CMLS, Cell. Mol. Life Sci. 55 (1999) 668–678
1420-682X/99/040668-11 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Research Article

Short-range linkage relationships, genomic organisation and sequence comparisons of a cluster of five *HSP70* genes in *Fugu rubripes*†

E. H. Lim[a,*] **and S. Brenner**[b]

[a]Royal Brompton Hospital, Respiratory Medicine, Fulham Road, London SW3 6NP (UK),
Fax + 44 171 7273243, e-mail: elim@clementi.demon.co.uk
[b]The Molecular Sciences Institute, Inc., 9894 Genesee Avenue, 3rd Floor, La Jolla (California 92037, USA)

Received 17 November 1998; received after revision 25 February 1999; accepted 26 February 1999

Abstract. Twelve cosmids containing sequences resembling genes encoding members of the 70-kDa heat-shock protein family, HSP70, have been isolated from *Fugu rubripes*. They can be broadly divided into three groups of overlapping cosmids. Restriction analysis and sequencing of one set of five cosmids have revealed five intronless *Fugu HSP70* genes spanning 42 kb, arranged in a combined head-to-head, tail-to-tail and head-to-tail orientation. The levels of DNA and amino acid identity are very high with respect to one another, and are most similar to *HSP70* sequences linked to the major histocompatibility complex (MHC) region in other species. Putative heat-shock consensus elements are identified. Non-*HSP70* sequences with homology to known genes have been found physically linked to this *Fugu HSP70* cluster: the *Drosophila melanogaster SOL* gene, the *Drosophila melanogaster nemo* gene, the *Caenorhabditis elegans T17E9.1* gene and the sequence encoding the serine protease domain. The linkage relationships described here so far bear no resemblance to those of *HSP70* in other organisms. Convergence of mammalian *HSP70* and MHC class I and II loci probably occurred after fish had diverged.

Key words. *Fugu*; *HSP70*; MHC; sequence analysis; linkage relationships.

Heat-shock proteins were originally identified as a set of proteins synthesised when organisms respond to stresses, for example sudden increases in temperature [1, 2]. This heat-shock response has been found to be highly conserved throughout evolution as a physiological phenomenon, and at the level of the individual proteins [3].

Members of the eukaryotic HSP70 (M_r 70 kDa) family [4, 5] contain an N-terminal adenosine triphosphatase (ATPase) domain of ~44 kDa, and a C-terminal peptide binding domain of ~27 kDa. They are subdivided into those localised in the cytosol (e.g. Hsp70, Ssa), the endoplasmic reticulum (e.g. BiP, Grp 78), the mitochondria (e.g. Grp 75, Ssc), the Hsp110/Sse class and a few additional minor classes, for example Stch, which contains only the ATPase domain without the peptide binding domain. They serve several functions which are

* Corresponding author.
† The work for this paper was performed at Molecular Genetics, Level 5, Department of Medicine, Addenbrookes Hospital, Cambridge CB2 2QQ, UK, which has since been moved to the Sanger Centre, Hinxton Hall, Cambridge CB10 1RQ, UK.
Accession numbers: The nucleotide sequences reported in this paper have been deposited in the EMBL Database under Accession Nos. X95332, X95333, X95334, X95335, X95336, X95337, X95338, Y08576, Y08577, Y08578, Y08579, Y08580, Y08581, Y08582.

related to their cellular locations, for example protein folding, uncoating of clathrin-coated vesicles, translocation of proteins across membranes in mitochondria, chloroplasts and endoplasmic reticulum, and are components of mitochondrial endonuclease *Sce* I and steroid hormone receptor complexes. The work presented here focusses on cytosolic HSP70.

Multiple genes encode members of the HSP70 family in various species [6]. Gene duplications have occurred not only within each class of *HSP70* but within various eukaryotic phyla. For example, six cytosolic *HSP70* genes, *Ssa 1-4* and *Ssb 1-2* [7–9] are present in *Saccharomyces cerevisiae*, which correspond to recent duplications, probably during the evolution of ascomycetes. Similarly, in mammals, at least five cytosolic *HSP70* genes are present [4]. *HSP70* genes are observed to occur in groups of two to three in mammalian genomes, but in greater numbers per cluster in nonmammalian genomes. There is also the recurring theme of close linkage between a cluster of *HSP70* genes and the mammalian major histocompatibility complex (MHC). In humans [10, 11], mouse [12], rat [13] and pig [14] three *HSP70* genes are clustered and closely linked to their MHC. At least two *HSP70* genes are found linked to the MHC in the cow [15] and at least one in the goat [16]. Two or three *HSP70* genes are found linked to the *Xenopus* MHC [17]. These duplication events probably occurred before the divergence of the amphibians. In humans, other members have been located on chromosomes 1 and 14 [18, 19]. Similarly, in the mouse [12], cow [15] and goat [16], *HSP70* sequences are located in two other locations in their genomes. Larger numbers of *HSP70* genes in a cluster are found in *Trypanosoma cruzi* [20], *Leishmania amazonensis* [21] and *S. cerevisiae* [5, 22].

Several aspects of the origins of the MHC were reviewed by Klein and Sato recently [23]. The human MHC on chromosome 6 is traditionally used as the point of reference, and the terminology is such that 'class I' and 'class II' refer to the two regions containing clustered class I and class II loci (antigen-presenting genes), respectively, and 'class III' refers to that region between class I and class II that contains non-antigen-presenting genes, for example, those encoding HSP70 and valyl-transfer RNA (tRNA) synthetase [24]. Amongst other reasons, Klein and Sato suggest that these views are not tenable as there are numerous non-antigen-presenting genes interspersed amongst the class I and II loci. Also, MHC organisation in mammalian and nonmammalian animals is greatly varied. For example, certain rabbit haplotypes have large deletions between their class I and class II regions [25]. The entire chicken MHC region is packaged into a microchromosome (400 kb) with no discrete class III equivalent [26]. Orthologues of human class III genes are not consistently linked to antigen-presenting genes in other species. For example, the zebrafish complement factor B gene is not linked to class I or class II loci [27], similarly for the pufferfish [28] valyl-tRNA synthetase gene [29]. Paralogous groups of these MHC-linked, non-antigen-presenting genes have been found in humans and other species [30, 31], giving rise to the hypothesis that an ancient linkage group existed which contained some of the non-antigen-presenting genes now residing in the mammalian MHC. Our earlier report on the linkage relationships of *Fugu* valyl-tRNA synthetase [29] supports this. It also showed that unlike mammals, there is no close physical linkage between *HSP70* and the valyl-tRNA synthetase gene in *Fugu*, both in cosmid-range and pulsed field gel electrophoresis.

The aims of this work were first, to establish the identities of coding sequences in the vicinity of *Fugu HSP70* to see if they bear resemblance to mammalian MHC genes, and second, to examine in detail the coding and noncoding sequences of this cluster of *Fugu HSP70*. It was hoped that the results would be of interest with respect to the evolution of *HSP70* genes and, more indirectly, the MHC.

Materials and methods

Isolation of a fragment of human *HSP70*. As detailed in our previous report [29], a 1-kb fragment of human *HSP70* was generated by polymerase chain reaction (PCR) using primers based on the sequence of *HSP70-1* close to its 3′ end. Its identity was confirmed by sequencing.

Screening of *Fugu* cosmid library for *HSP70*. As described in our previous report [29], radiolabelled probes were generated by PCR, using deoxycytidine triphosphate (dCTP) supplemented with 1/10 volume of α^{32}P-dCTP. The partial *Mbo* I/*dam* methylase Lawrist 4 cosmid genomic library representing an estimated four-fold coverage was screened. Filters were hybridised according to the Church and Gilbert protocol [32, 33]. Exposure to X-ray film lasted 18–24 h at -70 °C. Positives were picked and cosmid DNA was extracted via the alkaline lysis method [33].

Sequencing of *Fugu HSP70* and identification of sequences with homology to known genes by shotgun sequencing. After *Eco* RI digestion of DNA from cosmid cL9F14, the restricted fragments were separated in Tris-acetate/EDTA (TAE) agarose gel [33]. DNA was extracted from the gel slices with the Geneclean kit. After ligation with *Eco* RI-cut pBluescript II KS$^+$, relevant fragments were identified by sequencing the ends of the subclones. For complete sequence information, the DNA underwent sonication (3×20 s, maximum power of ~ 500 watts) in a 50-µl volume, followed by T4 DNA polymerase end-repair (0.5 units per microgram

of DNA, 2 h at 12 °C in supplier's buffer), polyethylene glycol (PEG) precipitation and ligation with *Eco* RV-cut pBluescript II KS$^+$. Sequencing was performed with an automated DNA sequencer (ABI 373A) using Dyedeoxy terminator chemistry [34]. Sequence assembly was carried out using DNAStar software (DNASTAR). Gaps in contigs were closed by PCR and sequencing with walking primers.

To identify sequences with homology to known genes, DNA of cosmids cL27E11 and cL46B17 underwent sonication, subcloning and sequencing as above. These subclones were also used as hybridising probes during restriction mapping of the cosmids.

Digestion of *Fugu* cosmid and genomic DNA and restriction mapping. Digestion of *Fugu* cosmid and genomic DNA was performed using restriction endonucleases according to the manufacturer's recommendations (New England Biolabs). Restriction enzymes used included *Bgl* II, *Eco* RI, *Hind* III, *Mae* I, *Pst* I, *Sac* I, *Sau* 3AI and *Xba* I. After separation in agarose gels, the restriction fragments were blotted onto Genescreen Plus (NEN) by Southern transfer [35]. Radiolabelled probes were generated and hybridised onto these filters as above.

Preparation of dot-blot filters. After ligation and transformation [33], positives were picked and allowed to grow overnight in a 37 °C incubator. About 0.5 µl of each culture was dotted onto Genescreen Plus (NEN), and sequentially blotted with denaturant (0.5 M NaOH, 1.5 M NaCl) for 5 min and neutralising solution (1.5 M NaCl, 0.5 M Tris-HCl pH 7.4, 10 mM EDTA) for 5 min, twice. After washing in 2 × SSC for 5 min, the filters were baked at 80 °C for 2 h. Radiolabelled probes were generated and hybridised onto these filters as above.

Phylogenetic analysis. The ClustalW [36] package was used to align the *HSP70* sequences (complete) of the different species (obtained from SWISSPROT and EMBL, see fig. 4 for accession numbers) and to implement the phylogenetic analysis based on applying the neighbor-joining method [37] to the calculated distance matrix. Bootstrap analysis was carried out by the method of Felsenstein [38]. One hundred bootstrap resampling replicates were performed. From sequences retrieved from the database, only one of identical clustered sequences (e.g. MHC-linked) was included in the analysis. Partial sequences were excluded.

Results

Preliminary fingerprinting of twelve *Fugu* cosmids containing *HSP70* sequences. Twelve cosmids containing *HSP70* sequences were isolated after screening the *Fugu* cosmid library with a human *HSP70* probe. They underwent single digestions with the restriction enzymes *Sau* 3AI, *Pst* I and *Mae* I, and their Southern blots were probed with the same human *HSP70* fragment described above.

According to the restriction and hybridising patterns obtained, the 12 cosmids were grouped into three sets. One set of 5 overlapping cosmids (cL46B17, cL20B2, cL9F14, cL18P18, cL27E11) was selected for further analysis, as it gave rise to the greatest number of hybridising fragments.

Generation of *Fugu*-specific probes for *HSP70*. To obtain *Fugu*-specific probes for *HSP70*, cL20B2 was randomly chosen for shotgun sequencing. Those fragments that hybridised to the human *HSP70* probe (dot-blot hybridisation) were sequenced to confirm their homology to the 3′ end of *HSP70*. The smallest was 134 bp (38A4), and it was used as a probe for the 3′ end.

Sequencing of *Eco* RI fragments of cosmid cL9F14 led to the discovery of several *Eco* RI subclones containing the 5′ portion of *HSP70*. The smallest, 60E9, which is 2.8 kb in size, was used to identify the positions of the 5′ ends of *HSP70* in the cosmid contig.

Restriction mapping of five overlapping *Fugu* cosmids containing *HSP70*. Cosmid DNA was digested with various restriction enzymes, including *Bgl* II, *Eco* RI, *Hind* III, *Sac* I and *Xba* I. The Southern blots were serially probed with 38A4 (*Fugu*-specific probe for 3′ end of *HSP70*), 60E9 (*Fugu*-specific probe for 5′ end of *HSP70*), an *Eco* RI fragment subcloned from cL9F14 whose translated amino acid sequence is similar to the serine protease domain (see later), and several fragments obtained from shotgun sequencing of cL27E11 and cL46B17 (see later). The entire cosmid contig measures about 98 kb, whereas the five *HSP70* genes, *HSP70-1*, *HSP70-2*, *HSP70-3*, *HSP70-4* and *HSP70-5*, span about 42 kb. The restriction maps are shown in figure 1.

Identification of non-*HSP70* coding sequences physically linked to the *Fugu HSP70* cluster. The end-clones of the cosmid contig, cL27E11 and cL46B17, were chosen for shotgun sequencing to look for non-*HSP70* coding sequences in linkage with the cluster of *HSP70* sequences. Dot-blot hybridisation with 38A4 and 60E9 identified those subclones which did not contain sequence homologous to *HSP70*. In total, 41 subclones from cL27E11 and 31 subclones from cL46B17 were sequenced.

cL46B17 yielded two subclones, insert sizes being 279 bp (European Molecular Biology Laboratory, EMBL, accession number X95332, fig. 2A) and 396 bp (accession number X95333, fig. 2B), whose translated amino acid sequences bear similarities to the small optic lobe (SOL) protein found in *Drosophila melanogaster* [39]. These two subclones do not overlap in sequence. The 396 bp subclone was subsequently mapped onto the cosmid contig.

cL27E11 gave three subclones, one (accession number X95334, fig. 2C) whose translated amino acid sequence is similar to the nemo protein of *D. melanogaster* [40], and the other two (accession numbers X95335 and X95336, figs 2D, E) to the T17E9.1 gene product in cosmid T17E9 of chromosome III of *Caenorhabditis elegans* [41]. They measure, respectively, 266 bp, 333 bp and 385 bp. There is no overlap in sequence. All were mapped onto the cosmid contig.

Sequencing of both ends of *Eco* RI subclones from cL9F14 revealed two subclones, 1.8 kb and 1.7 kb, the translated amino acid sequence of one end of each possessed similarities to the serine protease domain of many proteins (accession numbers X95337 and X95338, figs 2F and 2G). The 1.8-kb subclone was subsequently mapped onto the cosmid contig.

Sequence comparisons of *Fugu HSP70* genes. Sequences of *HSP70-1*, *HSP70-2*, *HSP70-3*, *HSP70-4* and *HSP70-5* were obtained by shotgun-sequencing *Eco* RI fragments (sizes ranged from about 1.9 to 5 kb) subcloned from cL9F14. *HSP70-2* and *HSP70-4* have been completely sequenced. Sequences of *HSP70-1*, *HSP70-3* and *HSP70-5* are incomplete at their 3′ ends.

Comparisons of their translated amino acid sequences show high levels of similarity with respect to one another, upwards of 94% (not shown, nucleotide sequences deposited in EMBL database). *HSP70-2* contains a 249-bp deletion at its 3′ end, with a stop codon preceding the start of the deletion by 54 bases.

Comparing translated amino acid sequences of *Fugu HSP70-4* with *HSP70* of other species, it is more similar to the human MHC-linked *HSP70-1* (83.3% amino acid identity) and its orthologues in the cow (83.2%), pig (83.3%) and mouse (83.0%), than the non-MHC-linked cytosolic *HSP70* of these organisms (78–82%).

Sequence comparisons of the contiguous non-coding 5′ (fig. 3A, B) and 3′ (not shown) regions were also performed. Similarities were identified in the 5′ regions, but not the 3′ regions. A search for regulatory elements was conducted, specifically, the heat-shock consensus elements. They can be divided into two groups: *HSP70-1*, *HSP70-4* and *HSP70-5* forming one, *HSP70-2* and *HSP70-3* the other. A segment of homology between *HSP70-1*, *HSP70-4* and *HSP70-5* extends from −1 to −38 (1 base and 38 bases upstream of the start codon, respectively). Thereafter, it breaks down between

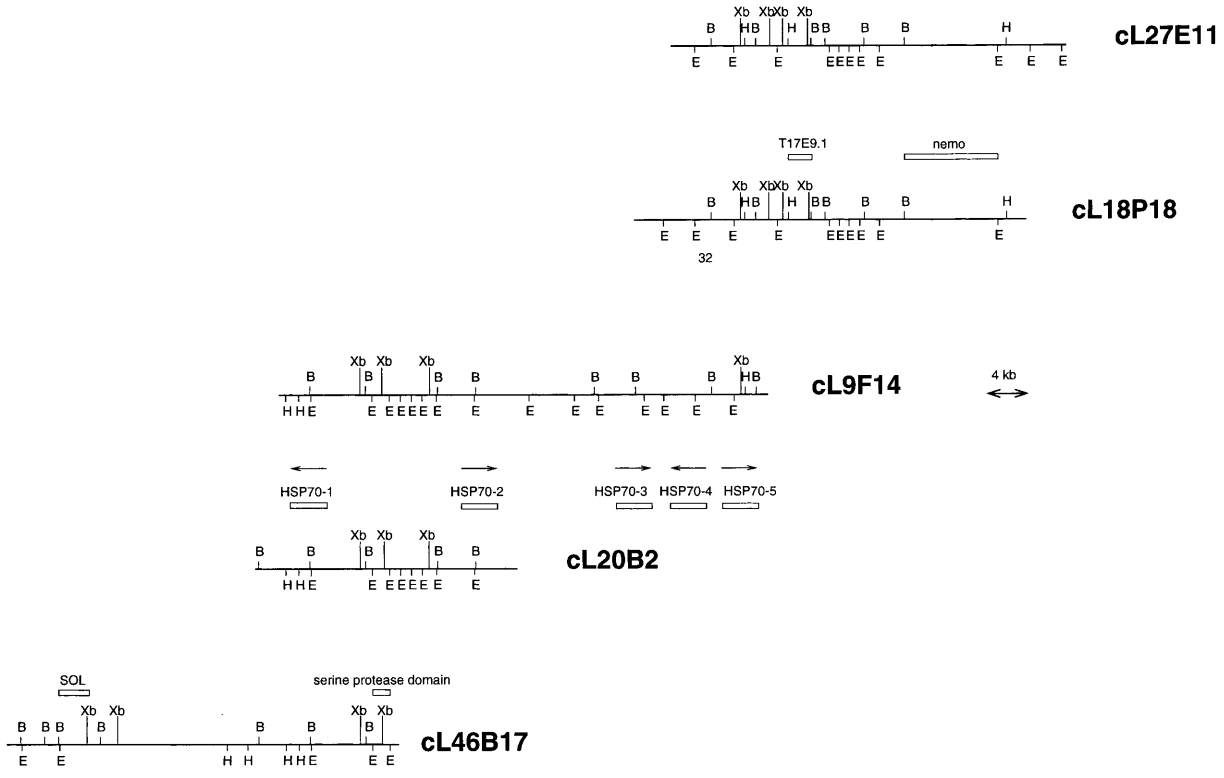

Figure 1. Cosmid contig and restriction maps of the five *Fugu* cosmids containing *HSP70* sequences. Boxes define locations of the five *HSP70* genes as well as the smallest restriction fragments to which the respective probes hybridised. Arrows indicate deduced directions of transcription. Restriction sites are marked according to the key E, *Eco* RI; H, *Hind* III; B, *Bgl* II; S, *Sac* I; Xb, *Xba* I.

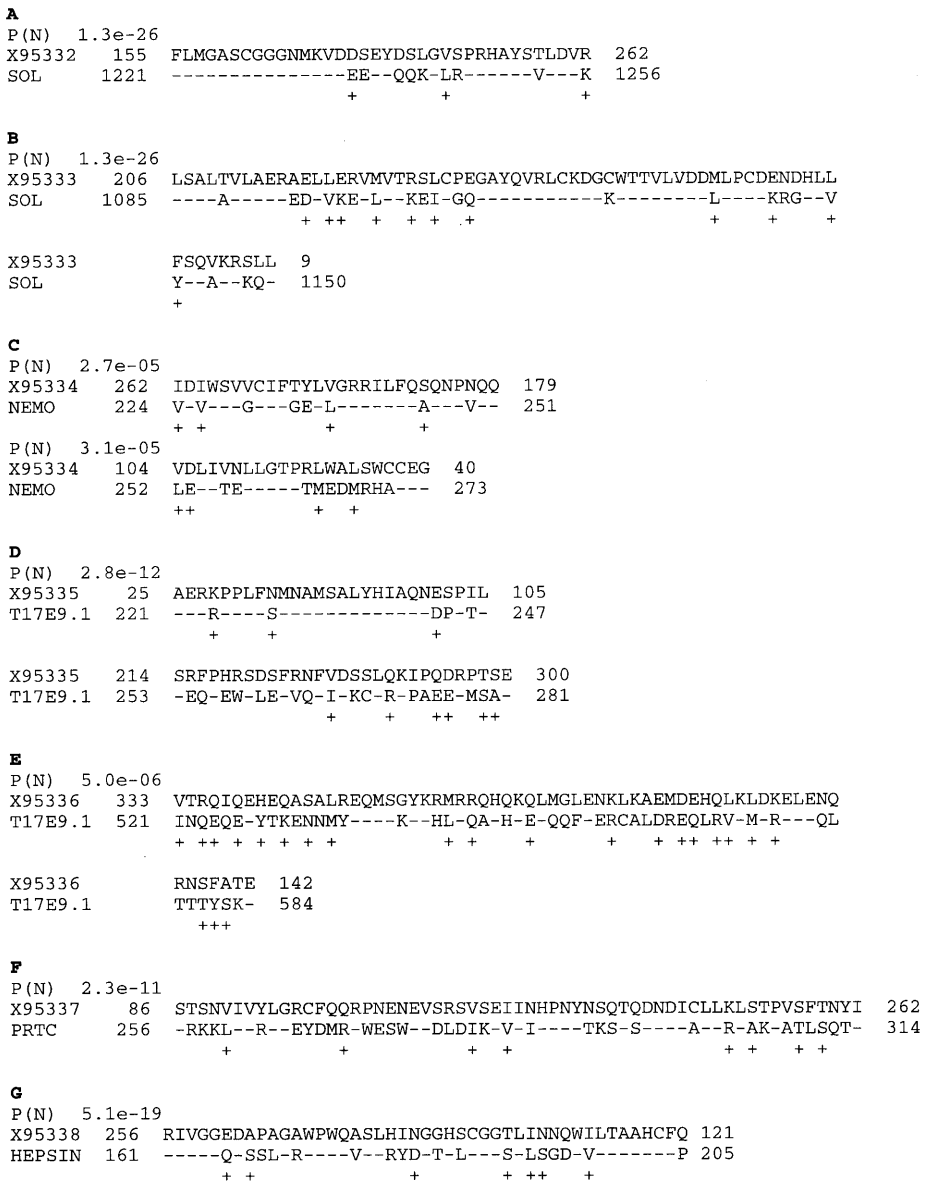

Figure 2. Pairwise alignments of translated amino acid sequences of *Fugu* subclones and genes of highest homology (Swissprot and EMBL database searches). Accession numbers are used to denote *Fugu* sequences. (*A* and *B*) Comparisons with *Drosophila SOL* (small optic lobes). (*C*) Comparison with *Drosophila nemo*. (*D* and *E*) Comparisons with *C. elegans T17E9.1*. (*F*) Comparison with bovine protein C gene. (*G*) Comparison with rat hepsin gene. Dashes (–) denote identity. Numbers indicate the nucleotide positions of the *Fugu* sequence and the amino acid positions of the segment of similarity of its best match. Plus signs (+) indicate a difference of one nucleotide base. P(N) refers to the smallest Poisson probability.

HSP70-4 and the other two, where the similarity between *HSP70-1* and *HSP70-5* continues, extending to −424 in *HSP70-1* and −365 in *HSP70-5*. However, this homology is discontinuous, the stretches of similarity being punctuated by insertions in *HSP70-1*, or deletions in *HSP70-5*. The homology between *HSP70-2* and *HSP70-3* extends to −1835 in *HSP70-3* and −891 in *HSP70-2*, and is also discontinuous.

Putative heat-shock elements (HSEs, see 'Discussion') have been identified in the five *Fugu HSP70* genes, and have been highlighted in figures 3A, B. Other putative regulatory elements have also been highlighted, namely TATA boxes and CAAT boxes. *HSP70-2* does not possess a TATA box. Together with data mentioned above, it is likely to be a pseudogene.

Discussion

Mutations in the *SOL* gene in *D. melanogaster* are known to cause specific classes of columnar neurons to degenerate in the developing optic lobes [39], leading to specific alterations in flight and walking manoeuvres. The SOL protein has a C-terminus similar to calpain, a calcium-activated protease. This is the region to which the translated amino acid sequences of the two subclones from cL46B17 show similarity. About 18–20 kb separate *HSP70-1* from the 2.6-kb *Eco* RI/*Xba* I fragment to which the 396 bp subclone hybridised. There is no data with regards the linkage relationships of *SOL* in other species for purposes of comparison.

The translated amino acid sequences of the two subclones from cL27E11 similar to the T17E9.1 gene product, a protein kinase, from *C. elegans* [41] correspond to different regions of the protein. The 333-bp subclone contains a 108-bp intron, based on the presence of splice donor and acceptor sites, and the breakdown in amino acid similarity in this interval. As both subclones colocalised to the same 2.4-kb *Hind* III/*Bgl* II restriction fragment, it is possible that they originate from the same *Fugu* orthologue of the *T17E9.1* gene. This restriction fragment is about 5 kb away from *HSP70-5*. The *T17E9.1* gene was isolated from *C. elegans* cosmid T17E9, which was cloned from chromo-

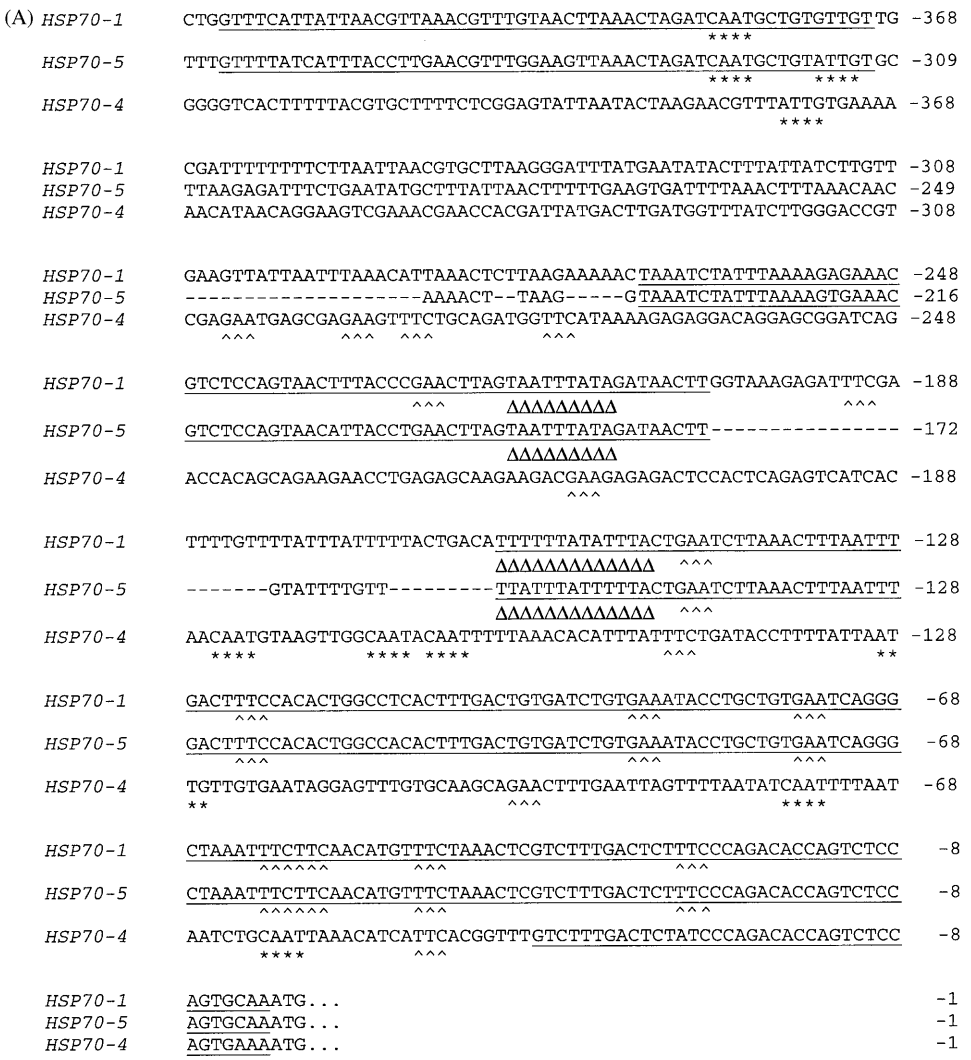

Figure 3. Alignment of 5′ noncoding regions of *Fugu HSP70* genes. (*A*) *HSP70-1*, *HSP70-4* and *HSP70-5*. (*B*) *HSP70-2* and *HSP70-3*. Regions of similarity are underlined. Gaps are represented by dashes (–). Putative heat shock consensus elements are marked with arrowheads (^). Putative TATA boxes are marked with triangles (△). CAAT and ATTG boxes are marked with asterisks (*).

some III. Another gene discovered in this *C. elegans* cosmid encodes N-myristoyltransferase.

Mutations of the *Drosophila nemo* gene affect eye development, causing blockage of dorsoventral rotation of the photoreceptor cell clusters [40]. The nemo protein shows similarity to a family of serine/threonine protein kinases, containing the 11 designated subdomains characteristic of this family. The translated amino acid sequence of a subclone isolated from cL27E11 is similar to subdomains IX and X. It is likely that an intron measuring 74 bp separates the segment coding for subdomain IX from the segment coding for subdomain X, as the similarity in translated amino acid sequence breaks down in an interval flanked by splice donor and acceptor sites. This subclone maps to an 8.5-kb *Bgl* II/*Eco* RI fragment about 9 kb away from the *Fugu*

```
(B) HSP70-2     AAGCTGCACTGTACAACGAAAATAGTTTAAACGAATAAAACCTTTAAAAAGCCCATGTCA  -831
    HSP70-3     TAGCTGCACTGTACAATGAAAATAGTTTAAACGAACAAAACCTTTAAAAAACCCATGTCA -1774

    HSP70-2     ATATTTGGTTCATTAATATTCCCAGTTTTGTAAGAGGTCT----------------GGAC  -787
    HSP70-3     ATATTTGGGTCATTAATATTCCCAGTTTTGTGAGAGGTCTTAAACATATCCCAAATTCTA -1714

    HSP70-2     GAGTAGCTGGGGAGAGGGAAGTCTGGGCTTCTCTCCTTAGGCTGCTGCCCCGTGACCCGA  -727
    HSP70-3     TCTATAATTATTAATTGTTTTCCGGTTGCCTTACTGGAAGTTTAAATAATAAAAAGAAAA -1654

    HSP70-2     CCCCGGATAAGCGGTAGAGAATGGATGGATGGATGGATGGATGGAAGTTCACAAAGCAGG  -667
    HSP70-3     AGCATAATACATGTTAAAGTAGCAGTAGTTGAGATTAAAGACTGAAGTTCACAAAGCAGG -1594

    HSP70-2     CTGCTTGTCCTGGCTCAAAAAATGTCTCTACACAAGAAACCTGTTGGGACAGAGACATAA  -607
    HSP70-3     CTGCTTGTCCTGGCTCAAAAAATGTCTCTACACAAGAAGCCTGTTGAGACAGA---CATAA -1536

    HSP70-2     TGTGTTGTAACCTGCAATTACTACAAGTCACACCCAGACAGCATCAGCAAATCCACCTCC  -547
    HSP70-3     TGTGTTGTAACCTGCAATTACTACACGTCACACCCAGACAGCATCAGCAAATCCACCTCC -1476

    HSP70-2     AACTTGGTCTGTTTTCCAGTAATACTGCATACTAAATACGAATCAGATTGACATCAGTGC  -487
    HSP70-3     AACTTGGTCTGTTTTCCAGTAATACTGCATACTAAATATGAACCAGATTGACATCAGTGC -1416

    HSP70-2     TGGGCGAATCATTTTATTTTTTTCTACTTATTAAGACATCTAAAACACTCACTGGTC---  -430
    HSP70-3     TGGGCGAATCATTTTATTTTTTTCTACTTATTAAGACATCTAAAACACTCACTGGTCTTT -1356

    HSP70-2     ------------------------------------------------------------
    HSP70-3     TGTTTGTACGATCTTGTCATTATTTTATTTATTCATTTCTGATTATCATGTCTAGCTTCC -1296

    HSP70-2     ------------------------------------------------------------
    HSP70-3     CGTGTGCAACTCCAAAGTATGATGGGATATATTTCTTGGTTAGGGGCGTTACTTGTTCCC -1236

    HSP70-2     ------------------------------------------------------------
    HSP70-3     TATTTTTCACAATGCACGATGTAATACATGAGTTCACTACACAGTGATCCTCACGCAACA -1176

    HSP70-2     ------------------------------------------------------------
    HSP70-3     TTCTGACAACCGTACAAAATGGTGTACACACTATTGAGTGTACTATAGAGAGGACAGGGA -1116

    HSP70-2     ------------------------------------------------------------
    HSP70-3     GTACTTTAAAAGTGCATGTCATGCCTGATCTGCGCCATACCACTTCTGAGCATACCCACA -1056

    HSP70-2     ------------------------------------------------------------
    HSP70-3     TTTTTAACACATGCAGCCCTCCCCCTAAACACACACACACACACACACACACACACAC  -996

    HSP70-2     ------------------------------------------------------------
    HSP70-3     ACACACACACAGAGTCAGTCCATTTACATCACAAGTAGAGATGTGAGTAGTTTGAACAGC  -936

    HSP70-2     ------------------------------------------------------------
    HSP70-3     TTCTTATTGTAAAAAAAAAACCCCAAAATTACGTTAATTGCTGGGAAATATGCATTGTGA  -876

    HSP70-2     ------------------------------------------------------------
    HSP70-3     TTGGGGGATTTGGCAAAAGCCCCAAATGAAATCATAAAATATTTATTGCTTTTTTGTTAT  -816

    HSP70-2     ------------------------------------------------------------
    HSP70-3     CATAAATCCTAATTAATGCTAACATGAGGTGATGTCTTTCCAGAGTGTGTATTAATTCAC  -756

    HSP70-2     ------------------------------------------------------------
    HSP70-3     AAAAACTAAAACATGTTGATTTTAGGGTTTATATCACGAAAAACTTCACATTAAAGTCAC  -696

    HSP70-2     --------TTTATATACCTGGTGAGACTCTGCATAATTTTATTGACGAGATCACTCACAC  -378
    HSP70-3     AGTAATTATTTTTTATACCTGGTGAGAAACTGCATAATTTTATTGACAAGATCAATCAAAC  -636
```

Fig. 3. (Continued)

orthologue of *T17E9.1*. There is no linkage data relating to *nemo* available from other species for comparison.

The translated amino acid sequences of two *Eco* RI restriction fragments subcloned from cL9F14 showed similarity to the serine protease domain of many proteins. The best database matches were to blood proteins such as protein C (involved in the prevention of thrombus formation), hepsin, kallikrein precursor (kinin system) and factors of the blood-clotting system. The C2 and factor B components of the complement system also contain this domain, and the genes encoding them reside in the MHC class III region of several organisms. However, until the entire gene is sequenced, there is insufficient data to confirm its identity. The 1.8-kb *Eco* RI fragment was mapped and found to be between *HSP70-1* and *HSP70-2*, about 2.5 kb from the former and 8 kb from the latter.

Unlike the *Fugu* valyl-tRNA synthetase gene [29], which has certain physical linkage similarities to the MHC region in other organisms, there does not appear to be any with this cluster of *Fugu HSP70* genes. In addition, a fragment of a *Fugu* MHC class II antigen-presenting gene [42] failed to hybridise to any of the 12 *HSP70*-containing *Fugu* cosmids. Also, PCR with primers based on the sequence of this *Fugu* MHC fragment, using the DNA of these cosmids as template, failed to amplify any product. Better understanding of linkage relationships between *Fugu HSP70* and *Fugu* orthologues of mammalian MHC genes will be obtained when the remaining seven cosmids are analysed in similar fashion.

Gene amplification has been shown to occur in tandem head-to-tail arrays, as well as inverted head-to-head or tail-to-tail arrays in eukaryotic cells. A number of mechanisms have been proposed to explain the former, for example unequal crossing over and disproportionate replication. A model for inverted duplication was proposed by Passananti and co-workers [43] which mimics the amplification of the 2 μm circle plasmid in *S.*

```
HSP70-2   TATGGCGCAAATACAGACACAGATACACACACAGAGATAATTAGAGTGCAGGG-------  -325
HSP70-3   AATGAAACAAATACGGAAACATATACACATACCAAGAAAATTATATTGCAGGGGGAAAAA  -576

HSP70-2   ------------------------------------------------------------
HSP70-3   GGCGTGAGAAAGGCCAACAAGTTAAGGCCCAACACCTCCAACTGCTGTTGCAAAGAGCTA  -516

HSP70-2   ------------------------------------------------------------
HSP70-3   CAGGTGAAACTGGCTGAACAGGTTCAGAAATGCCTGGCCTCAAAGTAGAAATAACATCAA  -456

HSP70-2   ------------------------------------------------------------
HSP70-3   GCAGGAAATGTGATAATCAGAGGGTCAGAGATCCGTGATCTTTCAAGGATGAAGGCCAAG  -396

HSP70-2   ------------------------------------------------------------
HSP70-3   GTCCTTGATCAGAAAGCAACGTACTATGTTATGTGTTTCTACTTGTGATTCTACAGCCTC  -336

HSP70-2   -------------------GGTTAACTTGTAGTTTTTACACTCCACAACCTCCATATGTA  -284
HSP70-3   CATATACGATACTGGTACTGGTTAACTTGTAGTTTTTACATTCAACAACCTCCATATCTA  -276


HSP70-2   GGTAACTCAGATCAATTAGAAAAA--CTTTATCTACTTTTAACTCCTCCTATCTTCCTTA  -226
                    ****  ^^^                                 ^^^
HSP70-3   GGTAACTCAGATCAATTAGAAAAACACTTTATCTACTTTTAACTCCTCCTATCTTCCTTA  -216
                    ****  ^^^                                 ^^^

HSP70-2   GTTAATTGCTACTGTTTGTAAATTCTTGCCACATCATTCTGCTCTATATACAGTATATAT  -166
          ****
HSP70-3   GTTAATTGCTACTGTTTGTAAATTCTTGCCACATCATTCTGCTCTATATACAGTATATAT  -156
          ****

HSP70-2   AGAATTATATATTTTAATAATTTGAATATTTAGGGTTGGAAAAATGAACAAATTCTTCCT  -106
          ΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔ                ^^^    ^^^    ^^^^^^
HSP70-3   AGAATTATATATTTTAATATTTA----------GGTTGGAAAAATGAACAAATTCTTCCT  -106
          ΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔ                ^^^    ^^^    ^^^^^^

HSP70-2   AAATCTAAACTCTGCCTCTTTTAACCCTGAACGGCTGCAGTTACTTTCAAATTTTACCTC   -46
                                  ^^^                ^^^
HSP70-3   AAATCTAAACTCTGCCTCTTTTAACCCTGAACGGCTGCAGTAACTTTCAAATTTTACCTC   -46
                                  ^^^                ^^^

HSP70-2   CTAATTTTATTTATCAACTTTTATAGGTAAAAAAAAGTAATCAAGATG...             -1
                       ΔΔΔΔΔΔΔ
HSP70-3   CTAATTTTATGTATCAACTTTTATAGGTAAAAAAAAGTAATCAAGATG...             -1
                       ΔΔΔΔΔΔΔ
```

Fig. 3. (Continued)

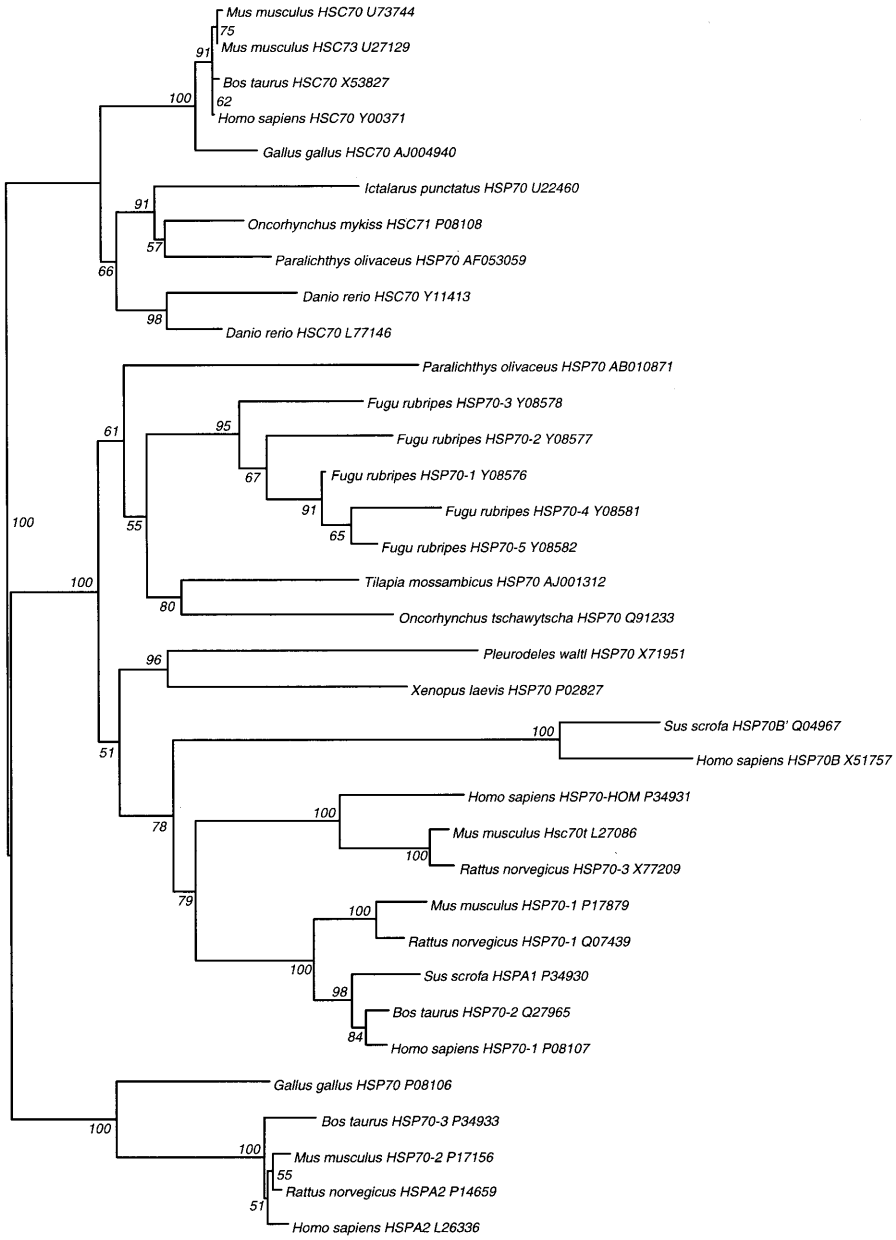

Figure 4. Phylogenetic tree of vertebrate cytosolic 70-kDa heat-shock proteins. Bootstrap support values are indicated at the nodes. The scale bar represents the number of substitutions per site. Gene designations and accession numbers are included.

cerevisiae. In humans, the three *HSP70* genes in the MHC class III region are arranged such that *HSP70-1* and *HSP70-Hom* are in a head-to-head orientation, and *HSP70-2* is upstream of *HSP70-1* and in the same direction [11]. In *D. melanogaster*, two *HSP70* genes are in a head-to-head orientation [44]. In *T. cruzi*, seven *HSP70* genes are in a head-to-tail tandem array [20]. *L. amazonensis* also has seven *HSP70* genes clustered and arranged in a head-to-tail fashion [21]. The five *Fugu HSP70* genes examined here are in a combined head-to-tail, head-to-head and tail-to-tail arrangement. In terms of gene orientation, they resemble humans. As regards the number of genes per cluster, they are similar to the nonvertebrate species mentioned above.

The early observations that transfected heat shock genes were inducible in heterologous systems suggested

the existence of common regulatory elements in these genes. These elements reside in a region of about 400 nucleotides upstream of the start site [45–49], and include the repeatedly occurring HSEs. The HSE [50] was initially described as a palindromic sequence, CnnGAAnnTTCnnG. Later, it was thought to be inverted adjacent pentamers containing the primary sequence 5′-nGAAn-3′, with a variable number of intervening bases. They serve as binding sites for heat-shock factors, HSFs [50, 51]. Multiple copies of the HSE exist upstream of *HSP* genes, and are thought to function in a cooperative manner, the efficiency of transcriptional activation being related to the number of HSEs present. The HSEs can also be positioned at different distances from the TATA box, reminiscent of enhancer elements. Although no functional assays have been performed to verify their involvement in transcriptional regulation, these putative *Fugu* HSEs possess several characteristics which compare well with other species. First, they occur within 250 bases upstream of the start codon, the least value being 146 bases, which is not significantly closer to the start codon in this compact genome. Second, they occur in multiple copies. Third, the number of bases between GAA and TTC varied between 2 and 65.

The translated amino acid sequence of *Fugu HSP70-4* is more similar to the MHC-linked bovine, porcine and murine orthologues of human cytosolic *HSP70-1* than other *HSP70* genes, cytosolic and noncytosolic, elsewhere in these genomes. There is about the same level of similarity (82.2%) with the translated amino acid sequence of *Xenopus HSP70*, which is also MHC-linked [51, 52]. Much less similarity is observed with *D. melanogaster* (71.3%), *L. amazonensis* (70.2%) and *T. cruzi* (69.0%). Not surprisingly, the highest amino acid identity is with other fish, for example salmon (88.2%). Whether the *HSP70* sequences in these fish are linked to their MHC regions is not yet known.

Based on the high level of coding and noncoding sequence similarity amongst themselves, their tightly clustered pattern and the absence of *Fugu* orthologues of mammalian MHC genes in their vicinity, it is possible that these five *Fugu HSP70* genes are the result of duplications which occurred during the evolution of the percomorphs, after the divergence of *Fugu* (fig. 4). These events could have taken place independent of the duplications of the *HSP70* which are MHC-linked in mammals.

These data lend support to the postulation that a fragmented gene pattern existed prior to convergence in the ancestral mammalian MHC region [23]. They support the hypothesis that the existence of the mammalian MHC class III region was brought about by chromosomal translocations during the process of converging the class I and II loci, which had probably arisen separately [27]. The convergence of mammalian *HSP70* genes and antigen-presenting genes probably occurred after fish had diverged.

Acknowledgments. This work was supported by a Wellcome Trust studentship and an Overseas Research studentship awarded to E. H. Lim.

1 Lindquist S. (1986) The heat-shock response. Annu. Rev. Biochem. **55:** 1151–1191
2 Pelham H. R. B. (1986) Speculations on the functions of the major heat shock and glucose-regulated proteins. Cell **46:** 959–961
3 Boorstein W. R., Ziegelhoffer T. and Craig E. A. (1994) Molecular evolution of the HSP70 multigene. J. Mol. Evol. **38:** 1–17
4 Feige U. and Polla B. S. (1994) Hsp70 – a multi-gene, multi-structure, multi-function family with potential clinical applications. Experientia **50:** 979–986
5 Rassow J., von Ahsen O., Bomer U. and Pfanner N. (1997) Molecular chaperones: towards a characterization of the heat-shock protein 70 family. Trends Cell Biol. **7:** 129–133
6 Gunther E. and Walter L. (1994) Genetic aspects of the hsp70 multigene family in vertebrates. Experientia **50:** 987–1001
7 Cyr D. M., Langer T. and Douglas M. G. (1994) DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. Trends Biochem. Sci. **19:** 176–181
8 Craig E. A., Gambill B. D. and Nelson R. J. (1993) Heat shock proteins: molecular chaperones of protein biogenesis. Microbiol. Rev. **57:** 402–414
9 Gao B., Biosca J., Craig E. A., Greene L. E. and Eisenberg E. (1991) Uncoating of coated vesicles by yeast hsp70 proteins. J. Biol. Chem. **266:** 19565–19571
10 Sargent C. A., Dunham I., Trowsdale J. and Campbell R. D. (1989) Human major histocompatibility complex contains genes for the major heat shock protein HSP70. Proc. Natl. Acad. Sci. USA **86:** 1968–1972
11 Milner C. M. and Campbell R. D. (1990) Structure and expression of the three MHC-linked HSP70 genes. Immunogenetics **32:** 242–251
12 Hunt C. R., Gasser D. L., Chaplin D. D., Pierce J. C. and Kozak C. A. (1993) Chromosomal localization of five murine HSP70 gene family members: *Hsp70-1*, *Hsp70-2*, *Hsp70-3*, *Hsc70t and Grp78*. Genomics **16:** 193–198
13 Wurst W., Benesch C., Drabent B., Rothermel E., Benecke B. J. and Gunther E. (1989) Localization of heat shock protein 70 genes inside the rat major histocompatubulity complex close to class III genes. Immunogenetics **30:** 46–49
14 Nunes M., Peelman L., Vaiman M., Bourgeaux N. and Chardon P. (1994) Characterization of six new loci within the swine major histocompatibility complex class III region. Mamm. Genome **5:** 616–622
15 Grosz M. D., Womack J. E. and Skow L. C. (1992) Syntenic conservation of HSP70 genes in cattle and humans. Genomics **14:** 863–868
16 Cameron P. U., Tabarias H. A., Pulendran B., Robinson W. and Dawkins R. L. (1990) Conservation of the central MHC genome: PFGE mapping and RFLP analysis of complement, *HSP70* and *TNF* genes in the goat. Immunogenetics **31:** 253–264
17 Salter-Cid L., Kasahara M. and Flajnik M. F. (1994) *Hsp70* genes are linked to the *Xenopus* major histocompatibility complex. Immunogenetics **39:** 1–7
18 Goate A. M., Cooper D. N., Hall C., Leung T. K. C., Solomon E. and Lim L. (1987) Localization of a human heat-shock HSP70 gene sequence to chromosome 6 and detection of two other loci by somatic-cell hybrid and restriction fragment length polymorphism analysis. Hum. Genet. **75:** 123–128

19 Leung T. K. C., Hall C., Rajendran M., Spurr N. and Lim L. (1992) The human heat-shock genes HSPA6 and HSPA7 are both expressed and localize to chromosome 1. Genomics **12:** 74–79
20 Requena J. M., Lopez M. C., Jimenez-Ruiz A., de la Torre J. C. and Alonso C. (1988) A head-to-tail tandem organization of hsp70 genes in *Trypanosoma cruzi*. Nucleic Acids Res. **16:** 1393–1406
21 Bock J. H. and Langer P. J. (1993) Sequence and genomic organization of the *hsp70* genes of *Leishmania amozonensis*. Mol. Biochem. Parasitol. **62:** 187–197
22 Ingolia T. D., Slater M. R. and Craig E. A. (1982) *Saccharomyces cerevisiae* contains a complex multigene family related to the major heat shock-inducible gene of *Drosophila*. Mol. Cell. Biol. **2:** 1388–1398
23 Klein J. and Sato A. (1998) Birth of the major histocompatibility complex. Scand. J. Immunol. **47:** 199–209
24 Campbell R. D. and Trowsdale J. (1993) Map of the human MHC. Immunol. Today **14:** 349–352
25 Chouchane L. and Kindt T. J. (1992) Mapping of the rabbit MHC reveals that class I genes are adjacent to the DR subregion and defines in insertion/deletion-related polymorphism in the class II region. J. Immunol. **149:** 1216–1222
26 Guillemot F., Kaufman J. F., Skjoedt K. and Auffray C. (1989) The major histocompatibility complex in the chicken. Trends Genet. **5:** 300–304
27 Bingulac-Popovic J., Figueroa F., Sato A., Postlethwait H. and 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
28 Brenner S., Elgar G., Sandford R., Macrae A., Venkatesh B. and Aparicio S. (1993) Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. Nature **366:** 265–268
29 Lim E. H. and Brenner S. (1997) Short-range linkage relationships of the valyl-tRNA synthetase gene in *Fugu rubripes*. Immunogenetics **46:** 332–336
30 Katsanis N., Fitzgibbon J. and Fisher E. M. C. (1996) Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. Genomics **35:** 101–108
31 Trachtulec Z., Hamvas R. M., Forejt J., Lehrach H. R., Vincek V. and Klein J. (1997) Linkage of TATA-binding protein and proteasome subunit C5 genes in mice and humans reveals synteny conserved between mammals and invertebrates. Genomics **44:** 1–7
32 Church G. M. and Gilbert W. (1984) Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991–1995
33 Sambrook J., Fritsch E. F. and Maniatis T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
34 Rosenthal A. R., Coutelle O. and Craxton M. (1993) Large-scale production of DNA sequencing templates by microtitre format PCR. Nucleic Acids Res. **21:** 173–174
35 Southern E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–517
36 Thompson J. D., Higgins D. G. and Gibson T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680
37 Saitou N. and Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425
38 Felsenstein J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791
39 Delaney S. J., Hayward D. C., Barleben F., Fischbach K. F. and Miklos G. L. (1991) Molecular cloning and analysis of small optic lobes, a structural brain gene of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **88:** 7214–7218
40 Choi K. W. and Benzer S. (1994) Rotation of photoreceptor clusters in the developing *Drosophila* eye requires the nemo gene. Cell **78:** 125–136
41 Du Z., Cope M. J. T. V. and Kendrick-Jones A. (1995) EMBL accession no. P46549
42 Lim E. H. and Brenner S. (1995) Sequence analysis of Mhc class II β-like fragments in the pufferfish *Fugu rubripes*. Immunogenetics **42:** 432–433
43 Passananti C., Davies B., Ford M. and Fried M. (1987) Structure of an inverted duplication formed as a first step in a gene amplification event: implications for a model of gene amplification. EMBO J. **6:** 1697–1703
44 Goldschmidt-Clermont M. (1980) Two genes for the major heat-shock protein of *Drosophila melanogaster* arranged as an inverted repeat. Nucleic Acids Res. **8:** 235–252
45 Karch F., Torok I. and Tissieres A. (1981) Extensive regions of homology in front of the two hsp70 heat shock variant genes in *Drosophila melanogaster*. J. Mol. Biol. **148:** 219–230
46 Ingolia T. D., Craig E. A. and McCarthy B. J. (1980) Sequence of three copies of the gene for the major *Drosophila* heat shock induced protein and their flanking regions. Cell **21:** 669–679
47 Dudler R. and Travers A. A. (1984) Upstream elements necessary for optimal function of the hsp70 promoter in transformed flies. Cell **38:** 391–398
48 Torok I. and Karch F. (1980) Nucleotide sequences of heat shock activated in *Drosophila melanogaster*. I. Sequences in the regions of the 5′ and 3′ ends of the hsp70 gene in the hybrid plasmid 56H8. Nucleic Acids Res. **8:** 3105–3123
49 Udvardy A., Sumegi J., Toth E. C., Gausz J., Gyurkovics H., Schedl P. et al. (1982) Genomic organization and functional analysis of a deletion variant of the 87A7 heat shock locus of *Drosophila melanogaster*. J. Mol. Biol. **155:** 267–280
50 Pelham H. R. B. (1982) A regulatory upstream promoter element in the *Drosophila* Hsp70 heat-shock gene. Cell **30:** 517–528
51 Morimoto R. I., Tissieres A. and Georgopoulos C. (1994) The Biology of Heat Shock Proteins and Molecular Chaperones, pp. 417–455, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
52 Bienz M. (1984) *Xenopus hsp* 70 genes are constitutively expressed in injected oocytes. EMBO J. **3:** 2477–2483