PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

A Molecular Analysis of the Phylogenetic Affinities of Saccoglossus cambrensis Brambell & Cole (Hemichordata)

Peter W. H. Holland, Adam M. Hacker and Nicola A. Williams

Phil. Trans. R. Soc. Lond. B 1991 **332**, 185-189
doi: 10.1098/rstb.1991.0048

A molecular analysis of the phylogenetic affinities of *Saccoglossus cambrensis* Brambell & Cole (Hemichordata)

PETER W. H. HOLLAND, ADAM M. HACKER AND NICOLA A. WILLIAMS

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

SUMMARY

Traditional approaches to phylogeny reconstruction have not allowed precise resolution of the evolutionary relationships between the major deuterostome phyla (chordates, hemichordates, echinoderms). Here we report the use of a molecular approach to investigate deuterostome phylogeny. We have used a polymerase chain reaction-based strategy to amplify, clone and sequence parts of the genes coding for 18S ribosomal RNA from *Saccoglossus cambrensis* (Hemichordata), *Arbacia sp.* (Echinodermata) and, for comparison, *Mytilus edulis* (Mollusca). We report the results of phylogenetic reconstructions using these, and homologous sequences from other eukaryotes. The results of our analyses are consistent with the hypothesis that *S. cambrensis* and vertebrates share a common ancestor not shared by echinoderms.

1. INTRODUCTION

The phylogeny of the Animal Kingdom has been a subject for vigorous discussion and investigation for over a century. However, while there is general agreement concerning the limits of most phyla, many of the evolutionary relationships between these major taxa remain unresolved. For example, there is little dispute as to which living species comprise the Chordata, Echinodermata and Hemichordata, or that these deuterostome phyla are more closely related to each other than they are to any other extant animals, but there is controversy regarding their precise evolutionary relationships. Thus many zoologists have considered a chordate:hemichordate clade probable (see, for example, Hyman (1959); Schaeffer (1987)), while others have recently suggested a chordate:echinoderm clade (see, for example, Inglis (1985); Jefferies (1986); Willmer (1990)).

Molecular biology offers potential as an additional source of comparative data for resolving such phylogenetic questions. However, while many RNA and DNA sequences have been obtained from chordates and echinoderms, only one sequence, that of 5S ribosomal RNA (rRNA) has previously been reported from a hemichordate (Ohama *et al.* (1984)). Unfortunately, 5S rRNA has generally proved of limited use for elucidating metazoan relationships, primarily because of its small size and its highly variable rate of sequence divergence (Hendriks *et al.* (1986)). In contrast, 18S rRNA sequences have proved much more useful for phylogenetic inference (see, for example, Field *et al.* (1988); Raff *et al.* (1988); Abele *et al.* (1989); Lake (1990)). Here we use 18S ribosomal gene sequences to examine the relationships between the chordates, echinoderms and hemichordates. Complete or partial 18S rRNA sequences have previously been reported from several chordate and echinoderm species, but from no hemichordate (Dams *et al.* 1988; Raff *et al.* 1988). Therefore, to allow a molecular investigation of deuterostome phylogeny, we have cloned part of the gene coding for 18S rRNA from the hemichordate *Saccoglossus cambrensis* Brambell & Cole, and the equivalent rDNA regions from the echinoderm *Arbacia* and the mollusc *Mytilus edulis* L. The cloning strategy we designed employed the polymerase chain reaction (PCR; Saiki *et al.* 1988) to amplify 18S rDNA to near purity, so that it could readily be cloned by recombinant DNA techniques, and its exact nucleotide sequence determined. Analysis of these sequences provides tentative support for the hypothesis that the hemichordates are the sister group of the chordates.

2. MATERIALS AND METHODS

(*a*) *Specimen collection and DNA extraction*

Saccoglossus cambrensis were collected by dredging mudflats at a depth of 11 m in Whitsand Bay, Cornwall, U.K. Fragmented tissue samples from several individuals were preserved in 2% SDS, 250 mM EDTA, 50 mM Tris. Cl (pH 8) at ambient temperature for 1 or 2 days, before refrigeration at 4 °C. Genomic DNA was isolated by digestion with 250 μg ml^{-1} proteinase K in 1% SDS, 100 mM NaCl, 5 mM EDTA, 50 mM Tris. Cl (pH 8) (55 °C, overnight) followed by multiple extractions with phenol, phenol:chloroform and chloroform, and ethanol precipitation. *Arbacia sp.* (*punculata* Lamarck?) DNA was kindly provided by Dr T. Hunt, Cambridge, U.K. Laboratory bred 21-day larvae from *Mytilus edulis* were used.

(*b*) *PCR amplification of ribosomal DNA*

By using published sequences from vertebrates, one arthropod and yeast (Huysman & DeWachter 1986),

Phil. Trans. R. Soc. Lond. B (1991), **332**, 185–189
Printed in Great Britain

Vol. 332. B 1264 13 [Published 29 June 1991]

we identified two highly conserved sequences flanking a variable 500–550 b.p. section of 18S rDNA. Two primers, complementary to the conserved flanking regions, were synthesized on a Milligen/Biosearch 7500 DNA synthesizer, to enable PCR-mediated amplification of metazoan 18S rDNA. The primer sequences were (5′ to 3′):

GCCAGTAGCATATGCTTGTCTC and

AGACTTGCCTCCAATGGATCC.

Preliminary experiments showed that a dramatic increase in the efficiency of PCR-mediated amplification of *Saccoglossus* rDNA could be achieved if the template DNA was first enriched for high molecular fragments by preparative gel electrophoresis. 50–500 ng of *Saccoglossus* or *Arbacia* genomic DNA, or a single *Mytilus* larva (frozen after lysis in distilled water), was added to 50 µl PCR buffer (10 mM Tris. Cl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01 % gelatin, 200 µM dATP, dCTP, dGTP, dTTP), 0.6 units *Taq* DNA polymerase (Perkin-Elmer Ltd.), and 500 ng of each of the two primers. Reactions were heated to 94 °C for 2.5 min and cycled 30–40 times through 94 °C (45 s), 50 °C (1 min), and 72 °C (2 min), followed by 10 min at 72 °C. Aliquots of each reaction were analysed by electrophoresis through 1.5 % agarose gels, to assess purity of the amplified products, before cloning.

(*c*) *Cloning and sequencing of PCR products*

The rDNA fragments amplified from *Saccoglossus* and *Mytilus* were blunt-ended with Klenow fragment, isolated by preparative gel electrophoresis, phosphorylated with T4 polynucleotide kinase, digested with *Bam*HI, ligated into pUC13 vector digested with either *Sma*I and *Bam*HI (*Saccoglossus*) or *Hinc*II and *Bam*HI (*Mytilus*), and transformed into competent *E. coli* DH5α.

The *Arbacia* rDNA fragment was cloned by a simpler strategy. The reaction products were phosphorylated by using T4 polynucleotide kinase, blunt-ended with Klenow fragment, and size fractionated by gel electrophoresis. The single band was isolated by electrophoresis onto DEAE-cellulose paper (Schleicher and Schuell NA45 paper; Dretzen *et al.* 1981), ligated into phosphatased *Sma*I-digested pUC13 vector (Pharmacia), and transformed into competent *E. coli* DH5α.

For each species, only inserts of the predicted size were present in recombinant plasmids selected by blue/white screening. Inserts were sequenced by using T7 DNA polymerase and 7-deaza dGTP sequencing mixes (Pharmacia), following the manufacturer's instructions. Four independent *Saccoglossus* clones, two independent *Arbacia* clones and one *Mytilus* clone, were sequenced.

(*d*) *Sequence analysis*

Phylogenetic analyses were performed by using 18S rDNA sequences from *Saccoglossus*, *Arbacia*, *Mytilus* and representatives from two other animal phyla for which complete RNA or DNA sequence is available over the entire amplified region. These were two arthropods (*Artemia salina* (L.) and *Tenebrio molitor* (L.)) and three chordates (human, mouse and *Xenopus laevis* (Daudin)). To ensure that homologous sites were used for phylogenetic inference, only regions that could be clearly aligned between all species were analysed. These were identified by using the multiple sequence alignment programme CLUSTAL (Higgins & Sharp 1988; kindly provided by Dr P. Sharp, Dublin, Ireland), followed by manual adjustment based on the 18S rRNA secondary structure models of Dams *et al.* (1988). Regions of ambiguous alignment were then removed, the CLUSTAL analysis repeated, and minor adjustments again made manually to optimize sequence similarities. One difficulty encountered was in defining the precise limits of the regions of uncertain alignment. A rule we employed was that the regions of putative homology to be used in phylogenetic reconstructions were flanked by at least one invariant site. This resulted in 457 conserved and semi-conserved sites being available for phylogenetic analyses. Previously reported partial 18S rRNA sequences (Field *et al.* 1988) were not used in the analyses, since they overlap incompletely with this region.

Regions putatively capable of forming double-stranded helices were identified by comparison with the secondary structure model of Dams *et al.* (1988).

Phylogenetic reconstruction was performed by using maximum likelihood (Felsenstein 1981), the distance method of Fitch & Margoliash (1967) and maximum parsimony (Eck & Dayhoff 1966). The programmes used were DNAML87 and FITCH87 (from the PHYLIP 3.2 and 3.3 packages, respectively; kindly provided by Dr J. Felsenstein, University of Washington, Seattle, U.S.A.) and PAUP 3.0 k (Swofford 1990). In DNAML87, a transition:transversion ratio of 2 was used, the 'frequencies' option selected and 'global optimization' was used. The robustness of the trees produced by the Fitch–Margoliash and parsimony methods was assessed by bootstrapping with 500 replicates (Felsenstein 1985).

3. RESULTS

(*a*) *Amplification, cloning and sequencing of 18S rDNA*

PCR-mediated amplification of 18S rDNA was performed on genomic DNA extracted from *Saccoglossus cambrensis* and *Arbacia sp.*, and on a single *Mytilus edulis* larva. Pure products were produced from each species, and cloned into plasmids as described.

Preliminary DNA sequencing analyses revealed heterogeneity between *Saccoglossus* clones; this was therefore investigated by completely sequencing four independent clones. Two clones had identical sequences (shown in figure 1), differing from a third clone by a single nucleotide over the 534 b.p. internal to the primers (G at site 21 of figure 1). It is possible that this difference simply reflects a copying error introduced by *Taq* DNA polymerase during PCR amplification; an event which we estimate occurs less

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

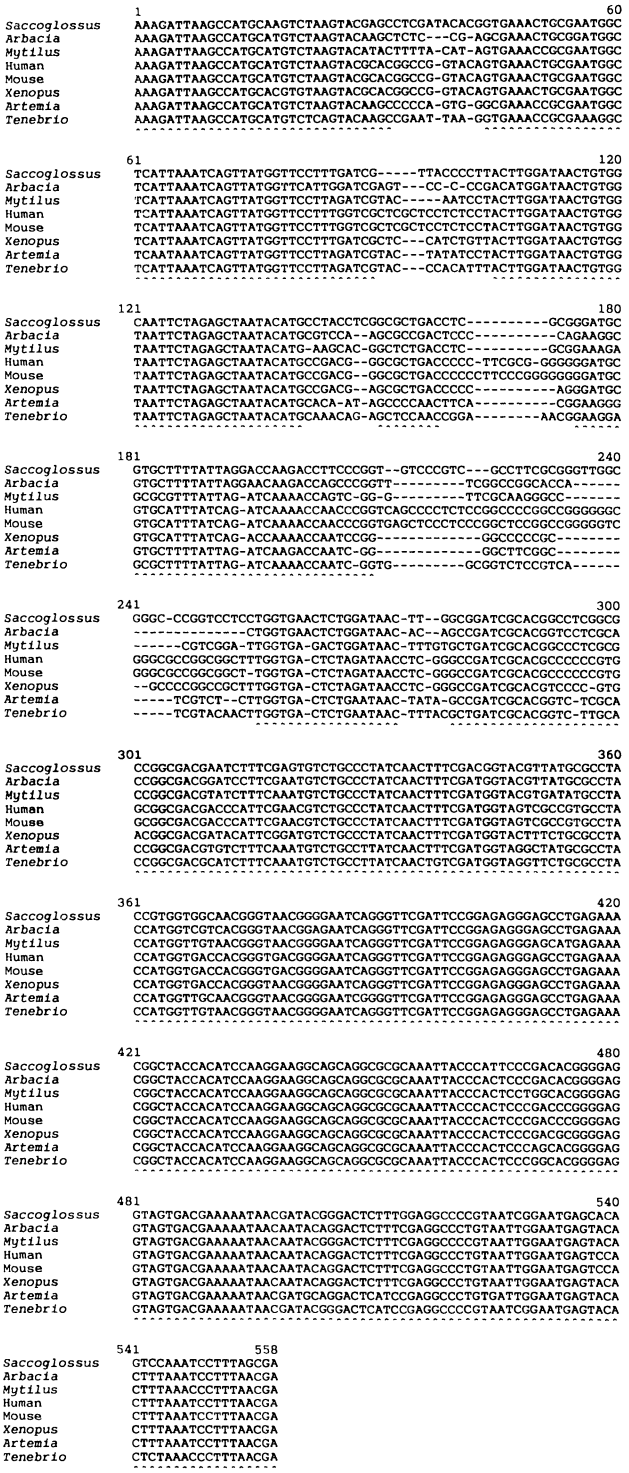

Figure 1. DNA sequences of *Saccoglossus*, *Arbacia* and *Mytilus* 18S rDNA clones, aligned with homologous sequences from vertebrates and arthropods. Dashes indicate gaps, arrowheads indicate sites used in phylogenetic analyses.

than once per 1000 bases (P.W.H.H. and N.A.W., unpublished observations). A fourth clone had four differences from the consensus (C at site 51, G at site 62, A at site 181, T at site 200); too many to be attributed to PCR errors. Although this level of divergence is approximately that which might be expected between two species within a genus, we do not believe that our sequences represent two species since (i) *S. cambrensis* is the only enteropneust species

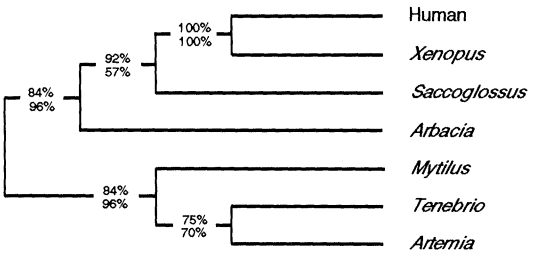

Figure 2. Consensus phylogenetic tree for 18S rDNA sequences from seven metazoan species. The figures denote the percentage of times the taxa to the right of each node were grouped into a monophyletic assemblage from 500 bootstrap resamplings (upper figure, Fitch–Margoliash method; lower figure, maximum parsimony). The tree is rooted by reference to classical and molecular phylogenies (Barnes *et al.* 1988; Field *et al.* 1988).

recorded from Whitsand Bay, and (ii) three of the differences result in substitution at sites otherwise highly conserved throughout the Metazoa. The explanation we currently favour for the origin of the divergent clone is that it represents sequence from a non-functional 18S rDNA gene from *S. cambrensis*.

Two independent rDNA clones from *Arbacia* were also completely sequenced. These differed by only a single substitution over the 505 b.p. internal to the primers (C at site 466 in figure 1). We suggest that the difference reflects a PCR error, and have used the sequences shown in figure 1 in phylogenetic analyses. A previously reported partial 18S rRNA sequence from *Arbacia punctulata* (Field *et al.* 1988) differs from the sequence reported here at seven sites within a 406 nucleotide region of overlap. A single rDNA clone from *Mytilus edulis* was obtained and sequenced (figure 1).

(*b*) *Phylogenetic affinities of Saccoglossus*

To investigate the phylogenetic relationships between chordates, hemichordates, and echinoderms, the *Saccoglossus* and *Arbacia* 18S rDNA sequences obtained in this study were aligned with rRNA sequences from three vertebrates, and for outgroup comparison, two arthropods and *Mytilus edulis*. 457 homologous sites were identified in the eight species, although mouse and human were identical at these sites. Phylogenetic analyses were performed using the sequences, by maximum likelihood, maximum parsimony and the distance method of Fitch–Margoliash. The three methods yielded trees of identical topology, within which *Saccoglossus* shares a common ancestor with the vertebrates, not shared by *Arbacia* (figure 2). To investigate the robustness of the trees produced by parsimony and distance methods, bootstrap resampling was performed (500 replicates for each method). As shown in figure 2, bootstrapped distance analysis clearly supports a *Saccoglossus*: vertebrate clade over alternative phylogenies (92% of trees), but bootstrapped parsimony analysis does so only marginally (57%). Differential weighting of transversions over transitions (10:1), or of putative single-stranded regions over helical regions (2:1), altered the bootstrapped parsimony confidence limits by 2% or less.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

4. DISCUSSION

Comparison between the primary structures of homologous genes or proteins has been widely used for investigation of phylogenetic relationships. The larger ribosomal RNA molecules (18S and 28S rRNA in Metazoa) are useful sources of data for revealing distant evolutionary relationships, primarily due to the conserved nature of their sequences. Most studies that have utilized rRNA for investigating metazoan phylogeny have determined these sequences by direct sequencing from total cellular RNA (see, for example, Field *et al.* (1988); Raff *et al.* (1988); Abele *et al.* (1989)).

Here we report the use of an alternative strategy involving PCR to amplify DNA sequences coding for rRNA, before cloning by recombinant DNA techniques. We used this strategy to overcome several technical problems inherent in the direct rRNA sequencing approach. First, direct rRNA sequencing rarely yields unambiguous sequences, whereas recombinant DNA can easily be purified for optimal, exact, sequence determination. Second, it has been reported that purification of RNA is extremely difficult from some species, including a hemichordate (Ghiselin 1988). In contrast, DNA extraction is easily optimized, and PCR is tolerant to variations in DNA quality. Two further advantages of PCR and cloning not essential to this study, but potentially of great importance, are applicability to heterogeneous samples and to minute amounts of tissue. The PCR strategy described allowed us to amplify and sequence approximately one-third of the 18S ribosomal RNA gene (spanning helices 1–9, and the variable V1, V2, and V3 regions; Dams *et al.* 1988), from one hemichordate, one echinoderm and one mollusc species. Three alternative phylogenetic reconstruction strategies were performed on these sequences, plus homologous regions from three vertebrates to represent the chordates, and two other invertebrates as outgroups. Maximum likelihood, maximum parsimony and Fitch–Margoliash analyses produced congruent results, each suggesting that the hemichordates are the sister group of the chordates. However, the low confidence level given by bootstrapped parsimony analysis indicates that this conclusion must be treated as tentative.

It is interesting to consider whether a chordate:hemichordate clade, as suggested by our data, is compatible with other evidence (reviewed by Barnes *et al.* (1988) and Willmer (1990)).

Larval morphology is strikingly similar between enteropneusts and some echinoderms, perhaps reflecting common ancestry. However, since the tadpole larva of urochordates clearly has a specialized, derived morphology, the similarity is not evidence for a sister group relationship between echinoderms and hemichordates.

Several morphological features supposedly shared by hemichordates and chordates have been proposed by earlier workers, but many are now thought to be either fallacious or superficial. These include supposed homology between a gut diverticulum in hemichordates and the notochord of chordates, and similarity in dorsal nervous tissue organization. However, serially repeated gill slits and gill bars are present in hemichordates and chordates, but not echinoderms. In contrast, some palaeontological evidence has been interpreted as supporting an echinoderm:chordate clade (Jefferies 1986).

In summary, we have presented molecular evidence to tentatively support the hypothesis that the lineage leading to the echinoderms diverged before the separation of hemichordates and chordates. This suggestion is consistent with many classical accounts, but does contradict some palaeontological evidence (Jefferies 1986). Further insight into deuterostome phylogeny could be obtained by determining additional, or more extensive, 18S rDNA sequences from hemichordates and from potentially related invertebrates, such as lophophorates and chaetognaths. These experiments are currently under way.

We thank Jeremy Lanfear, Quentin Bone, F.R.S., David Dixon and Tim Hunt for specimens and DNA samples, John McVey for primer synthesis, Max Telford, Jeremy Lanfear, Richard Griffiths, Colin Patterson and Clint Turbeville for advice on phylogeny reconstruction, Paul Sharp, Joe Felsenstein and David Swofford for computer programmes, and Jon Bartlett for assistance. This research was funded by SERC, NERC, The E.P.A. Cephalosporin Fund, and The Queen's College, Oxford.

REFERENCES

Abele, L. G., Kim, W. & Felgenhauer, B. E. 1989 Molecular evidence for inclusion of the Phylum Pentastomida in the Crustacea. *Molec. biol. Evol.* **6**, 685–691.

Barnes, R. S. K., Calow, P. & Olive, P. J. W. 1988 The invertebrates: a new synthesis. Blackwell Scientific Press.

Dams, E., Hendriks, L., Van de Peer, Y., Neefs, J.-M., Smits, G., Vandenbempt, I. & DeWachter, R. 1988 Compilation of small ribosomal subunit RNA sequences. *Nucl. Acid Res.* **16**, r87–r178.

Dretzen, G., Bellard, M., Sassone-Corsi, P. & De Chambon, P. 1981 A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *An. Biochem.* **112**, 295–298.

Eck, R. V. & Dayhoff, M. O. 1966 In *Atlas of protein sequence and structure* (ed. M. O. Dayhoff), pp. 161–169. Washington, D.C.: National Biomedical Research Foundation.

Felsenstein, J. 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. molec. Evol.* **17**, 368–376.

Felsenstein, J. 1985 Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Field, K. G., Olsen, G. J., Lane, D. J., Giovannoni, S. J., Ghiselin, M. T., Raff, E. C., Pace, N. R. & Raff, R. A. 1988 Molecular phylogeny of the animal kingdom. *Science, Wash.* **239**, 748–753.

Fitch, W. M. & Margoliash, E. 1967 Construction of phylogenetic trees. *Science, Wash.* **155**, 279–284.

Ghiselin, M. T. 1988 The origin of molluscs in the light of molecular evidence. *Oxford Surveys Evol. Biol.* **5**, 66–95.

Hendriks, L., Huysmans, E., Vandenberghe, A. & DeWachter, R. 1986 Primary structure of the 5S ribosomal RNAs of 11 arthropods and applicability of 5S RNA to the study of metazoan evolution. *J. molec. Evol.* **24**, 103–109.

Higgins, D. G. & Sharp, P. M. 1988 CLUSTAL: a package

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

for performing multiple sequence alignment on a microcomputer. *Gene* **73**, 237–244.

Huysmans, E. & DeWachter, R. 1986 Compilation of small ribosomal subunit RNA sequences. *Nucl. Acid Res.* **14**, r73–r118.

Hyman, L. H. 1959 *The invertebrates: smaller coelomate groups*, vol. 5. New York: McGraw-Hill.

Inglis, W. G. 1985 Evolutionary waves: patterns in the origins of animal phyla. *Aust. J. Zool.* **33**, 153–178.

Jefferies, R. P. S. 1986 *The ancestry of the vertebrates*. London: British Museum (Natural History).

Lake, J. A. 1990 Origin of the metazoa. *Proc. natn. Acad. Sci. U.S.A.* **87**, 663–766.

Ohama, T., Kumazaki, T., Hori, H. & Osawa, S. 1984 Evolution of multicellular animals as deduced from 5S rRNA sequences: a possible early emergence of the Mesozoa. *Nucl. Acid Res.* **12**, 5101–5108.

Raff, R. A., Field, K. G., Ghiselin, M. T., Lane, D. J., Olsen, G. J., Parks, A. L., Parr, B. A., Pace, N. R. & Raff, E. C. 1988 Molecular analysis of distant phylogenetic relationships in echinoderms. In *Echinoderm phylogeny and evolutionary biology* (ed. Paul, C. R. C. & Smith, A. B.). Oxford University Press.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S., Higuchi, R. H., Horn, G. T., Mullis, K. B. & Erlich, H. A. 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science, Wash.* **239**, 487–491.

Schaeffer, B. 1987 Deuterostome monophyly and phylogeny. *Evol. Biol.* **21**, 179–235.

Swofford, D. L. 1990 PAUP 3.0 k computer programme. Illinois Natural History Survey, Illinois, U.S.A.

Willmer, P. 1990 *Invertebrate relationships: patterns in animal evolution*. Cambridge University Press.

Received 27 April 1990; revised 7 December 1990; accepted 8 January 1991

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES