Hsp70 sequences indicate that choanoflagellates are closely related to animals

Elizabeth A. Snell*, Rebecca F. Furlong* and Peter W.H. Holland

Over 130 years ago, James-Clark [1, 2] noted a remarkable structural similarity between the feeding cells of sponges (choanocytes) and a group of free-living protists, the choanoflagellates. Both cell types possess a single flagellum surrounded by a collar of fine tentacles [3]. The similarity led to the hypothesis that sponges, and, by implication, other animals, evolved from choanoflagellate-like ancestors. Phylogenetic analysis of ribosomal DNA neither supports nor refutes this hypothesis [4–6]. Here, we report the sequence of an *hsp70* gene and pseudogene from the freshwater choanoflagellate *Monosiga ovata*. These represent the first nuclear-encoded protein-coding sequences reported for any choanoflagellate. We find that *Monosiga* and most bilaterian *hsp70* genes have high GC contents that may distort phylogenetic tree construction; therefore, protein sequences were used for phylogenetic reconstruction. Our analyses indicate that *Monosiga* is more closely related to animals than to fungi. We infer that animals and at least some choanoflagellates are part of a clade that excludes the fungi. This is consistent with the origin of animals from a choanoflagellate-like ancestor.

Address: School of Animal and Microbial Sciences, The University of Reading, Whiteknights PO Box 228, Reading RG6 6AJ, United Kingdom.

Correspondence: Peter W. H. Holland
E-mail: p.w.h.holland@reading.ac.uk

*These authors contributed equally to this work.

Received: **30 October 2000**
Revised: **27 April 2001**
Accepted: **27 April 2001**

Published: **26 June 2001**

Current Biology 2001, 11:967–970

0960-9822/01/$ – see front matter

Results and discussion

Choanoflagellate *hsp70* gene and pseudogene

The highly conserved gene-encoding cytoplasmic Hsp70 has been sequenced from many eukaryotic taxa and has been used for phylogenetic studies at deep nodes within the Metazoa [7]. Genomic DNA was extracted from the freshwater choanoflagellate *Monosiga ovata* and amplified using nested degenerate PCR. We obtained a single clear band of 1.4 kb, the predicted size if the *hsp70* gene contains no intron between the priming sites. After cloning and sequencing, we identified two distinct products. The first has the potential to encode a protein similar to the nuclear-encoded Hsp70 proteins of other eukaryotes and dissimilar to Hsp70 proteins of organelles and bacteria. We conclude that this represents the nuclear *hsp70* gene of *Monosiga ovata* (Figure 1).

The nucleotide sequence of the second amplified product also matched *hsp70* but did not possess a complete open reading frame throughout the sequence. Alignment with known *hsp70* genes revealed that there is a single nucleotide deletion in the sequence, resulting in a frame shift. When corrected in an alignment, a deduced protein sequence related to the first *M. ovata hsp70* sequence was obtained (Figure 1). The two *M. ovata hsp70* sequences differ at six nucleotide sites out of 1.4 kb, in addition to the deletion. A high proportion of the substitutions are nonsynonymous (four out of six), counter to the pattern observed for substitutions within functional genes. Taken together, these data argue that the second clone represents an *hsp70* pseudogene in the *Monosiga ovata* genome, produced recently in evolution from the functional *hsp70* gene. All subsequent analyses were restricted to the deduced functional *hsp70* gene.

Nucleotide sequence analysis

Neighbor-joining (NJ) and maximum likelihood (ML) phylogenetic trees were constructed from a nucleotide sequence alignment comprising the nuclear-encoded *hsp70* genes of *Monosiga* plus four bilaterian animals, three diploblastic animals (two cnidarians, one ctenophore), six sponges, two fungi, two plants, and two alveolates. Analysis was restricted to the first and second codon positions, to overcome mutation saturation at the third position. These analyses placed the choanoflagellate and all the animals in a monophyletic group excluding fungi, plants, and alveolates (98% bootstrap support). Within this clade, however, both NJ and ML placed choanoflagellates as a sister to the bilaterian or "higher" animals, but with low bootstrap support of 24% (data not shown).

The unusual (albeit weakly supported) phylogenetic position obtained for the choanoflagellate in these analyses led us to suspect that the nucleotide sequences may be violating one or more assumptions made by the phylogenetic methods. We examined GC content in all of the *hsp70* sequences used in these analyses, since unequal GC bias violates the assumption of equal probabilities for certain classes of nucleotide substitution. The choanofla-

Figure 1

The deduced protein sequence of *Monosiga ovata* Hsp70 (top) aligned with the translation of the *M. ovata hsp70* pseudogene. Dashes indicate identical amino acids; the delta marks the site of a single base pair deletion in the pseudogene, corrected for this translation. These are partial sequences; only the sequence internal to the PCR primers is shown. The GenBank accession numbers for the *hsp70* gene and pseudogene are AF316323 and AF316324, respectively.

```
Monosiga hsp70   RTTPSYVAFTETERLIGDAALAQVAMNPTNTVFDAKRLIGRNFDDTSVQSDMKHWPFNVV 60
pseudogene       ------------------------------------------------------------ 60

Monosiga hsp70   SVNGKPKIEVEYMGAKKQFFPEEISAMVLVKMKETAEAYIGKTVKDAVVTVPAYFNDSQR 120
pseudogene       -----------------------------------------------------T------- 120

Monosiga hsp70   QATKDAGTIAGLHVMRIINEPTAAAIAYGLDKKLKGAEKNILIFDLGGGTFDVSILSIDE 180
pseudogene       --------------------Δ--------------------------------------- 180

Monosiga hsp70   GVFEVKSTAGDTHLGGEDFDNRLVNHFVEEFKRKHKKDLSSNKRALRRLRTACERAKRTL 240
pseudogene       ------------------------------------------------------------ 240

Monosiga hsp70   STATQANIEIDSLFEGVDFYTNITRARFEELCADLFRGTLDPVEKSLRDAKMSKSDIHEI 300
pseudogene       -------T-----------------------------------------------L-- 300

Monosiga hsp70   VLVGGSTRIPKVQKLLQDLFNGKELNKSINPDEAVAYGAAVQAAILSGETHEAVQDVVLL 360
pseudogene       ------------------------------------------------------------ 360

Monosiga hsp70   DVAPLSLGLETAGGVMTSLIKRNTTIPTKQTQTFSTYSDNQPGVLIQVFEGERAMTRDNN 420
pseudogene       ------------------------------------------------------------ 420

Monosiga hsp70   LLGKFELSGTPPAPRGVPQIEVTFEVDANGILNVSAVDKSTGKSNKI 467
pseudogene       ---------I---------------------------------- 467
```

gellate *hsp70* gene has an unusually high GC content (>60%), as do three of the four bilaterian sequences (see Supplementary material available with this article online). Similarly, and possibly as a result of GC bias, the choanoflagellate *hsp70* gene has an aberrantly low effective number of codons (see Supplementary material). We suggest that a high GC content is a convergently acquired feature of *Monosiga ovata* and some bilaterian *hsp70* genes. This could have contributed to weak and artefactual support for a closer phylogenetic link.

Phylogenetic position based on protein sequence analysis

In view of the complications encountered with analysis of the nucleotide sequences, we turned to phylogenetic analysis based on deduced protein sequences. If the GC bias in some *hsp70* genes was sufficient to cause bias in amino acid usage, convergent evolution of GC content could drive convergent evolution of amino acid sequence and disrupt attempts to draw protein-based phylogenetic trees. We tested for deviations in amino acid composition using the chi-squared test implemented by Tree-Puzzle 5.0. All sequences passed this test with $p > 97.5\%$ (for *Monosiga*, $p = 100\%$). We also used two indirect measures: general average hydropathicity (the arithmetic mean of the hydropathic indices of each amino acid) and aromaticity (the frequency of aromatic amino acids in the sequence). Under both indices, the choanoflagellate Hsp70 protein was normal, lying in the middle of the range of values for Hsp70 proteins from the other taxa (data not

Figure 2

A neighbor-joining (NJ) phylogenetic tree constructed from amino acid sequence alignment. Figures above the nodes indicate the percentage of NJ bootstrap support; the lower figures give Tree-Puzzle support value. Figures below 50% are not shown.

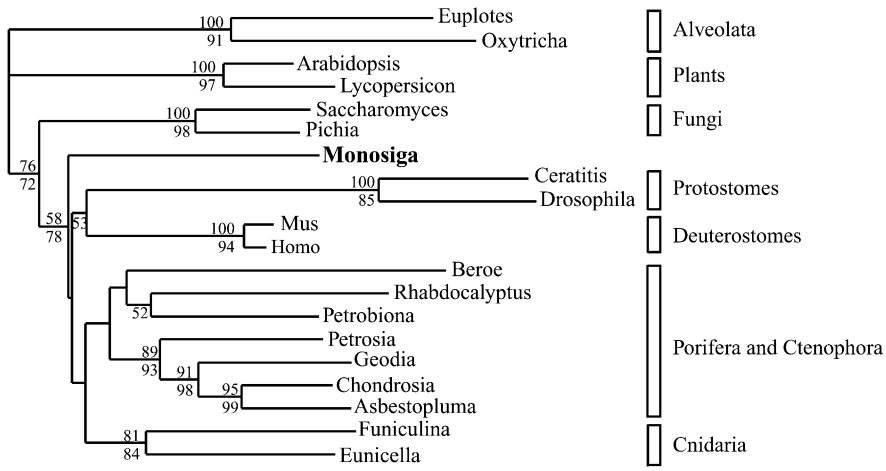


Table 1

ProtML analysis of user-defined trees.

Tree	Sister group to:	ln L	Difference in ln L	Standard error	Significantly worse?
A	animals	−5082.0	−1.1	1.9	no
B	fungi	−5090.2	−9.3	6.5	yes
C	animals/fungi	−5088.9	−8.0	6.9	yes
D	bilaterians	−5081.3	−0.4	2.5	no
E	diploblasts	−5080.9	0.0	–	best

ProtML analysis of user-defined trees showing the placement of choanoflagellate, the log likelihood, the difference in log likelihood from the best tree, and whether this difference equates to significantly worse than the best tree under Edward's criterion. Trees A, D, and E are significantly better than trees B and C, but they are not significantly different from each other. Of the three best trees, only tree A is compatible with previously hypothesized phylogenetic positions for choanoflagellates.

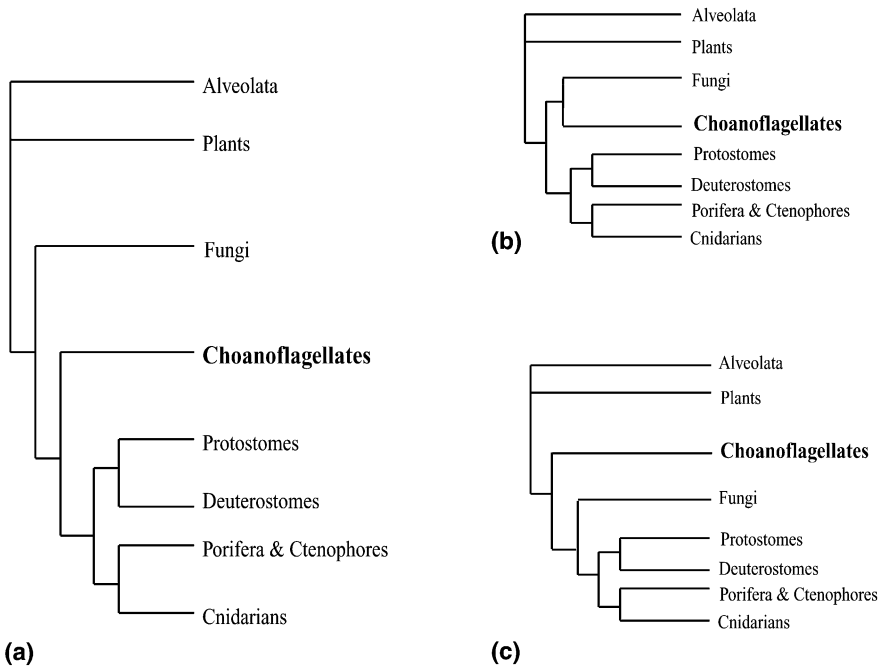
shown). We conclude that there is no strong amino acid usage bias; hence, phylogenetic analyses based on protein sequences should be reliable.

Phylogenetic analysis of amino acid sequences was performed using quartet-puzzling maximum likelihood (Tree-Puzzle) and a distance method (NJ). Both methods yielded optimal trees in which the choanoflagellate was grouped with the animals, excluding fungi (Tree-Puzzle = 78%; NJ = 58%). Within this clade, the NJ tree placed the choanoflagellate basal to the animals (Figure 2); however, this node had a bootstrap support of less than 50% and was collapsed by Tree-Puzzle. The support levels were highly dependant on the ingroup species chosen; for example, removal of *Rhabdocalyptus dawsoni*, *Beroe ovata*, and the insect species increased Tree-Puzzle support for the choanoflagellate-animal clade (excluding fungi) to 94%.

To test if this phylogenetic position is statistically significant, we used protein maximum likelihood (ProtML) to calculate and compare log likelihood values for three user-defined trees, using the original protein dataset (Table 1). These hypothetical trees place choanoflagellates as a sister group to animals (A), fungi (B), or both (C); these positions are the most biologically plausible positions for choanoflagellates (Figure 3). A fourth alternative, paraphyly, cannot be tested with our data. Tree A (sister to the animals) has the highest likelihood value; this tree is supported over the other two trees with statistical significance under Edward's criterion. Additional user-defined trees were constructed to test if choanoflagellates could

Figure 3

Three alternative phylogenetic placements for choanoflagellates. Tree A indicates a sister group relationship to the animals, tree B indicates a sister group relationship to fungi, and tree C indicates a sister group relationship to a clade including animals and fungi. Our analyses favor tree A.



be placed within the animals, either as basal to bilaterians or basal to diploblasts and sponges. Neither position was found to be significantly different from tree A, although both were more likely than trees B or C (Table 1).

Evolutionary implications

The data and analyses presented here argue that *Monosiga ovata* is more closely related to animals than to fungi. If Choanozoa is monophyletic, our conclusion could be safely extrapolated to all Choanozoa. An alternative possibility is that choanoflagellates are paraphyletic, with some species closely related to fungi, and some being the sister group to the animals [6, 8]. In this case, our conclusion concerning *Monosiga ovata* would reflect the position of just a subset of choanoflagellates. Either way, our data indicate that animals are closely related to some, or all, choanoflagellates. James-Clark's hypothesized relationship between sponges and choanoflagellates, proposed over 130 years ago [1, 2], is supported by our analyses.

Materials and methods

Choanoflagellate culture

Living cultures of *Monosiga ovata* Kent were purchased from American Type Culture Collection, ATCC. Cultures (10 ml) of *M. ovata* were grown in 25 cm^2 tissue culture flasks in sterile Sonneborn's Paramecium medium (SPM) (2.5 g dehydrated cereal leaves, 0.5 g Na_2HPO_4/l), which had been innoculated with *Enterobacter aerogenes* 24 hr in advance. Cultures were maintained at 20°C under a Sylvania Gro-Lux F20W/GRO light, with a light/dark cycle of 16 hr light and 8 hr darkness. Cultures were passaged every two weeks by transferring 0.25 ml into a fresh culture of *E. aerogenes* in SPM. Before extraction of choanoflagellate genomic DNA, high-density cultures were made axenic by the addition of Kanamycin (50 μg/ml), followed by incubation at 20°C for 48 hr.

PCR and cloning

Axenic *M. ovata* culture (1.5 ml) was removed, and cells were pelleted by centrifugation at 9000×g for 5 min, resuspended in 50 μl TlowE (10 mM Tris-Cl, 0.1 mM EDTA [pH 8.0]), and boiled for 10 min. The nested PCR strategy of Borchiellini et al. (1998) was used, with the same primer combinations, except that a lower annealing temperature of 45°C was used. After blunt-end cloning into pUC18, two recombinant clones were fully sequenced on both strands.

Molecular phylogenetics

The deduced protein sequence of *Monosiga ovata* Hsp70 was aligned with that of the following nucleotide sequences: Alveolata (U37280, L15292), plants (X77199, L41253), fungi (X12926, Z29379), Cnidaria (AF026516, AF026518), Ctenophora (AF026512), Porifera (AF026515, AF026514, AF026520, AF026517, AF026513, U20256), Arthopoda (Y08955, AJ001365), and Chordata (M12573, NM_005345). Alignment was carried out using CLUSTAL W [9] within the sequence editor package BioEdit [10]. Since many sequences did not cover the entire coding sequence, the ends were removed to create a blunt-ended alignment of 474 amino acids (1422 nucleotides) including gapped positions. Sites at which less than 25% of the sequences had an amino acid (75% or more contained a gap) were removed. The Phylip 3.57c package of programs was used for DNA-based phylogenetic analysis [11]. GC content, codon usage bias, and amino acid bias were calculated using codonW version 1.3 (J. Peden, Nottingham, UK). Neighbor-joining trees were constructed from a Dayhoff PAM distance matrix calculated from amino acid sequence using Phylip 3.57c, and robustness was assessed by 1000 bootstrap replicates. Quartet-puzzling maximum likelihood was applied to amino acid sequence data using Tree-Puzzle 5.0 [12], incorporating four gamma-distributed categories of between-site rate heterogeneity, a JTT matrix, and 1000 Puzzling steps. Comparison of user-defined trees for protein sequences was carried out using ProtML within MOLPHY 2.3 (J. Adachi and M. Hasegawa, Tokyo, Japan). Although Tree-Puzzle provides better models of sequence evolution, the program also automatically collapses poorly supported clades and was unable to distinguish between trees A and D, so it was not used in this instance. We used Edwards' criterion of a difference of two log-likelihood units to declare two likelihoods as different [13]. Edwards advocates this criterion when two models are not nested. More rigorous statistical tests such as the Kishino-Hasegawa test [14] were also carried out but were unable to provide statistical support for any of the alternative trees.

Supplementary material

Supplementary material including figures of percent GC content and effective number of codons used for Hsp70 proteins from diverse eukaryotes is available at http://images.cellpress.com/supmat/supmatin.htm.

Acknowledgements

We thank Mark Pagel and two referees for helpful suggestions. This work was funded by the Biotechnology and Biological Sciences Research Council.

References

1. James-Clark H: **Note on the infusoria flagellata and the spongiae ciliatae.** *Am J Sci* 1866, **1:**113-114.
2. James-Clark H: **On the spongiae ciliatae as infusoria flagellata; or observations on the structure, animality and relationship of *Leucosolenia botryoide*s, Bowerbank.** *Ann Mag Nat Hist* 1868, **1:**133-142.
3. Leadbeater BSC: **Choanoflagellate organization with special reference to loricate taxa.** In *The Biology of Free-Living Heterotrophic Flagellate.* Edited by Patterson DJ, Larsen J. Oxford: Clarendon Press; 1991:241-258.
4. Cavalier-Smith T: **Kingdom Protozoa and its 18 phyla.** *Microbiol Rev* 1993, **57:**953-994.
5. Wainwright PO, Hinkle G, Sogin ML, Stickel SK: **Monophyletic origins of the Metazoa, an evolutionary link with Fungi.** *Science* 1993, **260:**340-342.
6. Cavalier-Smith T: **Neomonada and the origin of animals and fungi**. In *Evolutionary Relationships among Protozoa.* Edited by Coombs GH, Vickerman K, Sleigh MA, Warren A. London: Chapman and Hall; 1998:375-407.
7. Borchiellini C, Boury-Esnault N, Vacelet J, Le Parco Y: **Phylogenetic analysis of the Hsp70 sequences reveals monophyly of Metazoa and specific phylogenetic relationships between animals and fungi.** *Mol Biol Evol* 1998, **15:**647-655.
8. Cavalier-Smith T: **The origin of fungi and pseudofungi.** In *Evolutionary Biology of the Fungi.* Edited by Raynor ADM, Brasier CM, Moore DM. Cambridge: Cambridge University Press; 1987:339-353.
9. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680.
10. Hill TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999, 41:95-98.
11. Felsenstein J: *PHYLIP (Phylogeny Inference Package) version 3.5c.* Seattle: University of Washington; 1993.
12. Strimmer K, von Haeseler A: **Quartet puzzling: a quartet maximum likelihood approach for reconstructing tree topologies.** *Mol Biol Evol* 1996, **13:**964-969.
13. Edwards AWF: *Likelihood.* London: Johns Hopkins University Press; 1972.
14. Kishino H, Hasegawa M: **Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea.** *J Mol Evol* 1989, **29:**170-179.