5q31; 6p21.3; 11q; 17q; 19q13; 21q22.11
BF (dup)			S004468	17379	4	7q35; 11q23
RDBP	*D6S45*, *NELF-E*, *RD*, *RDP*	P18615	S002594	35358	5	6p21.3
SKIV2L	*DDX13*, *G11A*, *SKI2*, *SKIV2*, *SKI2 W*	Q15477	S002594	35358	5	6p21.3
DOM3Z	*DOM3L*, *NG6*	Q9NPK4	S000468	123413	10	1q24.1; 6p21.3; 7p22; 16p; 22q13.2
STK19	*D6S60*, *G11*, *HLA-RP1*, *RD*	P49842	S000468	123413	10	1q24.1; 6p21.3; 7p22; 16p; 22q13.2
C4B	*C4F*, *CO4*	P01028	S000061	265888	12	1p33; 2q32; 6p21.3; 7p
CYP21A2	*CPS1*, *CYP21*, *CYP21B*, *P450-C21B*	Q16749	S000061	265888	12	1p33; 2q32; 6p21.3; 7p
TNXB	*HXBL*, *TENX*, *TNX*, *XB*, *XBS*	P22105	S000061	265888	12	1p33; 2q32; 6p21.3; 7p

Table 2 Continued

Gene name	Alternate names	SPTR	Scaffold	Scaffold size (bp)	Number of genes on scaffold	Human map data of genes on *Fugu* scaffolds
CREBL1	*CREB-RP*, *G13*	Q13269	S000146	198071	17	1; 5p15.3; 6p21.3; 7p21; 10p13; 19
FKBPL	*DIR1*, *NG7*	Q9UIM3	S000128	206279	11	1; 6p21; 11p15.5; Xp11
C6orf31	*NG5*	Q99946	S002594	35358	5	6p21.3
PPT2	*G14*	Q9UMR5	S001433	62791	7	6p21.3; 7p14; 9q34.1; 12q13; 19q13.4
PPT2 (dup)			S007225	7537	1	none: singleton
NG3	*C6orf8*	Q99944	NF			
AGPAT1	*G15*, *LPAAT*, *LPAATA*	Q99943	NF			
RNF5	*G16*, *NG2*, *RING5*	Q9UMQ2	S000021	345163	26	2q; 3q; 6q16; 7q; 12q; 17p13.3; 19q13.3; 22q
AGER	*RAGE*	Q15109	NF			
PBX2	*G17*, *HOX12*	P40425	S007803	3870	1	none: singleton
C6orf9	*G18*, *NG1*	P78548	NF			
NOTCH4		O00306	NF			
Class II						

Table 3 List of genes from the human MHC class II region used in this study with details of the *Fugu* scaffolds identified

Gene name	Alternate names	SPTR	Scaffold	Scaffold size (bp)	Number of genes on scaffold	Human map data of genes on *Fugu* scaffolds
Class III						
C6orf10	*TSBP*	Q60665	NF			
BTNL2	*BTL-II*	O70355	NF			
*HLA-D**		FSS	S011611	2098	1	none: singleton
ABCB3	*TAP2*	Q9DER8	S003776	22311	3	6p31.2
PSMB8	*LMP7*	Q9DES2	S007335	7216	2	6p21.3; 16q22.1
ABCB2	*TAP1*	Q9PT31	S005235	13931	1	none: singleton
PSMB9	*LMP2*	Q9DER9	S003776	22311	3	6p31.2
PPP1R2P	*IPP2*	P41236	NF			
BRD2	*RING3*	FSS	S003776	22311	3	6p31.2
COL11A2		FSS	S001220	72019	3	6p21.3
RXRB		FSS	S001220	72019	3	6p21.3
HKE4	*RNF5*, *RING5*	Q99942	NF			
FABGL	*KE6*, *RING2*	Q92506	S001220	72019	3	6p21.3
RING1		Q06587	NF			
SACM2L	*ARE1*	O43764	S000146	198132	17	1; 5p13.3, 6p; 7p21.1, 10p13; 19q13.11
RPS18	*KE3*	P25232	S000146	198132	17	1; 5p15.3, 6p; 7p21.1, 10p13; 19q13.12
B3GALT4		O96024	S000284	154134	6	1p32.1; 3; 6p21.3
C6orf11	*BING4*	O15213	S000284	154134	6	1p32.1; 3; 6p21.3
HKE2		O15212	S000817	95047	7	13q14.3; X
RAB2L	*RGL2*, *KE1*.5	Q92942	NF			
TAPBP		FSS	S005787	11732	3	6p21.3
ZNF297	*BING1*	FSS	S001569	58816	6	6p21.3; 17
DAXX		FSS	S005787	11732	3	6p21.3
LYPLA2L	*APT*	O95372	S004391	17820	2	1p36.11
KNSL2		FSS	S001569	58816	6	6p21.3; 17
Centromere						

availability of this data has presented a major opportunity to study the short-range linkage relationships of genes orthologous to the human MHC. The aim of this study was to discover if there were any ancient conserved linkage groups, by comparison with the data from zebrafish, or whether, with the exception of the piscine class I region, there are no constraints on gene rearrangement for this particular gene set within the teleosts.

Materials and methods

The *Fugu* draft sequence data is currently available on three web sites at the Joint Genome Institute in the USA (http://www.jgi.gov/fugu/index.htm), the UK HGMP-RC (http://fugu.hgmp.mrc.ac.uk) and an ensemble pipeline in Singapore (http://www.fugu-sg.org/). The analysis was carried out on the October 2001 freeze data set.

Analysis and annotation of the 4.6 Mb human MHC sequence (The MHC Sequencing Consortium 1999) was performed using the comprehensive suite of programs in NIX (http://www.hgmp.mrc.ac.uk/NIX/) available at the UK HGMP-RC. An MHC database of genomic sequence fragments and human gene annotations was created using data from SwissProt (Bairoch and Apweiler 2000), various public NCBI databases (UniGene, LocusLink and

Refseq), NIX and the Sanger Centre's chromosome 6 website (R. Horton, personal communication). The database analysis resulted in 219 known mRNA transcripts from MHC classes I, II, III, XI (extended class I) and XII (extended class II) being identified. The extended class I sequences were not used in the analysis. These sequences were then filtered for any repetitive elements using RepeatMasker (Smit, AFA and Green, P http://ftp.genome.washington.edu/RM/RepeatMasker.html). BLASTn (version 2.0.12; Altschul et al. 1997) was then used to query each MHC transcript against all the *Fugu* scaffolds. A number of genes were not identified using tBLASTn and these were re-screened for using tBLASTx (see database accession numbers in Tables 1, 2 and 3). In all cases, identified scaffolds were transferred to the NIX environment to facilitate gene identification and determine gene order.

A number of potentially duplicated genes were identified. The nucleotide sequences for these were isolated from the scaffolds and the amino acid sequence was extracted using the EMBOSS suite of programs (Rice et al. 2000). The predicted protein sequences were then aligned with other vertebrate gene family members present in the SPTR database (Bairoch and Apweiler 2000), identified via SRS (http://srs.hgmp.mrc.ac.uk/srs6/) using Clustal X (Higgins et al. 1992). Phylogenetic analysis was carried out from the Clustal X multiple alignment output using the neighbor joining method (Saitou and Nei 1987) via the PHYLO_WIN interface v1.2 (Galtier et al. 1996). Multiple alignment parameters for PHYLO_WIN were gap opening 15; gap extension 0.05; delay divergent sequences 40%; DNA transition weight 0.50, with 1000 bootstrap replicates.

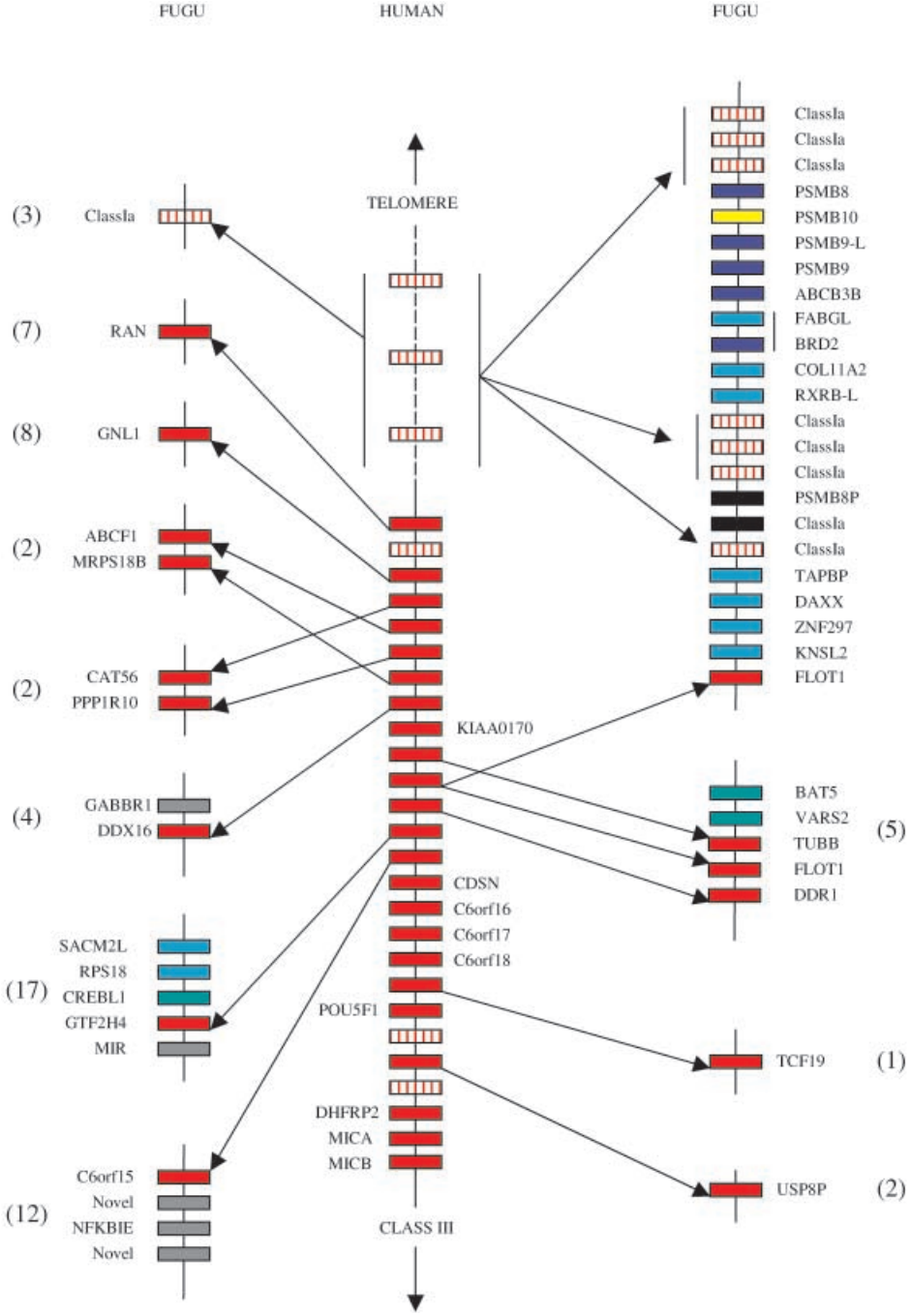

Fig. 1 Diagrammatic comparison of the human MHC class I region and the different identified *Fugu* scaffolds. The order of the scaffolds is not known. Genes not present on 6p are not shown. The total number of genes on each scaffold is shown by the *number in brackets* at the side of each contig. The main *Fugu* class I region contig does not have such a figure, as it is a representation of previous data and is currently split between several scaffolds. Color code: *red* class I region genes, *red striped* MHC class Ia genes, *dark blue*, human classical MHC class II region genes, *light blue* human extended MHC class II region genes, *green* human MHC class III region genes, *gray* genes found on human chromosome (Chr) 6, but not within the MHC region, *black* known *Fugu* pseudogenes, *yellow* gene involved in class I antigen presentation, but not mapping with the human MHC region

Results

Mapping of human MHC class I region orthologues

Of the 24 genes identified via SPTR within the class I region in human, 11 were not found (Table 1). Analysis of the scaffolds found to contain the remaining 13 genes and additionally the class Ia genes, confirmed previous findings (Clark et al. 2001) of one main class I region in *Fugu* (Fig. 1). This region contained a mixture of the class Ia genes with genes involved in class I antigen presentation, found within the human classical class II region, and four genes from the human extended class II region. This 300 kb region has been extensively described elsewhere (Clark et al. 2000) and was represented across seven different scaffolds. However, even with this limited data, the scaffolds did refine the mapping within this region. In the previous analysis, a single class Ia gene was found between *Col11A2* and *RXR*, based on the analysis of one BAC. Data from the scaffolds indicated that *Col11A2* and *RXR* were in fact neighbors and that this BAC was probably a co-ligation. Interestingly, the scaffolds may have identified a further class Ia locus. *S002818* shows a region of BLAST homology with the cyprinid nonclassical class I Z lineage sequences (*Q9MX31* and *Q9MX30*) and also the marbled lungfish class Ia sequence (*Q9TPO4*) (Sato et al. 2000b), covering the α1, α2 and α3 domains. Percentage identity of the *Fugu* sequence approximated to 50% for both of the cyprinid sequences and varied from 29% for the α2 and α3 domains to 42% for the α1 domain when compared to the lungfish sequence. Additional evidence for this sequence being a class I related molecule comes from the conservation of the invariant cysteine residues in the α2 and α3 domains and an N-glycosylation site at the end of the α1 domain. The latter is highly conserved in virtually all vertebrate class I sequences. However, no further sequence similarity was detected outside of these domains, indicating that if this is indeed a new class Ia locus, it is highly derived.

Of the remaining class I region genes identified, most were present either as singletons or pairs with other human MHC genes (Fig. 1). This was true even if the genes were present on large contigs containing in excess of six different genes. Most significant of the short range linkages were scaffolds *S000146* and *S000755*. The former contained two class II region genes (*SACM2L* and *RPS18*) with a class I region sequence, *GTF2H4* and a further gene, *MIR*, which maps to human Chr 6p. The latter contig contained two class III region genes (*BAT5* and *VARS2*) with three class I region genes (*TUBB*, *FLOT1* and *DDR1*). *FLOT1* or flotillin was also present on the major 300 kb class I region contig. Phylogenetic analysis shows that this is a duplicated gene and not two different genes, *FLOT1* and *FLOT2* (Fig. 2). The situation with *TUBB* is more complex. A further *TUBB* gene was identified on scaffold *S000971* linked to a human 6p gene (desmoplakin). Therefore they were both candidates for the MHC *TUBB* orthologue and also potentially duplicated genes. Both of the genes were very similar, showing 96.2% and 94.7% amino acid identity (for scaffolds *S000755* and *S000971* respectively) to the human MHC TUBB gene. However, phylogenetic analysis (results not shown) gave inconclusive results.

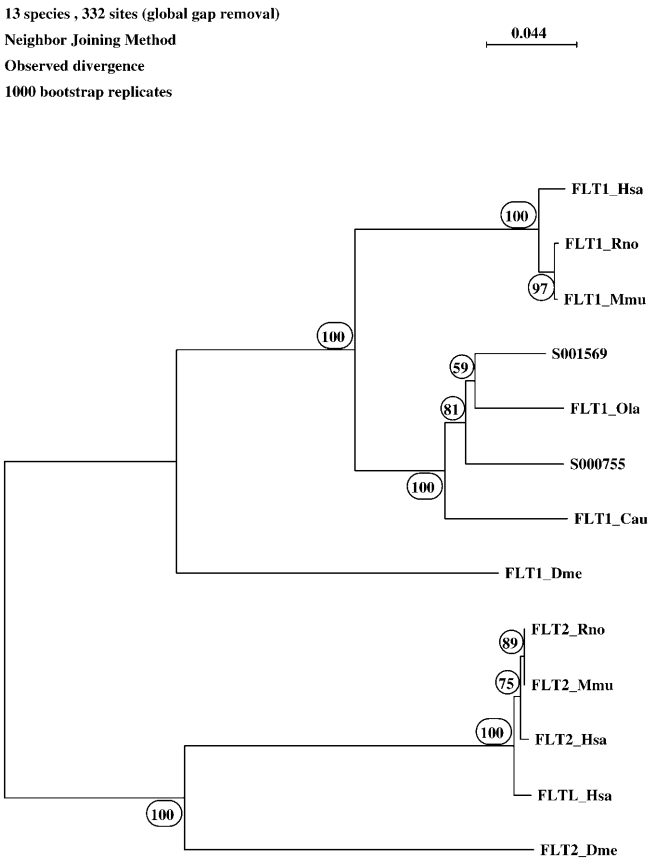

Fig. 2 Phylogenetic analysis of the two *Fugu* flotillin genes (*S001569* and *S000755*). Abbreviations and accession numbers: *FLT1* flotillin 1, *Hsa* human (*Homo sapien*) O75955, *Rno* rat *(Rattus novegicus*) Q9ZIE1, *Mmu* mouse (*Mus musculus*) O08917, *Ola* medaka (*Oryzias latipes*) BAB83856, *Cau* goldfish (*Carassius auratus*) O13127, *Dme* fruit fly (*Drosophila melanogaster*) O61491, *FLT2* flotillin 2; Hsa: Q14254, Rno: Q9Z2S9, Mmu: Q60634, Dme: O61492, FLTL similar to flotillin 2 Hsa: Q9BTI6

Mapping of human MHC class III region orthologues

Of the 58 genes identified in the human MHC class III region, 22 were not found in *Fugu* using BLAST sequence similarity searching (Table 2). Most are present as either singletons or in groups of two or more MHC-related genes (Fig. 3) across 33 different scaffolds. This figure takes into account duplicated MHC class III region genes in *Fugu*.

A total of 24 *Fugu* genes were found to be linked with other genes located in the human MHC. The largest contig comprised scaffold *S000061* with a cluster of five class III region genes (*CSNK2B*, *AIF1*, *CYP21A2*, *TNXB*, *C4B*) and an extended class II region gene (*RXRB*: or

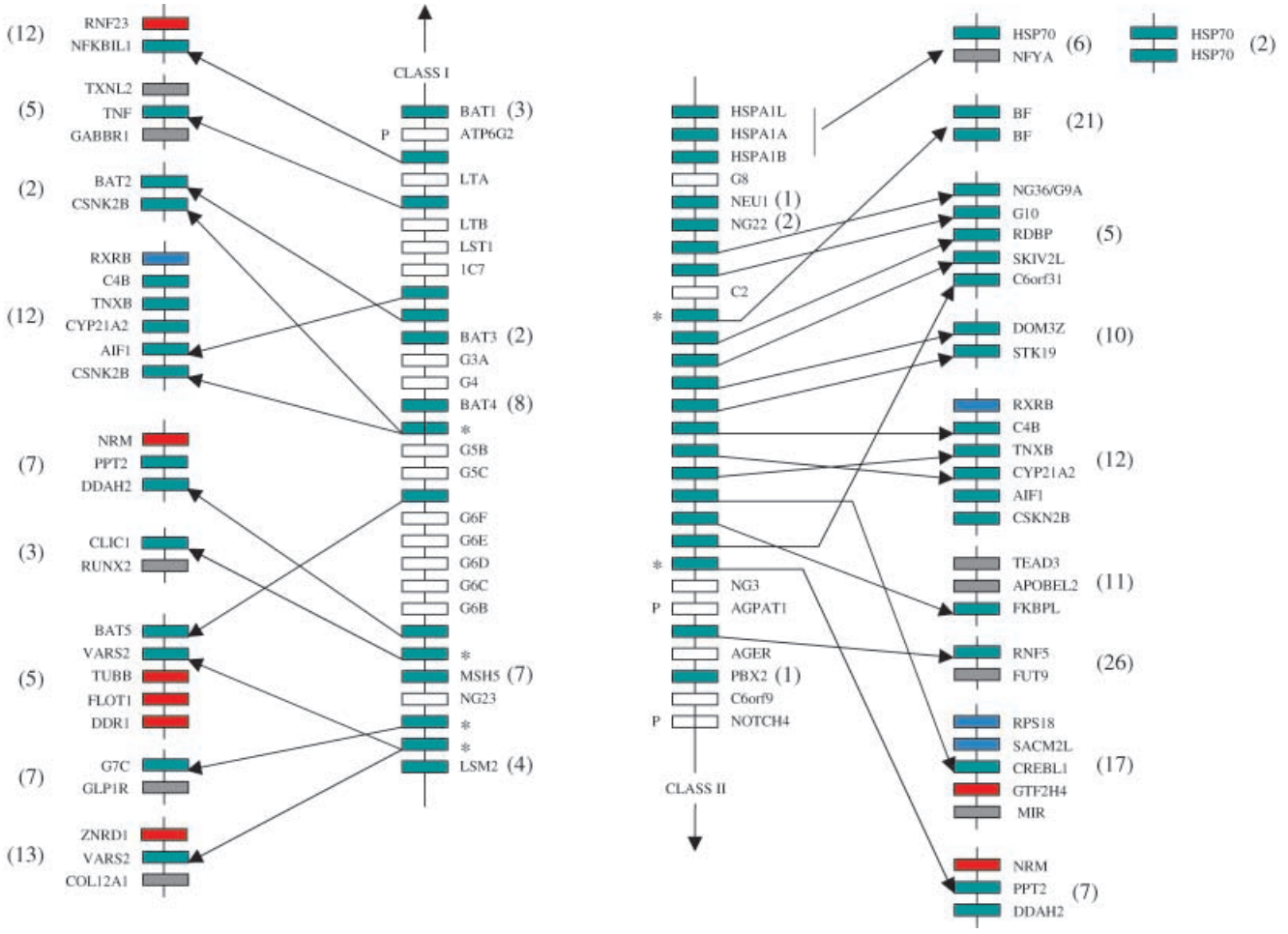

Fig. 3 Diagrammatic comparison of the human MHC class III region and the different identified *Fugu* scaffolds. The order of the scaffolds is not significant. Genes not present on 6p are not shown. The total number of genes on each scaffold is shown by the *number in brackets* at the side of each contig. Color code: *red* class I region genes, *dark blue* human classical MHC class II region genes, *light blue* human extended MHC class II region genes, *green* human MHC class III region genes, *gray* genes found on human Chr 6, but not within the MHC region. Duplicated genes in *Fugu* are marked by an *asterisk*. The symbol "*P*" refers to genes that have been identified as paralogues, but not as orthologues to the human class III region on 6p21.3. *ZNRD1*, *RNF23* and *NRM* have been mapped to the class I region using ensembl, but are not in the official MHC annotation

RXRε, as this is a duplicated teleost gene). Scaffold *S002594* also contained a group of five class III region genes (*NG36/G9A*, *G10*, *RDBP*, *SKIV2L*, *C6orf31*). Their relative gene order was equivalent to that found in human, however there are many genes present in human, which were missing in the *Fugu* contig. The only MHC class III region gene flanked by both class I and II region genes was *CREBL1* found in scaffold *S000146*. This scaffold contained 17 genes with four which mapped to Chr 6p21.3, as well as several genes that mapped to 1p34-p36, 1q21-q23 and 19p13.1.

Approximately a third of the *Fugu* genes were identified as singletons [*BAT1*, *BAT3*, *BAT4*, *CLIC1* (duplicated), *MSH5*, *G7C* (duplicated), *LSM2*, *NEU1*, NG22, *BF* (duplicated), *PPT2* (duplicated) and *PBX2*]. With the exception of the larger scaffolds above 60 kb containing *BAT4*, *MSH5* and *G7C*, the number of genes per scaffold was four or less and therefore analysis was limited. Four scaffolds (*S011760*, *S005829*, *S007225* and *S007803*) that were analysed only harbored singletons. Associations with human genes mapping to Chr 6, but not the MHC were observed in relation to *CLIC1*, *G7C*, *HSP70*, *FKBPL* and *RNF5*. Genes identified as pairs included *DOM3Z* and *STK19* that are found adjacent to each other in the human genome. These were linked to genes mapping mainly to Chr 7p22 and 16p11-p12. Scaffold *S000499* contained the MHC genes *NFKBIL1* and *RNF23* with an assortment of genes mapping to Chr 2, 3, 4, 8, 11, 12, 17 and 22. Scaffold *S001433* contained *PPT2* and *DDAH2*, both human class III region genes. *TNF* was found in scaffold *S000795* that also contained *GABBR1* and *TXNL2*, two human genes mapping to 6p21.3 and 6p25.3 respectively. Five previously characterized HSP70 genes (Lim and Brenner 1999) were identified in scaffolds *S002561*, *S006421*, *S007341*, and *S011760* with an additional member found in scaffold *S001038*. Human *HSPA1A* and *HSPA2B* genes encode protein products that are 99% identical. These are 88% identical with *HSPAIL* and hence their *Fugu* counterparts could not be discriminated. Therefore, in the *Fugu* data, these genes have all been collectively named *HSP70*.

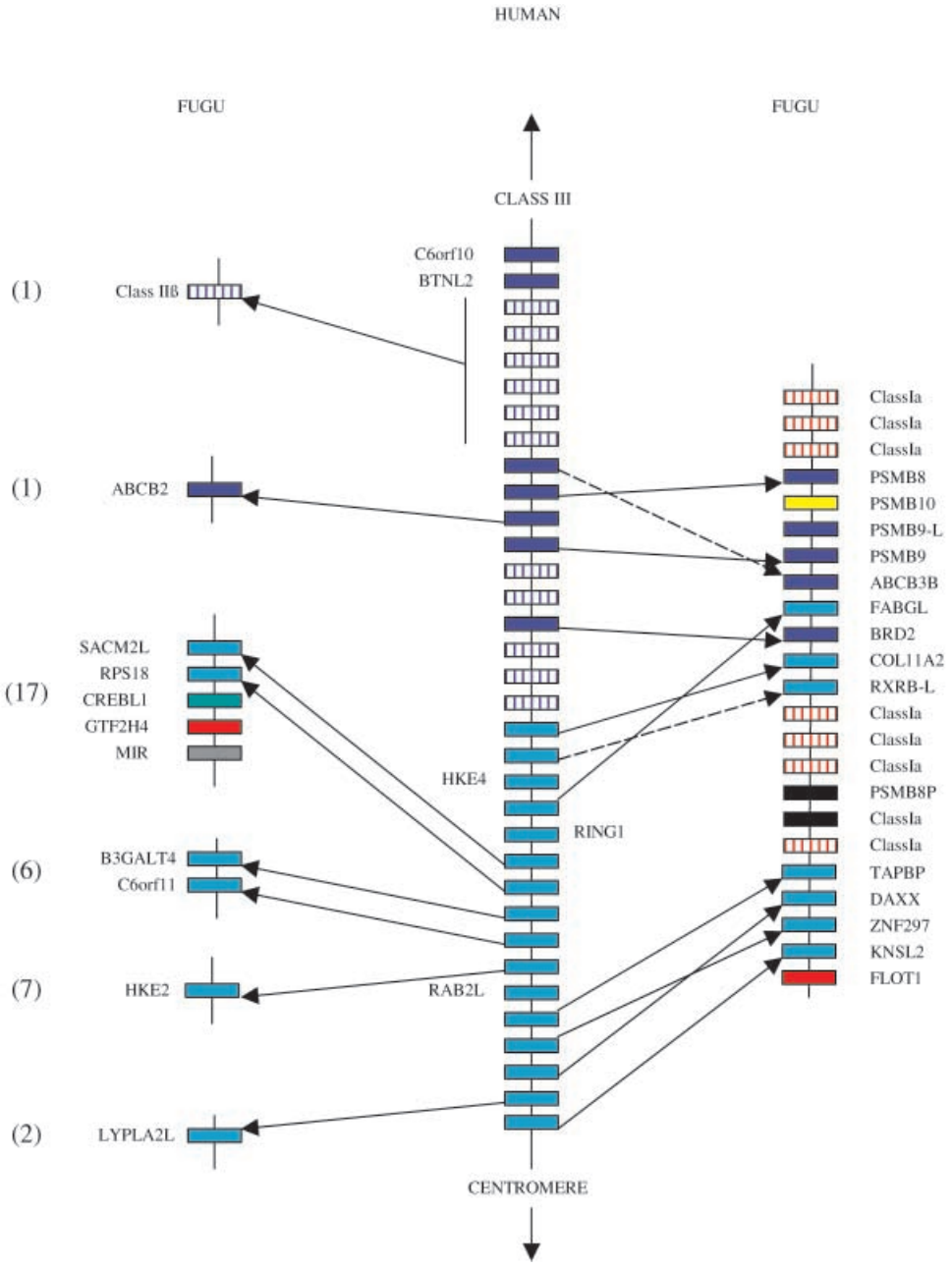

Fig. 4 Diagrammatic comparison of the human MHC class II region and the different identified *Fugu* scaffolds. The order of the scaffolds is not known. Genes not present on 6p are not shown. The total number of genes on each scaffold is shown by the *number in brackets* at the side of each contig. The main *Fugu* class I region contig, which contains class II region genes, does not have such a figure, as it is a representation of previous data and is currently split between several scaffolds. Color code: *red* class I region genes, *red striped* MHC class Ia genes, *dark blue* human classical MHC class II region genes, *dark blue striped* human class II antigen presenting genes, *light blue* human extended MHC class II region genes, *green* human MHC class III region genes, *gray* genes found on human Chr 6, but not within the MHC region, *black* known *Fugu* pseudogenes, *yellow* gene involved in class I antigen presentation, but not mapping with the human MHC region

The *Fugu HSP70* gene in scaffold *S001038* showed linkage to *NFYA*, a human gene mapping to 6p21.

Four of the identified MHC class III region genes were found to have two copies in *Fugu*, namely *CSNK2B*, *CLIC1*, *VARS2* and *PPT2*. The duplicates of *CSNK2B*, found in scaffolds *S000061* and *S003195*, were both found to be linked to different class III region genes. One association with *BAT2* is also found in zebrafish. Both *Fugu* copies of *VARS2* were linked to class I region genes (in scaffolds *S000755* and *S000553*). The smaller contig (*S000755*) contained the short-range linkage between *BAT5*, *VARS2*, *TUBB*, *FLOT1* and *DDR1* (discussed in the class I section above). The second *VARS* gene (in *S000553*) was flanked by *PPPIR11* and *ZNRD1*, two class II region genes, a fragment of *COL12A1* located in Chr 6q12-q13, and genes mapping predominantly to Chr 1. A previously identified *VARS2* gene (Lim and Brenner 1997) had highest sequence similarity to the one found in scaffold *S000755*, although data regarding the surrounding genes was contradictory. Unlike the complete sequencing data analysed here, Lim and Brenner (1997) used random shotgun sequencing which revealed *VARS2* to be linked to genes resembling tenascin X, the nuclear antigen A/Ro of Sjogren's syndrome and the Landsteiner-Weiner blood group glycoprotein and not to class I genes. This contradiction between the shotgun sequencing and scaffold data is most easily explained by a co-ligation of the cosmid used by Lim and Brenner (1997). *BF* and *G7C* were the only class III region genes to be found in triplicate. Two cop-

ies of *BF* were found in scaffold *S000298* (linked to *RPL23AP1*) and one copy found in *S004468*. The designation of these genes as *BF* is a matter of convenience in Table 2, as analysis shows these genes to be intermediate between *BF* and *C2*. Genes mapping mainly to human Chr 17 in scaffolds *S000951* and *S001291* surrounded *G7C*, although the third copy of *G7C* in scaffold *S000537* also contained a gene mapping to 6p21.

Mapping of human MHC class II region orthologues

Of the 24 genes identified from the human classical class II and extended class II regions, only six were not found (Table 3). Most of the genes which were identified were present in the 300 kb class I contig previously discussed (Clark et al. 2001) (Fig. 4). The scaffolds did reverse the order of the two class II region genes; *FABGL* and *BRD2*, though this result must still remain questionable. The previous arrangement of these two genes was based on comparison with zebrafish map data and most importantly long-range PCR on *Fugu* genomic DNA between *ABCB3B* and both *FABGL* and *BRD2*. Only primers between *ABCB3B* and *FAGBL* gave a PCR product, which was then sequence verified. Closer scrutiny of scaffold *S001220* revealed a tract of inverted repeats between *Col11A2* and *FABGL* and this is probably the cause of the incorrect assembly.

Of the remaining genes, the only conserved short-range linkages were those between *B3GALT4* and *C6orf11* and also *SACM2L*, *RPS18*, *GTF2H4* and *MIR* (discussed in the class I section above). The class II antigen presenting genes (designated *HLA-D** in Table 3) were conspicuous by their absence in the scaffolds. Only a single scaffold was identified (*S011611*) as containing a class II β chain and this gene is present on a small contig of 2098 bp with no other genes. Previous library screenings (data unpublished) confirmed this with only one class II-containing cosmid identified (*C025F12*).

One of the problems during the course of this work has been the determination of orthology. A prime example of this was the *RING1* (ring finger protein 1) gene in the human class II region. BLAST analysis of the human sequence against the *Fugu* scaffolds revealed strong sequence similarity to a single scaffold, *S002396*. This scaffold showed highest sequence similarity to the zebrafish *RING1B* gene fragment and also the human *DING* gene (described as either RNF2, ring finger protein 2 or similar to RING1, and mapping to human Chr 1q25.1). Sültmann and co-workers (2000b) placed the zebrafish *RING1B* gene within the MHC. However phylogenetic analysis (Fig. 5) using the limited numbers of sequences within the database, clearly placed both the *Fugu* gene *S002396* and zebrafish *RING1B* with the human and mouse *RING2* sequences, indicating that both of these genes are orthologous to the human sequence on Chr 1q25.1, and not the MHC paralogue.

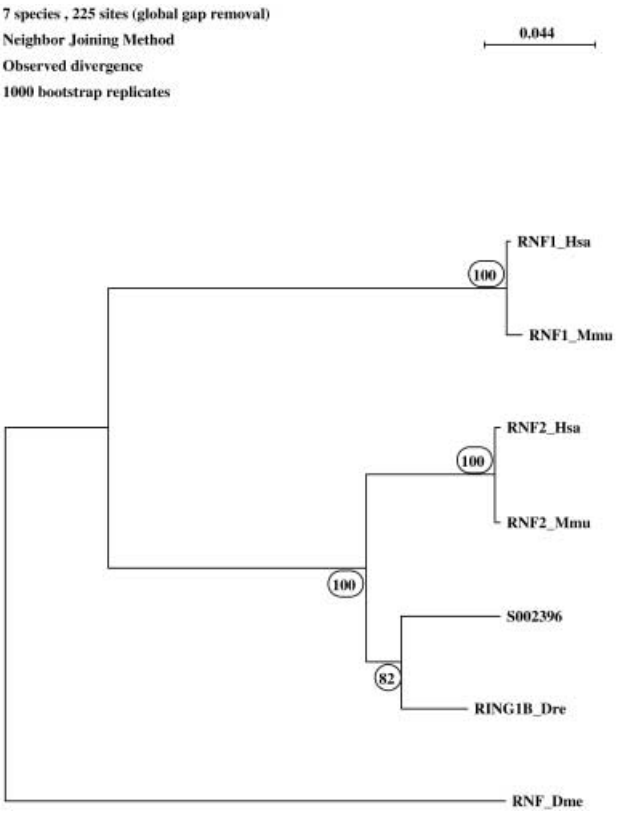

Fig. 5 Phylogenetic analysis of the *Fugu* ring finger protein gene (*S002396*). Abbreviations and accession numbers: *RNF1* ring finger protein 1, *Hsa* human (*Homo sapien*) Q06587, *Mmu* mouse (*Mus musculus*) Q921Z8, *RNF2* ring finger protein 2; Hsa: Q99496, Mmu: Q9CQJ4, *Dre* zebrafish (*Danio rerio*) (RING1B) Q9PTH4; outgroup: RNF ring finger protein, *Dme* fruit fly (*Drosophila melanogaster*) Q9VBO8

Discussion

Excluding the class Ia and class II antigen presenting genes, 108 genes from the human MHC region in 6p21.3 were searched for BLAST sequence similarity against the *Fugu* draft genome database. Of these, 36% were not identified. There are a number of explanations for this, primarily that the whole genome has not been completely sequenced to reference standard. *Fugu* has 22 pairs of chromosomes (Miyaki et al. 1995) and, therefore, if the genome had been completely sequenced, 22 giant scaffolds would be expected. However, at present there are 12,403 scaffolds in the current data set release (http://www.fugu.hgmp.mrc.ac.uk) and clearly gaps exist. Certainly, the lack of coverage in some cases could be due to the presence of retroelements, which are problematical when attempting to obtain universal coverage of genomic clones. Also to optimize the accuracy of the automated assembly, the Consortium masked out any repeat elements which could not be spanned by a single read (Fugu Genome Consortium 2002). Such elements have been found in high numbers associated with the *Fugu* MHC class I region (Clark et al. 2001) and are also present on the only class IIβ locus identified.

One of the main difficulties in identification is associated with levels of homology shared between orthologous *Fugu* and human proteins. For example, the *Fugu B3Galt4* was identified as sharing approximately 33% sequence identity with the human orthologue. If this gene had been on a scaffold on its own, it would probably have been overlooked, but identification was aided by its association with *C6orf11* (*BING4*), which is the neighboring gene in the human class II region. Genes, such as *G7C* and *BAT4* which were not associated with other MHC or Chr 6 genes, were facilitated in their identification by using the Genscan protein predictions. *G7C*, found in scaffolds *S000537*, *S000951* and *S001291*, yielded poor BLAST scores showing similarity to only half of the human gene, although the predicted *Fugu* protein sequence confirmed its homology. Unfortunately, the human *G7C* gene contains no obvious functional motifs or homologues in the database (Snoek et al. 2000) to establish its identification.

There is increasing evidence that certain gene families, particularly those involved in the immune response, such as the cytokine receptors and Ig families, are more rapidly evolving than others (Fugu Genome Consortium 2002). This rapid evolution could explain some of the problems in identifying all the genes. Certainly when listing the MHC genes not found in *Fugu* and then comparing the human and mouse orthologues, it is those which are involved in immune responses which differ the most in percentage identity. For example, many novel genes found in the human class III region are members of the immunoglobulin (Ig) or Ly6 super families (*G5B*, *G5C*, *G6F*, *G6E*, *G6D*, *G6C*, *G6B*, *1C7*) and percentage identities between mouse and human vary between 53% and 83%. Additional genes not found in *Fugu* which map within the human MHC class II region include *DHFRP2* which belongs to the carcinoembryonic antigen family and *BTNL2*, a member of the immunoglobulin super family, both of which share only 31–44% identity between human and mouse. Therefore, by extrapolation, it can be seen that percentage identities of these genes between human and *Fugu* will be even lower and consequently more difficult to detect, particularly at the genomic level, where weak BLAST matches to protein domains may be the only sites of sequence similarity detected. Other unidentified genes, such as *SPR1* and *HCR* have no known function and, therefore, it is difficult to hypothesize as to their conserved homology between *Fugu* and human. They both share a maximum of 74% identity between human and mouse and, whilst their absence in *Fugu* could be due to their divergent nature, perhaps because of some immune-related function, it is more likely that they are simply not in the assemblies.

There is also the situation of species independent evolution, as occurred with the class Ia genes (Klein et al. 1993; Nei and Hughes 1992; Nei et al. 1997). Therefore, genes such as *MICA* and *MICB* which are stress inducible class I homologues would also be expected to show a similar type of evolution and consequently reduced levels of protein similarity to their human orthologues. Certainly, these two genes only share a maximum of 43% identity between human and mouse. A further example of this is *TNF*. Phylogenetic analysis on the Japanese flounder *TNF* gene (Hirono et al. 2000) suggested teleosts possess one ancestral *TNF* as the mammalian lymphotoxin-α and *TNF-*α genes duplicated and evolved after mammals diverged from teleosts. This would explain the identification of *TNF* in *Fugu* but not of *LTA* or *LTB*. In mammals, *BF* and *C2*, located in tandem in the human class III region, share extensive amino acid similarity. They are thought to have arisen by gene duplication from an ancestral gene. Both zebrafish (Seeger et al. 1996) and medaka (Kuroda et al. 1996) possess genes equally similar to both *BF* and *C2*, even though in zebrafish there are two copies of this putative *BF*/*C2* ancestral gene (Gongora et al. 1998). Protein sequences derived from scaffolds *S000298* (where the *Fugu* gene is duplicated) and *S004468* in *Fugu* demonstrated an overall amino acid identity ranging from 32% to 34% with the human *BF* and *C2* sequences making the definition of orthology difficult. This adds further evidence to the hypothesis that teleosts possess an ancestral *BF*/*C2* gene, which could potentially be involved in both the classical and alternative pathways of innate immunity.

An additional point to make is that the scaffolds have been assembled automatically and although, in general, the standard of assembly is high (as judged using already mapped regions), occasionally the assembly may not be 100% accurate. One example of misassembly concerning the gene order of *FABGL* and *BRD2* has already been cited. Another example of this is exemplified by the assembly of *IPP2* (protein phosphatase inhibitor 2), found in the human class II region. Only one match for this protein was found using tBLASTx to scaffold *S000868*, but close analysis indicated that amino acids 41–75 were duplicated. Dotter analysis indicates that this was due to a missassembly.

Sometimes, orthology within this study was difficult to ascribe, as the MHC contains numerous examples of paralogous loci. This may be resolved either using phylogenetic analysis (Figs. 2 and 5) or by examination of closely linked genes. Using the latter approach, paralogues for *RalGDS* (Bouchireb et al. 2001), *ATP6G2*, *AGPAT1* and *NOTCH4* were identified in *Fugu*, but not the MHC orthologous sequences. Neither the orthologues for *AGPAT1* nor *NOTCH4* were identified in zebrafish using a variety of approaches including degenerate PCR, PCR using known zebrafish EST sequences, and heterologous hybridization using fish/amphibian cDNAs, although the paralogue for *ATP6G2* was found (Sültmann et al. 2000b). Family gene members found as singletons in the scaffold data, such as *PBX*, were subjected to phylogenetic analysis to establish orthology (data not shown). This approach was also taken with *NG22* and *CLIC1* where mapping data of flanking genes was limited. The identification of MHC paralogous loci does indicate that in fish similar gene duplication events have taken place in this region compared to human. Whether this is due to block duplication (Kasahara 2000)

or independent duplication events (Hughes 1998, 1999) is open to debate and will require a more in-depth analysis. The identification of conserved paralogous blocks will certainly be more difficult in fish, due to the numerous rearrangements that have occurred compared to human (Postlethwait et al. 2000). Whilst syntenic paralogous loci can be identified, short range conservation of gene order and position compared to human is uncommon (Bouchireb et al. 2001; data unpublished). This is not a phenomenon restricted to fish, as there is increasing evidence to show that numerous changes in gene order do occur between even closely related species (Carver and Stubbs 1997; Johansson et al. 1995; Nanda et al. 1999; Puech et al. 1997; Rattink et al. 2001; Schibler et al. 1998).

Overall, the results from the genome screen indicate that most of the human MHC orthologues are present in *Fugu*. What is becoming clear from this screen, along with the work on zebrafish, medaka, trout and *Fugu* (Clark et al. 2001; Hansen et al. 1999; Matsuo et al. 2002; Michalova et al. 2000), is that there is only one major MHC locus in fish. This contains the class Ia genes and additionally some of those in the human classical class II and extended class II regions, some genes of which, are involved in class I antigen presentation. The close linkage between class I and the class I antigen processing genes would appear to be logical in terms of a shared functional pathway, and regulation by α interferon, and this may have perpetuated this linkage across widely divergent teleost species.

From the data available in *Fugu*, zebrafish and medaka, the conserved linkage of the class I region appears to extend between *RXRB* (which is actually duplicated in teleosts and this locus is in fact *RXR*δ (Gongora et al. 1998) and flotillin (Clark et al. 2001; Matsuo et al. 2002; Michalova et al. 2000). The conservation of gene order starts to break down around the flotillin locus; in medaka, this gene is surrounded by *CIZ* and *TUBB* (Matsuo et al. 2002). On *Fugu* scaffold *S001569*, there is a zinc finger protein between *KNSL2* and *FLOT1*. The exact definition of this gene is difficult, but it is not the orthologue of the medaka *CIZ* gene (tBLASTx analysis of this gene against the *Fugu* scaffolds indicates this is most likely to reside in scaffold *S000627*). The putative *Fugu* orthologue of *TUBB* has been localized to a separate contig of class I and class II genes (including a duplicated flotillin gene) and cannot be contiguous with flotillin, as previous mapping data indicated an evolutionary breakpoint with Chr 17 (Clark et al. 2001).

The scaffold data also indicates the presence of a divergent or non-classical class Ia locus. Non-classical is a term used to describe genes with a typical class I structure that do not have the function or distribution of classical class I genes and they are not necessarily encoded in the MHC (Hedrick 1992; Klein and O'hUigin 1994). Although there is no functional data on this gene in *Fugu*, the structure is sufficiently different, with sequence similarity matches to both the cyprinid non-classical class I Z lineage sequences and the marbled lungfish, rather than the *Fugu* and closely related teleost class Ia genes in the database. This finding is substantiated by the identification of such a locus in salmonids (Shum et al. 1999) and sharks (Ohta et al. 2002). Interestingly, in medaka there is an additional class Ia locus which appears to stand alone on linkage group 22 (Naruse et al. 2000), although this has not been subjected to phylogenetic analysis.

Previous studies on the MHC of fish, particularly studying the locations of the class I and class II genes, have indicated that the genes in fish orthologous to the human MHC are dispersed onto several different chromosomes (Bingulac-Popovic 1997; Graser et al. 1998; Nonaka et al. 2001; Sato et al. 2000a; Sültmann et al. 2000b). A more detailed analysis of eight zebrafish class II region loci, of which six were not class II antigen presenting genes, revealed that these are dispersed over six linkage groups (Sültmann et al. 2000b). This would appear to be the case in *Fugu*, as apart from the main grouping of the class II region genes linked in the main *Fugu* MHC region, the remainder are dispersed across several different scaffolds. Figure 4 is simplified in terms of the gene content on each scaffold and the class II region genes are truly dispersed. For example, *B3Galt4* and *C6orf11* are surrounded by four other genes, which map to either human Chr 3 or 1p. Similarly, *HKE2* which is present as a singleton is actually surrounded by seven genes, most of which map to the human Chr X. The lack of success in identifying the class II antigen presenting genes is disappointing. However, fish appear to have very few of these loci, with an average of one to two class IIα and 1–2 class IIβ genes (reviewed in McConnell et al. 1998). A similar number can be expected in *Fugu*, but these genes are very small (250 amino acids) and may be associated with retrotransposable elements, which prove recalcitrant to cloning. Also, since the main *Fugu* MHC region containing the class Ia sequences is bounded by two breakpoints and class I and class II genes are linked in shark, it would appear that the class II sequences in fish have been translocated away from the main MHC region. These translocations would appear to involve other chromosomes, as is the case with zebrafish (Sato et al. 2000a) and they may be random, unrelated to any potential block duplication of the MHC into paralogous regions, as preliminary analysis of 6p21.3 paralogous sequences (J. Sambrook, personal communication; Bouchireb et al. 2001) has not revealed any further class II genes.

Investigation into the MHC class III region genes in fish has been relatively limited so far, centering on the cloning of the *BF*/*C2* gene (Kuroda et al. 1996; Nakao et al. 1998; Seeger et al. 1996; Sunyer et al. 1998). This is presumably due to the greater interest in the class I and II regions which are more closely associated with the immune response. An exception to the above example is the mapping of zebrafish class III orthologues by Sültmann and co-workers (2000b), in which 17 class III region genes were mapped using a radiation hybrid panel. These genes were dispersed over nine different linkage

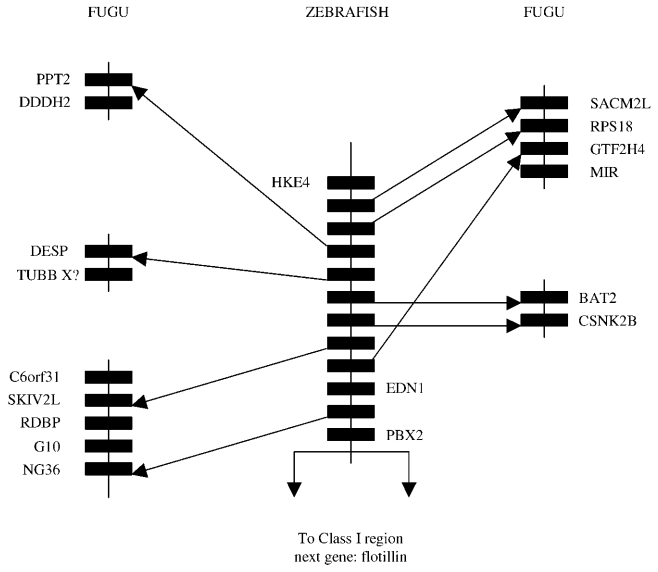

Fig. 6 A comparison of the cluster of genes in zebrafish mapped to *LG19*, the same chromosome which harbors the main MHC class I region (Sültmann et al. 2000b), to *Fugu* scaffold data

groups, with a cluster of genes mapped to *LG19*, the same chromosome which harbors the main MHC class I region (Sültmann et al. 2000b). They hypothesized that these additional human MHC orthologues (*HKE4*, *SACM2L*, *RPS18*, *PPT2*, *BAT2*, *CSNK2B*, *SKIV2L*, *GTF2H4*, *EDN1*, *G9A* and *PBX2*) formed two main blocks of conserved synteny with the human MHC class III region. These blocks potentially represented the vestiges of the vertebrate MHC; an ancient synteny group that existed before the divergence of the bony fish and tetrapod lineages more than 400 million years ago. Although the zebrafish data was in the form of radiation hybrid mapping and the *Fugu* data, as full sequence in scaffold format, the latter does confirm some of these short-range linkages, although there is a rearrangement involving *GTF2H4* (Fig. 6). As there is no physical localization of these scaffolds in *Fugu*, it is entirely possible that they map to the same chromosome as the class I region, confirming the status of these ancient conserved MHC blocks. Because of the dispersed nature of the genes in fish orthologous to those in the human MHC, this situation would benefit from further clarification in elasmobranchs, where it is known that the class I and class II genes are linked (Ohta et al. 2000, 2002). These organisms represent what is presumed to be the ancestral arrangement and in which there may also be continued linkage to an ancient class III region.

Of the remaining MHC class III region genes found in *Fugu*, there are two other major gene clusters. Scaffold *S000755* contained a mix of *BAT5* and a duplicated *VARS2* from the class III region linked with *TUBB*, *FLOT1* and *DDR1* from the class I region. Scaffold *S000061* contained one class II region gene *RXRB*/δ and four class III region genes, *C4B*, *TNXB*, *CYP21A2*, *AIF1* and a duplicated *CSNK2B*. Characterization of these genes does not reveal a common functionality, which could explain their continued linkage in *Fugu*. These scaffolds may also represent traces of an ancient MHC synteny group. Correlation of this data with other fish may provide more evidence to this effect. However, it should be noted that in the latter scaffold *RXRB*/δ maps to *LG16* and *AIF1* to *LG5* in zebrafish.

Although the fish may indicate remnants of an ancestral MHC, it would appear to have been considerably rearranged over a long evolutionary time scale, indicating that it is not essential for these genes to be in immediate proximity for their immune function (certainly as regards fish species). Teleost fish present the opportunity to examine vestiges of an ancestral MHC, despite the dispersed nature of the genes when compared to human. With reference to more applied aspects of this research; the MHC is the only fish genomic region to be subjected to full depth sequencing in a range of different species and there is great interest among fish geneticists in examining conserved synteny within fish species. This has obvious implications in the transfer of sequence information from model to commercial species where analysis of the immune system is vital for continued and successful production. The data from *Fugu*, *Tetraodon* and zebrafish will significantly aid our understanding of gene evolution, both within teleosts and vertebrates and also immunology within fish species.

Acknowledgements This work was supported by an MRC Programme Grant. J.G.S. and R.R. were in receipt of MRC studentships.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28:45–48

Beck TW, Menninger J, Voigt G, Newmann K, Nishigaki Y, Nash WG, Stephens RM, Wang Y, de Jong PJ, O'Brien SJ, Yuhki N (2001) Comparative feline genomics: a BAC/PAC contig map of the major histocompatibility complex class II region. Genomics 71:282–295

Bingulac-Popovic J, Figueroa F, Sato A, Talbot WS, Johnson SL, Gates M, Postlethwait JH, Klein J (1997) Mapping of MHC class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. Immunogenetics 46:129–134

Bouchireb N, Grutzner F, Haaf T, Stephens RJ, Elgar G, Green AJ, Clark MS (2001) Comparative mapping of the human 9q34 region in *Fugu rubripes*. Cytogenet Cell Genet 94:173–179

Carver EA, Stubbs L (1997) Zooming in on the human-mouse comparative map: genome conservation re-examined on a high-resolution scale. Genome Res 7:1123–1137

Chardon P, Rogel-Gaillard C, Peelman L, Yerle M, Velten FW, Renard C, Vaiman M (1999) Physical organization of the pig major histocompatibility complex class II region. Immunogenetics 50:344–348

Clark MS, Pontarotti P, Gilles A, Kelly A, Elgar G (2000) Identification and characterization of a beta proteasome subunit cluster in the Japanese pufferfish (*Fugu rubripes*). J Immunol 165: 4446–4452

Clark MS, Shaw L, Kelly A, Snell P, Elgar G (2001) Characterization of the MHC class I region of the Japanese pufferfish (*Fugu rubripes*). Immunogenetics 52:174–185

Fugu Genome Consortium (2002) The sequence, assembly and analysis of the compact vertebrate genome of *Fugu rubripes*. Science (in press)

Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 12:543–548

Gongora R, Zaleska-Rutczynska Z, Takami K, Figueroa F, Klein J (1998) Linkage of RXRB-like genes to class I and not to class II MHC genes in the zebrafish. Immunogenetics 48:141–143

Graser R, Vincek V, Takami K, Klein J (1998) Analysis of zebrafish MHC using BAC clones. Immunogenetics 47:318–325

Hansen JD, Strassburger P, Thorgaard GH, Young WP, Du Pasquier L (1999) Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, *Oncorhynchus mykiss*. J Immunol 163:774–786

Hedrick SM (1992) Dawn of the hunt for nonclassical MHC function. Cell 70:177–180

Herberg JA, Beck S, Trowsdale J (1998) *TAPASIN*, *DAXX*, *RGL2*, *HKE2* and four new genes (*BING 1*, *3* to *5*) form a dense cluster at the centromeric end of the MHC. J Mol Biol 277: 839–857

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Hirono I, Nam BH, Kurobe T, Aoki T (2000) Molecular cloning, characterization, and expression of TNF cDNA and gene from Japanese flounder *Paralychthys olivaceus*. J Immunol 165: 4423–4427

Hughes AL (1998) Phylogenetic tests of the hypothesis of block duplication of homologous genes on human chromosomes 6, 9, and 1. Mol Biol Evol 15:854–870

Hughes AL (1999) Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. J Mol Evol 48:565–576

Johansson M, Ellegren H, Andersson L (1995) Comparative mapping reveals extensive linkage conservation – but with gene order rearrangements – between the pig and the human genomes. Genomics 25:682–690

Kasahara M (2000) Genome paralogy: a new perspective on the organization and origin of the major histocompatibility complex. Curr Top Microbiol Immunol 248:53–66

Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923–925

Klein J, O'hUigin C (1994) The conundrum of nonclassical major histocompatibility complex genes. Proc Natl Acad Sci USA 91:6251–6252

Klein J, Sato A (1998) Birth of the major histocompatibility complex. Scand J Immunol 47:199–209

Klein J, Ono H, Klein D, O'hUigin C (1993) The accordian model of MHC evolution. In: Gergely J, Petranyi G (eds) Progress in immunology. Springer, Berlin Heidelberg New York, pp 137–143

Kuroda N, Wada H, Naruse K, Simada A, Shima A, Sasaki M, Nonaka M (1996) Molecular cloning and linkage analysis of the Japanese medaka fish complement Bf/C2 gene. Immunogenetics 44:459–467

Lim EH, Brenner S (1997) Short-range linkage relationships of the valyl-tRNA synthetase gene in *Fugu rubripes*. Immunogenetics 46:332–336

Lim EH, Brenner S (1999) Short-range linkage relationships, genomic organisation and sequence comparisons of a cluster of five *HSP70* genes in *Fugu rubripes*. Cell Mol Life Sci 55:668–678

Matsuo MY, Asakawa S, Shimizu N, Kimura H, Nonaka M (2002) Nucleotide sequence of the MHC class I genomic region of a teleost, the medaka (*Oryzias latipes*). Immunogenetics 53:930–940

McConnell TJ, Godwin UB, Cuthbertson BJ (1998) Expressed major histocompatibility complex class II loci in fishes. Immunol Rev 166:294–300

McShane RD, Gallagher DS, Jr. Newkirk H, Taylor JF, Burzlaff JD, Davis SK, Skow LC (2001) Physical localization and order of genes in the class I region of the bovine MHC. Anim Genet 32:235–239

Michalova V, Murray BW, Sültmann H, Klein J (2000) A contig map of the MHC class I genomic region in the zebrafish reveals ancient synteny. J Immunol 164:5296–5305

Miyaki K, Tabeta O, Kayano H (1995) Karyotypes in six species of pufferfishes genus *TakiFugu* (*Tetraodontidae*, *Tetraodontiformes*) Fish Sci 61:594–598

Murray BW, Sültmann H, Klein J (1999) Analysis of a 26-kb region linked to the MHC in zebrafish: genomic organization of the proteasome component beta/transporter associated with antigen processing-2 gene cluster and identification of five new proteasome beta subunit genes. J Immunol 163:2657–2666

Nakao M, Fushitani Y, Fujiki K, Nonaka M, Yano T (1998) Two diverged complement factor B/*C2*-like cDNA sequences from a teleost, the common carp (*Cyprinus carpio*). J Immunol 161: 4811–4818

Namikawa C, Salter-Cid L, Flajnik MF, Kato Y, Nonaka M, Sasaki M (1995) Isolation of Xenopus *LMP-7* homologues. Striking allelic diversity and linkage to MHC. J Immunol 155: 1964–1971

Nanda I, Shan ZH, Schartl M, Burt DW, Koehler M, Nothwang HG, Grutzner F, Paton IR, Windsor D, Dunn I, Engel W, Staeheli P, Mizuno S, Haaf T, Schmid M (1999) 300 million years of conserved synteny between chicken Z and human chromosome 9. Nat Genet 21:258–259

Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Kondo S, Hanamura N, Morita Y, Hasegawa K, Nishigaki R, Shimada A, Wada H, Kusakabe T, Suzuki N, Kinoshita M, Kanamori A, Terado T, Kimura H, Nonaka M, Shima A (2000) A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution. Genetics 154:1773–1784

Nei M, Hughes A (1992) Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: Tsuji K, Aizawa M, Sasazuki T (eds) HLA 1991, Proceedings of the eleventh international histocompatibility workshop and conference, vol 2. Oxford University Press, Oxford, pp 27–38

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci USA 94:7799–7806

Nonaka M, Namikawa C, Kato Y, Sasaki M, Salter-Cid L, Flajnik MF (1997a) Major histocompatibility complex gene mapping in the amphibian Xenopus implies a primordial organization. Proc Natl Acad Sci USA 94:5789–5791

Nonaka M, Namikawa-Yamada C, Sasaki M, Salter-Cid L, Flajnik MF (1997b) Evolution of proteasome subunits delta and LMP2: complementary DNA cloning and linkage analysis with MHC in lower vertebrates. J Immunol 159:734–740

Nonaka M, Matsuo M, Naruse K, Shima A (2001) Comparative genomics of medaka: the major histocompatibility complex (MHC). Mar Biotechnol 3:S141–S144

Ohta Y, Powis SJ, Coadwell WJ, Haliniewski DE, Liu Y, Li H, Flajnik MF (1999) Identification and genetic mapping of Xenopus *TAP2* genes. Immunogenetics 49:171–182

Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, Flajnik MF (2000) Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. Proc Natl Acad Sci USA 97:4712–4717

Ohta Y, McKinney EC, Criscitiello MF, Flajnik MF (2002) Proteasome, transporter associated with antigen processing, and class I genes in the nurse shark *Ginglymostoma cirratum*: evidence for a stable class I region and MHC haplotype lineages. J Immunol 168:771–781

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res 10:1890–1902

Puech A, Saint-Jore B, Funke B, Gilbert DJ, Sirotkin H, Copeland NG, Jenkins NA, Kucherlapati R, Morrow B, Skoultchi AI (1997) Comparative mapping of the human 22q11 chromo-

somal region and the orthologous region in mice reveals complex changes in gene organization. Proc Natl Acad Sci USA 94:14608–14613

Rattink AP, Faivre M, Jungerius BJ, Groenen MA, Harlizius B (2001) A high-resolution comparative RH map of porcine chromosome (*SSC*) 2. Mamm Genome 12:366–370

Renard C, Vaiman M, Chiannilkulchai N, Cattolico L, Robert C, Chardon P (2001) Sequence of the pig major histocompatibility region containing the classical class I genes. Immunogenetics 53:490–500

Rice R, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sato A, Figueroa F, Murray BW, Malaga-Trillo E, Zaleska-Rutczynska Z, Sültmann H, Toyosawa S, Wedekind C, Steck N, Klein J (2000a) Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. Immunogenetics 51:108–116

Sato A, Sültmann H, Mayer WE, Klein J (2000b) MHC class I gene of African lungfish. Immunogenetics 51:491–495

Schibler L, Vaiman D, Oustry A, Giraud-Delville C, Cribiu EP (1998) Comparative gene mapping: a fine-scale survey of chromosome rearrangements between ruminants and humans. Genome Res 8:901–915

Seeger A, Mayer WE, Klein J (1996) A complement factor B-like cDNA clone from the zebrafish (*Brachydanio rerio*). Mol Immunol 33:511–520

Shum BP, Rajalingam R, Magor KE, Azumi K, Carr WH, Dixon B, Stet RJ, Adkison MA, Hedrick RP, Parham P (1999) A divergent non-classical class I gene conserved in salmonids. Immunogenetics 49:479–490

Snoek M, Albertella MR, van Kooij M, Wixon J, van Vugt H, de Groot K, Campbell RD (2000) *G7c*, a novel gene in the mouse and human major histocompatibility complex class III region, possibly controlling lung tumor susceptibility. Immunogenetics 51:383–386

Stephens R, Horton R, Humphray S, Rowen L, Trowsdale J, Beck S (1999) Gene organisation, sequence variation and isochore structure at the centromeric boundary of the human MHC. J Mol Biol 291:789–799

Sültmann H, Murray BW, Klein J (2000a) Identification of seven genes in the major histocompatibility complex class I region of the zebrafish. Scand J Immunol 51:577–585

Sültmann H, Sato A, Murray BW, Takezaki N, Geisler R, Rauch GJ, Klein J (2000b) Conservation of MHC class III region synteny between zebrafish and human as determined by radiation hybrid mapping. J Immunol 165:6984–6993

Sunyer JO, Zarkadis I, Sarrias MR, Hansen JD, Lambris JD (1998) Cloning, structure, and function of two rainbow trout *BF* molecules. J Immunol 161:4106–4114

The MHC sequencing consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. Nature 401:921-923

Totaro A, Rommens JM, Grifa A, Lunardi C, Carella M, Huizenga JJ, Roetto A, Camaschella C, De Sandre G, Gasparini P (1996) Hereditary hemochromatosis: generation of a transcription map within a refined and extended map of the HLA class I region. Genomics 31:319–326

Walter L, Gunther E (1998) Identification of a novel highly conserved gene in the centromeric part of the major histocompatibility complex. Genomics 52:298–304