

# Concerted and Nonconcerted Evolution of the Hsp70 Gene Superfamily in Two Sibling Species of Nematodes

Nikolas Nikolaidis and Masatoshi Nei

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University

We have identified the Hsp70 gene superfamily of the nematode *Caenorhabditis briggsae* and investigated the evolution of these genes in comparison with Hsp70 genes from *C. elegans*, *Drosophila*, and yeast. The Hsp70 genes are classified into three monophyletic groups according to their subcellular localization, namely, cytoplasm (CYT), endoplasmic reticulum (ER), and mitochondria (MT). The Hsp110 genes can be classified into the polyphyletic CYT group and the monophyletic ER group. The different Hsp70 and Hsp110 groups appeared to evolve following the model of divergent evolution. This model can also explain the evolution of the ER and MT genes. On the other hand, the CYT genes are divided into heat-inducible and constitutively expressed genes. The constitutively expressed genes have evolved more or less following the birth-and-death process, and the rates of gene birth and gene death are different between the two nematode species. By contrast, some heat-inducible genes show an intraspecies phylogenetic clustering. This suggests that they are subject to sequence homogenization resulting from gene conversion-like events. In addition, the heat-inducible genes show high levels of sequence conservation in both intra-species and inter-species comparisons, and in most cases, amino acid sequence similarity is higher than nucleotide sequence similarity. This indicates that purifying selection also plays an important role in maintaining high sequence similarity among paralogous Hsp70 genes. Therefore, we suggest that the CYT heat-inducible genes have been subjected to a combination of purifying selection, birth-and-death process, and gene conversion-like events.

## Introduction

Heat shock 70 proteins (Hsp70s) function in a diverse set of biochemical processes, including protein folding, protein transportation across membranes, and regulation of the heat shock response (Hartl and Hayer-Hartl 2002). Each Hsp70 molecule consists of an N-terminal adenosine triphosphatase domain (ATPase; ~400 aa), a substrate-binding domain (SBD; ~180 aa), and a C-terminal domain of variable length.

The members of the *Hsp70* gene family may be heat/stress inducible or constitutively expressed (Boorstein, Ziegelhoffer, and Craig 1994). Recently, Hsp110 proteins (Hsp110s) have been identified in a wide variety of eukaryotic organisms. They are larger and more divergent proteins but are considered members of the *Hsp70* gene superfamily on the basis of structural and functional similarities (Easton, Kaneko, and Subject 2000). Hsp70s are localized in the cytoplasm (CYT), endoplasmic reticulum (ER), and mitochondria (MT), and Hsp110s are localized in the CYT and ER. The protein sequences in different cellular compartments are considerably different, and in general orthologous genes are more similar than paralogous genes (Boorstein, Ziegelhoffer, and Craig 1994; Easton, Kaneko, and Subject 2000).

In *Drosophila* and mosquito species CYT Hsp70 genes (*hsp70s*) often exist as a pair of genes with inverted transcriptional directions (inverted gene pair). The DNA sequences of the two member genes are often identical or very similar (e.g., Leigh Brown and Ish-Horowitz 1981; Benedict, Cockburn, and Seawright 1993; Bettencourt and Feder 2001). For this reason, *hsp70s* are generally believed to be subject to concerted evolution (Smith 1974; Arnheim

1983). However, all multicellular eukaryotes, including *D. melanogaster*, have several members of CYT *hsp70s*, and the sequence similarity between them is not necessarily high, suggesting that this gene family might also be subject to a birth-and-death process of evolution (Nei, Gu, and Sitnikova 1997). Furthermore, ER and MT genes appear to evolve following the model of independent evolution (Hughes and Nei 1990) or divergent evolution (Ota and Nei 1994). For these reasons, we have decided to investigate the evolutionary relationships of the Hsp70 gene superfamily from two sibling species of nematodes, *Caenorhabditis elegans* and *C. briggsae*.

The complete genome sequence of the nematode *C. elegans* (the *C. elegans* Sequencing Consortium 1998) and the nearly complete sequence of its sibling species *C. briggsae* (Washington University Genome Sequencing Center, unpublished data) are now available, and each of these genomes contains many members of the Hsp70 gene superfamily. Therefore, it is possible to study the evolutionary mechanism of *hsp70s* by examining the evolutionary relationships of the genes from these closely related species.

For this purpose, we identified all member genes of the Hsp70 gene superfamily from the two nematodes. Fortunately, some *hsp70s* from *C. elegans* have already been characterized (reviewed in Heschl and Baillie 1990a), and the fully annotated genomic sequence of *C. elegans* is available from WormBase (Harris et al. 2003). We therefore conducted the gene search for *C. briggsae* using the annotated genes of *C. elegans*. The genes identified in this way were then subjected to phylogenetic analyses to study their evolutionary relationships. In these phylogenetic analyses we included genes from *Drosophila melanogaster* and *Saccharomyces cerevisiae* to clarify the long-term evolution of *hsp70s*.

## Materials and Methods

### Sources of Sequence Data

The *C. elegans* genome, release WS102, has 13 genes containing the Hsp70 domain (www.WormBase.org,

Key words: divergent evolution, birth-and-death, purifying selection, gene conversion.

E-mail: nxn7@psu.edu.

*Mol. Biol. Evol.* 21(3):498–505, 2004

DOI: 10.1093/molbev/msh041

Advance Access publication December 23, 2003

*Molecular Biology and Evolution* vol. 21 no. 3

© Society for Molecular Biology and Evolution 2004; all rights reserved.

**Table 1**  
**Orthologous Relationships of the Hsp70 and Hsp110 Genes in *C. elegans* and *C. briggsae***

<i>C. elegans</i>			<i>C. briggsae</i>			Some Characteristics		
Clone	Gene	Codons	Clone	Gene	Codons	Identity (nt/aa)	Family	Localization
F26D10	<i>hsp70-1</i>	640	Fpc0010	<i>hsp70-1</i>	639	88/98	Hsp70	CYT
K09C4	<i>hsp70-2<math>\psi</math></i>	310	—	—	—	—	Hsp70	CYT
C15H9	<i>hsp70-3</i>	661	Fpc3857	<i>hsp70-3</i>	661	92/98	Hsp70	ER
F43E2	<i>hsp70-4</i>	658	Fpc2969	<i>hsp70-4</i>	658	82/94	Hsp70	ER
C37H5	<i>hsp70-6</i>	657	Fpc2114	<i>hsp70-6</i>	658	88/98	Hsp70	MT
F44E5	<i>hsp70-7*</i>	645	Fpc2976	<i>hsp70-7</i>	644	86/92	Hsp70	CYT
F44E5	<i>hsp70-8*</i>	645	Fpc2976	<i>hsp70-8</i>	629	86/94	Hsp70	CYT
C12C8	<i>hsp70-9*</i>	643	Fpc4321	<i>hsp70-9</i>	646	76/90	Hsp70	CYT
F11F1	<i>hsp70-10</i>	607	Fpc0305	<i>hsp70-10</i>	607	77/86	Hsp70	CYT
F54C9	<i>hsp70-11</i>	469	Fpc0011	<i>hsp70-11</i>	381	80/75	Hsp70	CYT
—	<i>hsp70-7</i>	—	Fpc2976	<i>hsp70-12</i>	60	87/96	Hsp70	CYT
—	<i>hsp70-7</i>	—	Fpc2976	<i>hsp70-13</i>	235	87/98	Hsp70	CYT
—	<i>hsp70-7</i>	—	Fpc4079	<i>hsp70-14</i>	644	86/94	Hsp70	CYT
—	<i>hsp70-7</i>	—	Fpc4079	<i>hsp70-15</i>	126	81/84	Hsp70	CYT
—	<i>hsp70-9</i>	—	Fpc4127	<i>hsp70-16</i>	297	77/57	Hsp70	CYT
—	<i>hsp70-9</i>	—	Fpc4127	<i>hsp70-17</i>	484	74/68	Hsp70	CYT
—	<i>hsp70-9</i>	—	Fpc4127	<i>hsp70-18</i>	81	76/77	Hsp70	CYT
—	<i>hsp70-9</i>	—	Fpc4127	<i>hsp70-19</i>	274	76/75	Hsp70	CYT
—	<i>hsp70-9</i>	—	Fpc4127	<i>hsp70-20</i>	597	68/81	Hsp70	CYT
—	<i>hsp70-7</i>	—	Fpc4044	<i>hsp70-21</i>	132	75/81	Hsp70	CYT
C30C11	<i>hsp110-1</i>	776	Fpc4079	<i>hsp110-1</i>	777	81/92	Hsp110	CYT
T24H7	<i>hsp110-2</i>	925	Fpc0058	<i>hsp110-2</i>	922	76/82	Hsp110	ER
T14G8	<i>hsp110-3</i>	926	Fpc0003	<i>hsp110-3</i>	907	77/85	Hsp110	ER

NOTE.—*C. briggsae* does not contain a sequence orthologous to the *C. elegans* pseudogene *hsp70-2 $\psi$*  gene (indicated by dashes). Highly heat-inducible genes are indicated by \*. Gene *hsp70-12~13* and *hsp70-15~21* do not encode complete Hsp70 proteins and are present only in *C. briggsae* (indicated by dashes). Their sequences were most similar to the genes *hsp70-7* and *hsp70-9* of *C. elegans*. Localization, the subcellular compartment in which the Hsp70 protein molecule resides. CYT, cytoplasm; ER, endoplasmic reticulum; MT, mitochondrion, nt and aa, indicate nucleotides, and amino acids, respectively (identity is given in percentage).

07/01/03; Harris et al. 2003), and five of them have been previously characterized (reviewed in Heschl and Baillie 1990a). There are 10 Hsp70 and 3 Hsp110 genes in *C. elegans* (table 1). For *C. briggsae* we used the sequence data available from the Web site genome.wustl.edu/gsc (version cb25.agp8; Washington University Genome Sequencing Center, unpublished data). This genome sequence is estimated to cover the 98% of the briggsae genome. The Hsp70 gene sequences for yeast were obtained from the *Saccharomyces* Genome Database (genome-www.stanford.edu/Saccharomyces; Goffeau et al. 1996), and those for *Drosophila* were obtained from the Berkeley *Drosophila* Genome Project (www.fruitfly.org; Adams et al. 2000).

#### Identification and Genomic Organization of *C. briggsae* *hsp70* Genes

We used the 13 gene sequences of *C. elegans* as queries for several rounds of BLAST search (tBLASTn, BLASTn; Altschul et al. 1990) against the *C. briggsae* genome sequence. We downloaded the contig sequences that contained a BLAST hit and extracted a 40-kb region surrounding each hit. We then identified the genes, including non-Hsp genes, in this region using the programs GeneMark and GenScan (Borodovsky and McIninch 1993; Burge and Karlin 1997) and any preliminary annotation from the ENSEMBL database (www.ensembl.org/Caenorhabditis\_briggsae; unpublished data). Because of the absence of dense gene maps for *C. briggsae* and the preliminary nature of the genome assembly, we made an effort to assign the *hsp70*s and

*hsp110*s of this species on the chromosomal regions that are homologous to the *C. elegans* chromosomes. For this purpose, the syntenic relationships observed between the *C. briggsae* and *C. elegans* genes were determined by examining the 40-kb genomic region surrounding the Hsp genes and other available data from the WormBase (Harris et al. 2003). Pairwise alignments were performed by using programs BLAST2seq, Spidey (www.ncbi.nlm.nih.gov) and BioEdit (www.mbio.ncsu.edu/BioEdit/bioedit.html), and promoter elements were identified with the TESS software (Transcription Element Search Software; www.cbil.upenn.edu/tess).

#### Multiple Sequence Alignments and Phylogenetic Tree Construction

Multiple sequence alignments were performed by using ClustalX v1.81 (Thompson et al. 1997). Phylogenetic trees were constructed by the Neighbor-Joining (NJ) and maximum parsimony (MP) methods, both included in MEGA2 (Kumar et al. 2001). In the NJ method we used *p*-distances to construct trees, because our main purpose was to find the branching pattern (topology) rather than to estimate branch lengths, and in this case *p*-distances are known to be generally more efficient than other distances (Nei and Kumar 2000, Pp. 17–32; Takahashi and Nei 2000). The accuracy of reconstructed trees was examined by the bootstrap test with 1,000 replications. The phylogenetic trees were rooted with the *Escherichia coli* HscC sequence (Yoshimune et al. 2002). The trees were obtained by using both DNA and amino acid sequences, but we present only those obtained by the latter sequences,

because both trees were similar and amino acid sequences usually give more reliable trees than DNA sequences for distantly related species (Nei and Kumar 2000, Pp. 17–33; 138–140).

## Results

### Hsp70 and Hsp110 Genes in Two Nematode Species

We have identified 22 genes belonging to the Hsp70 gene superfamily in the genome of *C. briggsae*. There were 19 Hsp70 genes and 3 Hsp110 genes (table 1; fig. 1). Ten of the 19 Hsp70 genes were full-length genes, and the remainder were partial sequences. Hsp110 genes were all full-length genes. As mentioned earlier, *C. elegans* has 10 Hsp70 genes and 3 Hsp110 genes. Five of the *C. elegans* *hsp70s* are designated as *hsp70-1*, *hsp70-2*, etc., in order of discovery (Heschl and Baillie 1990a, b). We named the remaining genes of *C. elegans* in a similar way (table 1; fig. 1). One of them (*hsp70-2ψ*) is known to be a truncated pseudogene. *Hsp70-5* has been reported (see Heschl and Baillie 1990a) but has not been identified in subsequent studies. The three Hsp110 genes are named *hsp110-1*, *hsp110-2*, and *hsp110-3* (table 1).

Study of sequence similarity between the genes from the two nematode species showed that *C. briggsae* contains all the genes orthologous to *C. elegans* genes except *hsp70-2ψ* (table 1; see also figs. 1 and 2 in the Supplementary Material online at [www.mbe.oupjournals.org](http://www.mbe.oupjournals.org)). We therefore used the *C. elegans* notations for these genes (table 1; fig. 1). All Hsp70 genes in the two nematode species were located in homologous chromosomal regions except genes *hsp70-7*, *-8*, and *-10* (see table 1 in the Supplementary Material online at [www.mbe.oupjournals.org](http://www.mbe.oupjournals.org)). This finding is supported by the syntenic relationships of other non-Hsp genes in the regions currently available from WormBase (Harris et al. 2003). *C. briggsae* has another full-length Hsp70 gene, which is closely related to *hsp70-7*, and we call it *hsp70-14* (table 1). The remaining nine genes are truncated genes either because they are pseudogenes containing termination codons or because sequencing is incomplete (table 1; fig. 1).

The number of introns and intron phases (0, 1, and 2) varied considerably with gene. In the case of CYT genes, these properties were generally conserved between the orthologous gene pairs of the two nematode species. There were two exceptions: One is *hsp70-8*, which has one intron in *C. elegans* and two in *C. briggsae* (fig. 1), and the other is *hsp70-11*, of which the intron/exon organization is quite different between the two nematode species. Interestingly, genes *hsp70-10* and *-11* are very different in the intron/exon organization from the other CYT genes. By contrast, genes *hsp70-7~9* and *hsp70-13~21* have their first intron in the same position with the same intron phase (0) (fig. 1). The nucleotide sequence of this intron is identical for genes *hsp70-7~8* of *C. elegans*. Similarly, the intron sequence is the same for *hsp70-7*, *-8*, *-14* and for *hsp70-12~13* of *C. briggsae*. This suggests that these genes have a recent common ancestor or experienced a gene conversion-like event.

In the cases of ER and MT genes, the intron/exon organization differs even between the orthologous genes of

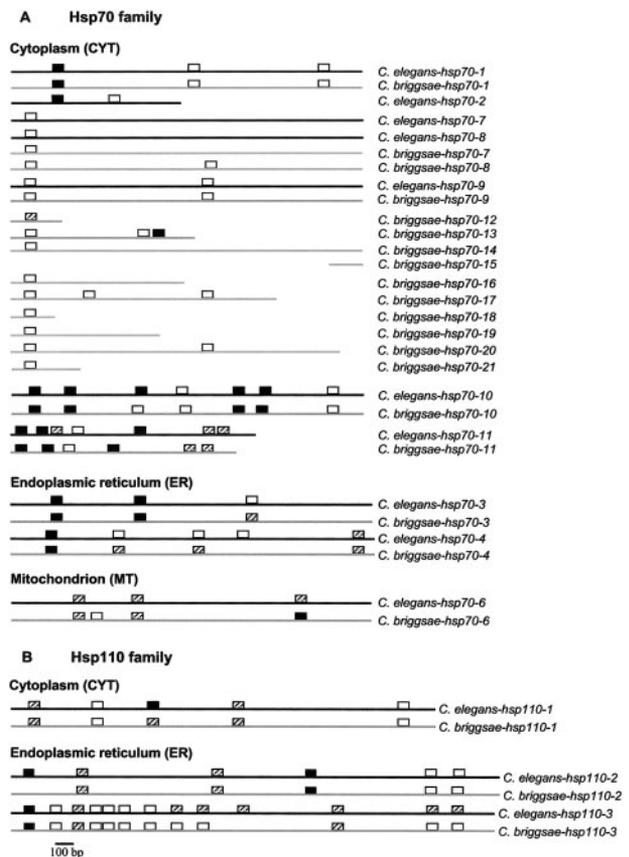


FIG. 1.—Intron positions and phases of the *hsp70* genes (A) and the *hsp110* genes (B) in *Caenorhabditis elegans* (black line) and *C. briggsae* (gray line). The *hsp70* coding regions are to scale; introns are not. An intron is of phase 0 if it lies between two codons (open box), of phase 1 if it lies between the first and second nucleotides of a codon (closed box), and of phase 2 if it lies between the second and the third nucleotides of a codon (marked boxes).

the two species. Thus, the number of exons and the intron phases of *hsp70-4* are different in the two species. Furthermore, the intron phase of the third intron of *hsp70-3* is 0 in *C. elegans* but 2 in *C. briggsae*. The intron/exon organization of the MT gene *hsp70-6* differs considerably in the two species (fig. 1).

In the Hsp110 gene family there are one CYT and two ER genes in each of the two nematode species. The CYT gene (776 codons) is considerably shorter than the ER genes (925 codons). In both groups of genes the intron/exon organization is different in the two species (fig. 1).

Hsp70 proteins are highly conserved. Therefore, one would expect that the amino acid identity of the orthologs between the two nematode species would be higher than the nucleotide sequence identity. In fact, this was the case for most gene comparisons (table 1). However, comparisons of *hsp70-9* from *C. elegans* with the *hsp70-16~18* from *C. briggsae* showed a higher identity for nucleotide sequences than for amino acid sequences, although all of the *C. briggsae* genes were fragmentary. This suggests that these fragmentary genes are pseudogenes and therefore accumulated many nucleotide substitutions leading to amino acid substitutions. Another pair of orthologous genes that showed a higher nucleotide sequence identity

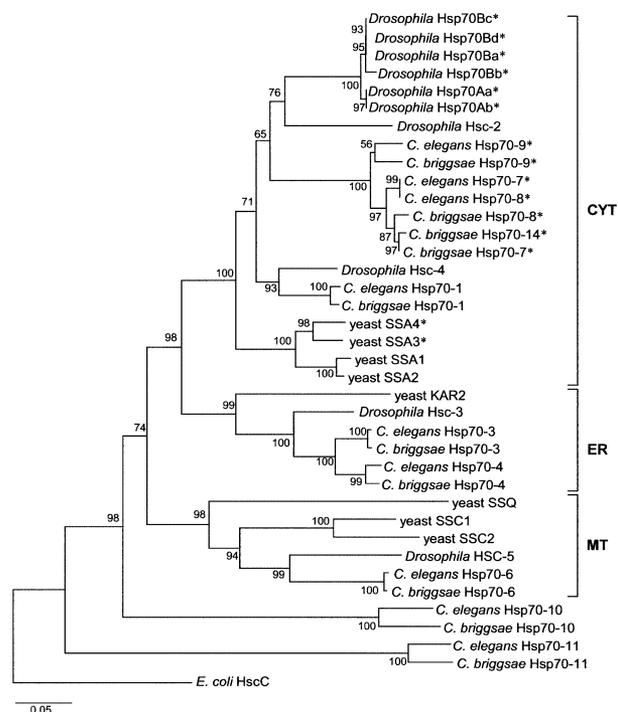


FIG. 2.—Phylogenetic relationships of the Hsp70 proteins from *Caenorhabditis elegans*, *C. briggsae*, *Saccharomyces cerevisiae* (yeast) and *Drosophila melanogaster* (*Drosophila*). The numbers for interior branches represent bootstrap values. The sites with alignment gaps were completely excluded, and the amino acid sites used to construct the tree numbered 311. The phylogenetic relationships of the Hsp110 family are presented in a separate figure (fig. 3) because the two subfamilies are very divergent. Accession numbers of the sequences used: *Escherichia coli*: HscC, P77319. Yeast: KAR2, M25064; SSA1, P10591; SSA2, P10592; SSA3, S36753; SSA4, B36590; SSC1, M27229; SSC2, S50429; SSQ1, S44545. *Drosophila*: Hsp70Aa, AAA28640; Hsp70Ab, P02825; Hsp70Ba, AAG26902; Hsp70Bb, AAD15226; Hsp70Bc, AAK30242; Hsp70Bd, AC007668; Hsc-2, P11146; Hsc-5, P29845; Hsc-4, AAB59186; Hsc-3, P29844. Highly heat-inducible genes are indicated by \*. CYT, cytoplasm; ER, endoplasmic reticulum; MT, mitochondrion.

than amino acid sequence identity is *hsp70-11*. This gene is much shorter in *C. briggsae* (381 codons) than in *C. elegans* (469 codons). Therefore, some unusual form of nucleotide substitution might have happened.

#### Phylogenetic Relationships of Hsp70 Genes

Hsp70 and Hsp110 genes are very divergent and form two distantly related clusters (see fig. 2A in the Supplementary Material online at [www.mbe.oupjournals.org](http://www.mbe.oupjournals.org); Easton, Kaneko, and Subject 2000). We therefore present the results of our phylogenetic analysis for Hsp70 genes and Hsp110 genes, separately. Figure 2 shows the phylogenetic trees for Hsp70 genes including those from *Drosophila* and yeast. This tree shows that *hsp70s* can be classified into three monophyletic groups, CYT, ER, and MT, if we exclude two genes (*hsp70-10* and *hsp70-11*) from each of the two nematode species. Therefore, the functional differentiation of the CYT, ER, and MT genes is well supported by the phylogenetic analysis. There is no experimental study of the subcellular locations of the CYT, ER, and MT genes in *C. briggsae*, but they can be

identified by the presence of signal sequences at the carboxyl (C)-terminal of the protein molecules (Heschl and Baillie 1990a, b; Boorstein, Ziegelhoffer, and Craig 1994).

The two basal genes *hsp70-10* and *hsp70-11* from each of the two nematode species are unique to nematodes, and their orthologs have not been found in yeast and *Drosophila*. However, a putative ortholog of *hsp70-11*, also known as Stch-1, has been found in fugu, mice, rats, and humans (Otterson and Kaye 1997; N.N. unpublished results). The function of Hsp70-11 has not been well studied, but this Hsp70 protein does not contain the substrate-binding domain. By contrast, no genes orthologous to *hsp70-10* have been reported in any other organism listed in GenBank. This gene might have been lost from other organisms.

The MT clade of the gene tree in figure 2 shows that each species of nematodes and *Drosophila* has a single MT gene, but yeast has three genes. These genes seem to have evolved following the model of independent (Hughes and Nei 1990) or divergent evolution (Ota and Nei 1994), in which each gene acquires a new function and its sequence diverges with time. Essentially the same evolutionary pattern is observed in the ER genes, though the two nematode species have duplicate genes. The nematode gene *hsp70-3* and the *Drosophila* gene *hsc-3* have the ER retention amino acid motif KDEL and are not heat-inducible, whereas the nematode *hsp70-4* and the yeast *KAR-2* have the amino acid motif HDEL and are heat-inducible (Heschl and Baillie 1990a, b). All vertebrate species so far studied have one ER homolog with the KDEL motif (Kaufmann 1999). Most of the MT and CYT genes are expressed constitutively, but some are moderately heat-inducible (Heschl and Baillie 1990a). However, these heat-inducible genes do not show a pattern of concerted evolution, unlike the CYT genes discussed below.

The evolutionary pattern of CYT genes is somewhat complicated. These genes can be divided into heat-inducible (marked with \*) and constitutively expressed genes (some with moderate heat inducibility; Kim et al. 2001; fig. 2; see also table 2 in the Supplementary Material online at [www.mbe.oupjournals.org](http://www.mbe.oupjournals.org)). The phylogenetic tree of constitutively expressed genes from the two species of nematodes, *Drosophila*, and yeast is generally consistent with the species tree and suggests that these genes are subject to divergent evolution. However, the heat-inducible genes from the nematodes form a tight monophyletic clade and so do the *Drosophila* heat-inducible genes. In particular, the protein sequence of *hsp70-7* is identical with that of *hsp70-8* in *C. elegans*, as is the nucleotide sequence (fig. 2). It is interesting to note that *hsp70-7* and *hsp70-8* exist in the genome as an inverted pair of genes with different transcriptional directions (fig. 3). The *C. briggsae* genes *hsp70-7* and *hsp70-8* also exist as an inverted gene pair, but they are slightly different in amino acid and nucleotide sequences. Interestingly, these gene pairs exist in close neighborhood with the *hsp70-12*~*hsp70-13* gene pair (fig. 3). By contrast, *hsp70-9* exists as a singleton in both *C. elegans* and *C. briggsae*. These genes make a single clade in the tree in figure 2, even though their amino acid sequences are considerably different.

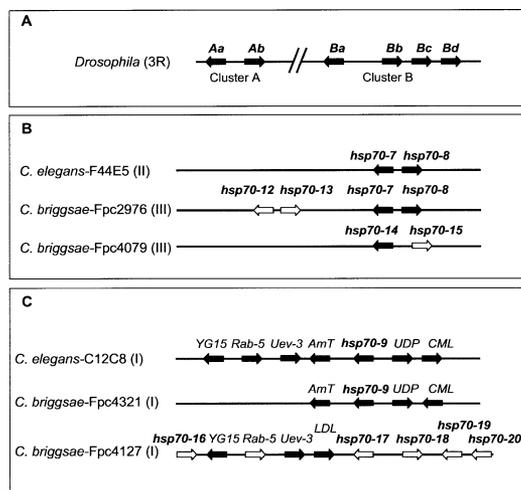


FIG. 3.—A. Genomic organization of the *hsp70* genes at bands 87A (genomic cluster A) and 87C (genomic cluster B) in *Drosophila melanogaster* (*Drosophila*) (only *hsp70* genes are shown). B. Genomic organization of the genomic clone F44E5 (chromosome II) in *Caenorhabditis elegans*, and of clones Fpc2976 and Fpc4079 (chromosome III) in *C. briggsae* (only *hsp70* genes are shown). C. Genomic organization of the genomic clone C12C8 (chromosome I) in *C. elegans*, and of clones Fpc4231 and Fpc4127 (chromosome I) in *C. briggsae*. Complete genes are indicated by closed arrows and incomplete genes by open arrows (arrowheads show the transcription orientation). Drawing is not to scale. Chromosome number is given in parentheses (Roman letters). Gene *hsp70-9* in both nematode species is encoded by the complementary strand of a large intron of the UDP gene. YG-15, shows similarity with the yeast hypothetical gene YG-15; Rab-5, a member of the RAS GTPase superfamily; Uev-3, ubiquitin E2 (conjugating enzyme); AmT, phosphoserine amino transferase (also known as PSAT); LDL, low-density lipoprotein receptor; UDP, UDP-glucose-glycoprotein glycosyltransferase; CML, cystathionine beta-lyase/cystathionine gamma-synthase.

As mentioned above, *C. briggsae* has several fragmentary genes, which are probably pseudogenes but could be parts of functional genes (table 1; fig. 3; see also fig. 1 in the Supplementary Material online at [www.mbe.oupjournals.org](http://www.mbe.oupjournals.org)). We therefore constructed another tree for CYT genes including the genes with relatively long sequences (see fig. 2B in the Supplementary Material online). This tree was constructed by the pairwise deletion option of MEGA2. In this tree genes *hsp70-7* and *hsp70-13* in *C. briggsae* have identical amino acid sequence (and identical nucleotide sequence as well) and are closely related to *hsp70-14*. Note that *hsp70-14* belongs to a genomic cluster of Hsp70 genes that is separated from the genomic cluster of *hsp70-7~8* by almost 2 Mb. *C. briggsae* has another genomic cluster (see fig. 3C). This cluster has five *hsp70*s, but two of them (*hsp70-16* and *-18*) were not included in the tree because they were too short. The tree in figure 2B of the online Supplementary Material shows that these genes form a separate clade and evolved relatively fast except the *hsp70-19*.

#### Phylogenetic Relationships of Hsp110 Genes

Figure 4 shows the phylogenetic tree of the Hsp110 genes. As mentioned earlier, Hsp110 genes can be classified into two functionally different groups: ER and CYT genes. As in the case of Hsp70 genes, these two groups of genes

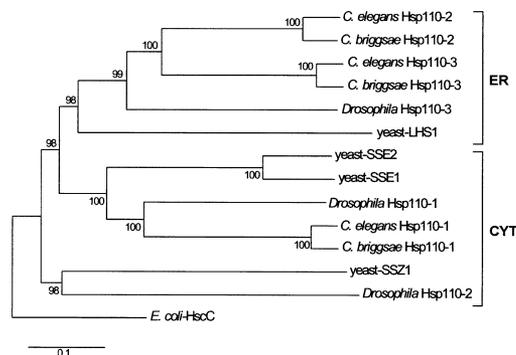


FIG. 4.—Phylogenetic relationships of the Hsp110 proteins from *Caenorhabditis elegans*, *C. briggsae*, *Saccharomyces cerevisiae* (yeast) and *Drosophila melanogaster* (*Drosophila*). The numbers for interior branches represent bootstrap values. The alignment gaps were completely excluded from the phylogenetic analysis, and the amino acids used to construct the tree were 418. Accession numbers of the sequences used: *Escherichia coli*: HscC, P77319. Yeast: LHS1, S37895; SSE1, P32589; SSE2, P32590; SSZ1 (PDR13), S46712. *Drosophila*: HSP110-1, CAB38171; HSP110-2, Q96NG7; HSP110-3, QVUC1. CYT, cytoplasm; ER, endoplasmic reticulum.

diverged long before the separation of animals and fungi. The ER genes are monophyletic. Yeast and *Drosophila* have only one gene, but *C. elegans* and *C. briggsae* each have two (*hsp110-2* and *-3*), both of which were generated by gene duplication after nematodes and *Drosophila* diverged. The CYT genes are polyphyletic, because the *Drosophila hsp110-2* and the yeast SSZ1 separated from the other CYT and ER genes before the latter two groups diverged. If we exclude yeast SSZ1 and *Drosophila hsp110-2*, the rest of the CYT genes as well as the ER genes diverged following the species tree. Therefore, these genes experienced the model of divergent evolution (Ota and Nei 1994) rather than concerted evolution. Among all metazoan species so far studied, *C. elegans* and *C. briggsae* are exceptional in that they have only one CYT gene (Hsp110) and two ER genes. Yeast, *Drosophila*, and all animals so far studied have at least two CYT and only one ER gene (Easton, Kaneko, and Subject 2000).

#### Discussion

We have seen that Hsp70 genes in nematodes can be divided into CYT, ER, and MT groups and that each of these groups of genes forms a monophyletic cluster. There are two additional genes, *hsp70-10* and *hsp70-11*, which diverged from the other Hsp70 genes much earlier than the divergence of animals and fungi. As mentioned earlier, *hsp70-11* appears to be present in many vertebrate species, but its function is not well understood. By contrast, gene *hsp70-10* has been identified only in nematodes, and its orthologous gene seems to have been lost from other animals so far studied.

We have also seen that the non-heat-inducible genes in the ER and MT groups have evolved following the model of divergent evolution. The CYT constitutively expressed Hsp70 genes probably evolved under the birth-and-death model of evolution, and the rates of gene birth and death are different between *C. elegans* and *C. briggsae*. However, many CYT heat-inducible genes show

**Table 2**  
**Numbers of Synonymous ( $p_S$ ) (below diagonal) and Nonsynonymous ( $p_N$ ) Differences per Site (above diagonal) for Comparisons of Full-Length Heat-Inducible Hsp70 Genes from *C. elegans* and *C. briggsae***

	<i>C. elegans</i>			<i>C. briggsae</i>			
	<i>hsp70-7</i>	<i>hsp70-8</i>	<i>hsp70-9</i>	<i>hsp70-7</i>	<i>hsp70-8</i>	<i>hsp70-9</i>	<i>hsp70-14</i>
<i>C. elegans</i>							
<i>hsp70-7</i>		0	7.8	3.4	4.1	10.6	3.6
<i>hsp70-8</i>	0		8.0	3.4	4.1	10.6	3.6
<i>hsp70-9</i>	53.2	53.2		7.8	8.3	8.8	8.0
<i>C. briggsae</i>							
<i>hsp70-7</i>	42.5	42.5	61.7		1.2	9.9	0.4
<i>hsp70-8</i>	41	41	61.5	5.3		10	1.5
<i>hsp70-9</i>	70.4	70.4	73.4	68.4	68.8		10.2
<i>hsp70-14</i>	42.6	42.6	61.8	0	5.3	68.5	

NOTE.—All values are multiplied by 100.

a high degree of amino acid sequence similarity, and also in some inverted gene pairs the two copies show identical amino acid sequence.

This high similarity in amino acid sequence can be explained either by purifying selection or by gene conversion (Arnheim 1983). A simple way to distinguish between the two hypotheses is to compare the number of synonymous ( $p_S$ ) and nonsynonymous ( $p_N$ ) nucleotide differences per site for all heat-inducible genes from the two nematode species. If purifying selection is the main factor, one would expect that  $p_S$  will be greater than  $p_N$ . By contrast, if gene conversion is the main factor, both synonymous and nonsynonymous sites are homogenized, and therefore  $p_S$  will be nearly the same as  $p_N$ . Using this type of approach, Nei, Rogozin, and Piontkivska (2000), Rooney, Piontkivska, and Nei (2002), and Piontkivska, Rooney, and Nei (2002) showed that the sequence homogeneity among paralogs of ubiquitins and histones H3 and H4 is caused primarily by purifying selection. In this study we used the modified version of the Nei-Gojobori method as included in MEGA2 to calculate  $p_S$  and  $p_N$ .

The results obtained are presented in table 2, and they show that  $p_S$  is generally much higher than  $p_N$ , although there are a few exceptions. Intraspecific comparisons show that gene *hsp70-9* is very close to *hsp70-7* and *hsp70-8* in terms of nonsynonymous differences, but they are very different in terms of synonymous differences in both *C. elegans* and *C. briggsae*. Interspecific comparison of genes between *C. elegans* and *C. briggsae* also indicates that  $p_S$  is about 10 times greater than  $p_N$ . This indicates that purifying selection plays an important role in maintaining the sequence homogeneity among paralogous Hsp70 genes.  $p_S$  is also higher than  $p_N$  when the *Drosophila hsp70s* of cluster A are compared with those of cluster B (data not shown), suggesting that purifying selection plays an important role in the evolution of these genes in *Drosophila* species as well.

However, *hsp70-7* and *hsp70-8* in *C. elegans* have identical amino acid and nucleotide sequences. Similarly, there is not much difference between  $p_S$  and  $p_N$  for the gene pairs *hsp70-7* and *hsp70-8*, *hsp70-7* and *hsp70-14*, and *hsp70-8* and *hsp70-14* in *C. briggsae*. These results appear to be due either to gene conversion or to recent

gene duplication. We have already mentioned that genes *hsp70-7* and *hsp70-8* form an inverted gene pair in the two nematode species. In both gene pairs  $p_S$  is nearly equal to  $p_N$ . Similar inverted and tandem gene pairs with small  $p_S$  and  $p_N$  have also been observed in *Drosophila* (Konstantopoulou, Nikolaidis, and Scouras 1998; Bettencourt and Feder 2001, 2002), mosquito (Benedict, Cockburn, and Seawright 1993), fugu (Lim and Brenner 1999), rat (Walter, Rauh, and Günther 1994), and human (Tavaria et al. 1996). Therefore, sequence homogenization is likely due to some kind of gene conversion. It appears that both purifying selection and gene conversion are involved in the evolution of CYT heat-inducible Hsp70 genes.

However, some CYT heat-inducible genes appear to be subject to even birth-and-death evolution. In figure 2 the *C. briggsae* gene *hsp70-9* clusters with the *C. elegans hsp70-9* rather than with other *C. briggsae* genes. If gene conversion is the major mode of evolution, one would not expect this type of clustering, because the two nematode species, despite their morphological similarity, appear to have diverged about 20–40 mya (Thacker et al. 1999) or even more (Coghlan and Wolfe 2002). Furthermore, *C. briggsae* has many duplicate genes compared with *C. elegans*, and several of them are apparently pseudogenes, as mentioned above. For example, the gene pairs *hsp70-12*~*hsp70-13* and *hsp70-14*~*hsp70-15* appear to be recent duplications of the gene pair *hsp70-7*~*hsp70-8*, because they have high sequence similarity (see figs. 1 and 2 in the Supplementary Material online at [www.mbe.oupjournals.org](http://www.mbe.oupjournals.org)). However, only gene *hsp70-14*, appears to be functional now. These genes are apparently the result of block duplications, unlike their orthologs in *Drosophila* in which the additional copies are probably the result of transposition (Bettencourt and Feder 2001). The *C. briggsae* Fpc4127 genomic region has five Hsp70 genes, and the homologous segments of the genes have high sequence similarity with one another. This suggests that they are recent duplicate genes, but all of them appear to be pseudogenes. If this interpretation is correct, the birth-and-death model of evolution also plays a significant role in the evolution of CYT Hsp70 genes.

Nevertheless, the hallmark of CYT *hsp70s* molecular evolution is the concept of gene conversion, mainly because

of the high degree of nucleotide sequence similarity of two tandem or inverted copies. The number of tandemly arranged copies is usually two (Leigh Brown and Ish-Horowitz 1981; Bettencourt and Feder 2001), and if this number increases, some of the genes tend to diverge or, sometimes, to become inactivated, as in the case of *C. briggsae* genes. *D. mauritiana* has two inverted gene pairs, and each one has pseudogenes segregating (Bettencourt and Feder 2001, 2002).

The molecular mechanism of gene conversion has been unknown for a long time. It is now believed that gene conversion is a repairing mechanism of DNA breaks that occur at the time of replication or transcription, and this DNA repair results in homogenization of two sequences (Haber 1999; Saxe, Datta, and Jinks-Robertson 2000). However, how the sequence identity or virtual identity between inverted or tandem gene pairs is achieved by gene conversion is still mysterious. How does a single gene conversion event homogenize the entire gene region of about 2–3 kb, including introns? If this homogenization is caused by many conversion events for short gene fragments, how often is the sequence identity for the entire gene region achieved? Why does gene conversion occur only in heat-inducible CYT Hsp70 genes? Is it because the transcription machinery promotes DNA breaks (Saxe, Datta, and Jinks-Robertson 2000; González-Barrera, García-Rubio, and Aquilera 2002), and stimulates conversion events? These are difficult questions to answer at this moment. It is of paramount importance to clarify the molecular mechanism of gene conversion in order to understand the mechanism of concerted evolution.

### Acknowledgments

The Sanger Institute and the Genome Sequencing Center, Washington University, St. Louis, are gratefully acknowledged for making the *C. briggsae* sequence available. We also thank Zacharias G. Scouras, Wojciech Makalowski, and Dimitra Chalkia for their valuable comments and discussions. This work was supported by a grant from the National Institutes of Health (GM20293) to M.N.

### Literature Cited

Adams, M. D., S. E. Celniker, R. A. Holt et al. (192 co-authors). 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**:2185–2195.

Altschul, S. F., W. Gish, W. Schäffer, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.

Arnheim, N. 1983. Concerted evolution of multigene families. Pp. 38–61 in M. Nei and R. K. Koehn, eds. *Evolution of genes and proteins*. Sinauer Associates, Sunderland, Mass.

Benedict, M. Q., A. F. Cockburn, and J. A. Seawright. 1993. The Hsp70 heat-shock gene family of the mosquito *Anopheles albimanus*. *Insect Mol. Biol.* **2**:93–102.

Bettencourt, B. R., and M. E. Feder. 2001. *Hsp70* duplication in the *Drosophila melanogaster* species group: how and when did two become five? *Mol. Biol. Evol.* **18**:1272–1282.

———. 2002. Rapid concerted evolution via gene conversion at the *Drosophila hsp70* genes. *J. Mol. Evol.* **54**:569–586.

Boorstein, W. R., T. Ziegelhoffer, and E. A. Craig. 1994. Molecular evolution of the Hsp70 multigene family. *J. Mol. Evol.* **38**:1–17.

Borodovsky, M., and J. McIninch. 1993. Recognition of genes in DNA sequence with ambiguities. *Biosystems* **30**:161–171.

Burge, C., and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**:78–94.

Coghlan, A., and K. H. Wolfe. 2002. Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genome Res.* **16**:857–867.

Easton, D. P., Y. Kaneko, and J. R. Subjeck. 2000. The Hsp110 and Grp1 70 stress proteins: newly recognized relatives of the Hsp70s. *Cell Stress Chaperones* **5**:276–290.

Goffeau, A., B. G. Barrell, H. Bussey et al. (13 co-authors). 1996. Life with 6000 genes. *Science* **274**:546, 563–567.

González-Barrera, S., M. García-Rubio, and A. Aquilera. 2002. Transcription and double-strand breaks induce similar mitotic recombination events in *Saccharomyces cerevisiae*. *Genetics* **162**:603–614.

Haber, J. E. 1999. DNA recombination: the replication connection. *Trends Biochem. Sci.* **24**:271–275.

Harris, T. W., R. Lee, E. Schwarz et al. (21 co-authors). 2003. WormBase: a cross-species database for comparative genomics. *Nucleic Acids Res.* **31**:133–137.

Hartl, F. U., and M. Hayer-Hartl. 2002. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**:1852–1858.

Heschl, M. F., and D. L. Baillie. 1990a. The Hsp70 multigene family of *Caenorhabditis elegans*. *Comp. Biochem. Physiol. B* **96**:633–637.

———. 1990b. Functional elements and domains inferred from sequence comparisons of a heat shock gene in two nematodes. *J. Mol. Evol.* **31**:3–9.

Hughes, A. L., and M. Nei. 1990. Evolutionary relationships of class II major-histocompatibility-complex genes in mammals. *Mol. Biol. Evol.* **7**:491–514.

Kaufmann, R. J. 1999. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. *Genes Dev.* **13**:1211–1233.

Kim, S. K., J. Lund, M. Kiraly, K. Duke, M. Jiang, J. M. Stuart, A. Eizinger, B. N. Wylie, and G. S. Davidson. 2001. A gene expression map for *Caenorhabditis elegans*. *Science* **293**:2087–2092.

Konstantopoulou, I., N. Nikolaidis, and Z. G. Scouras. 1998. The *hsp70* locus of *Drosophila auraria* (*montium* subgroup) is single and contains copies in a conserved arrangement. *Chromosoma* **107**:577–586.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**:1244–1245.

Leigh Brown, A. J., and D. Ish-Horowitz. 1981. Evolution of the 87A and 87C heat-shock loci in *Drosophila*. *Nature* **290**:677–682.

Lim, E. H., and S. Brenner. 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.

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

Nei, M., and S. Kumar. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York.

Nei, M., I. B. Rogozin, and H. Piontkivska. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc. Natl. Acad. Sci. USA* **97**:10866–10871.

- Ota, T., and M. Nei. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol. Biol. Evol.* **11**:469–482.
- Otterson, G. A., and F. J. Kaye. 1997. A “core ATPase,” Hsp70-like structure is conserved in human, rat, and *C. elegans* STCH proteins. *Gene* **199**:287–292.
- Piontkivska, H., A. P. Rooney, and M. Nei. 2002. Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol. Biol. Evol.* **19**:689–697.
- Rooney, A. P., H. Piontkivska, and M. Nei. 2002. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Mol. Biol. Evol.* **19**:68–75.
- Saxe, D., A. Datta, and S. Jinks-Robertson. 2000. Stimulation of mitotic recombination events by high levels of RNA polymerase II transcription in yeast. *Mol. Cell. Biol.* **20**:5404–5414.
- Smith, G. P. 1974. Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symp. Quant. Biol.* **38**:507–513.
- Takahashi, K., and M. Nei. 2000. Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Mol. Biol. Evol.* **17**:1251–1258.
- Tavaria, M., T. Gabriele, I. Kola, and R. L. Anderson. 1996. A hitchhikers guide to the human Hsp70 family. *Cell Stress Chaperones* **1**:23–28.
- Thacker, C., M. A. Marra, A. Jones, D. L. Baillie, and A. M. Rose. 1999. Functional genomics in *Caenorhabditis elegans*: An approach involving comparisons of sequences from related nematodes. *Genome Res.* **9**:348–359.
- The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**:2012–2018.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTALX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
- Walter, L., F. Rauh, and E. Günther. 1994. Comparative analysis of the three major histocompatibility complex-linked heat shock protein 70 (Hsp70) genes of the rat. *Immunogenetics* **40**:325–330.
- Yoshimune, K., T. Yoshimura, T. Nakayama, T. Nishino, and N. Esaki. 2002. Hsc62, Hsc56, and GrpE, the third Hsp70 chaperone system of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **293**:1389–1395.

Diethard Tautz, Associate Editor

Accepted October 13, 2003