Insect Molecular Biology (1999) **8**(4), 527–530

Intron insertion as a phylogenetic character: the *engrailed* homeobox of Strepsiptera does not indicate affinity with Diptera

A. Rokas,* J. Kathirithamby† and P. W. H. Holland

School of Animal and Microbial Sciences, The University of Reading, U.K. and †Department of Zoology, University of Oxford, U.K.

Abstract

The phylogenetic relationships of the order Strepsiptera are unclear. Affiliation to Coleoptera has been proposed, however this implies that dipteran halteres and strepsipteran haltere-like organs evolved convergently. An alternative is a sister group relationship with Diptera. In this case, halteres could be homologous but a radical homeotic mutation may have switched their position to the Strepsipteran mesothorax. Ribosomal DNA sequence analysis has been used to support Dipteran affiliation, although this is controversial. Here we investigate the potential of an intron insertion site as a phylogenetic character. We find that the *en* homeobox gene of the strepsipteran *Stichotrema dallatorreanum* lacks a derived intron insertion shared by representatives of Diptera and Lepidoptera. We argue against a close affiliation between Strepsiptera and Diptera.

Keywords: Strepsiptera, *Stichotrema*, phylogeny, intron, *engrailed*.

Introduction

Insects of the order Strepsiptera are entomophagous endoparasitoids with extreme sexual dimorphism. Many species are known only from the free-living winged males, because females are permanently endoparasitic (with the exception of the family Mengenillidae where females emerge to pupate externally, but are still wingless as adults). Ever since the first strepsipteran was described in 1793, their phylogenetic affinities have been a subject of debate and confusion. In recent years, four alternative phylogenetic placements have been seriously suggested: as a sister group to Coleoptera (e.g. Kathirithamby, 1989; Kukalova-Peck & Lawrence, 1993), within the Coleoptera (Crowson, 1960), outside the Holometabola (Kristensen, 1991) and as a sister group to the Diptera (Whiting *et al.*, 1997). A proposed affinity to Coleoptera is based primarily on the use of hindwings for flight in both orders, several associated morphological characters and the pattern of hindwing venation (Crowson, 1960; Kathirithamby, 1989; Kukalova-Peck & Lawrence, 1993). The major morphological feature suggesting affinity to Diptera is the fact that strepsipteran forewings are reduced, resembling the metathoracic halteres of dipterans. The reduced forewings of strepsipterans play the same role as dipteran halteres; both act as gyroscopic balancing organs during flight (Pix *et al.*, 1993). Possession of a similar structure on different segments of the body is often taken as evidence against homology. This is not necessarily the case, because mutations in developmental control genes are capable of transforming the positions of structures along the body axis (homeotic mutations; Lewis, 1978). Whiting & Wheeler (1994) made the intriguing suggestion that the different segmental position of halteres and haltere-like organs in Diptera and Strepsiptera may reflect a natural homeotic mutation that caused a reversal of T2 and T3 segment identities in the ancestral strepsipteran lineage. This example has been cited as a very rare case of homeotic mutation contributing to body plan evolution (Raff, 1996).

In the face of conflicting morphological evidence, molecular data can give an independent insight into phylogenetic relationships. Molecular phylogenetic analysis of 18S ribosomal DNA (rDNA) sequences has been interpreted as supporting a sister group relationship between Strepsiptera and Diptera (Chalwatzis *et al.*, 1996; Whiting *et al.*, 1997). However, a thorough analysis of 28S and 5.8S rDNA sequences concluded that strong support could only be given to placement of Strepsiptera within the Holometabola; affinity with either Diptera or Coleoptera

*Present address: Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Building, West Mains Road, Edinburgh EH9 3JT, U.K.

Received 2 March 1999; accepted 10 June 1999. Correspondence: Professor Peter W. H. Holland, School of Animal and Microbial Sciences, The University of Reading, Whiteknights, PO Box 228, Reading RG6 6AJ, U.K. E-mail: p.w.h.holland@reading.ac.uk

could not be distinguished due to saturation of nucleotide substitutions (Hwang *et al.*, 1998). Furthermore, other authors have suggested caution in interpretation of the 18S rDNA data, noting the extremely fast rates of rDNA evolution in Diptera and Strepsiptera (Carmean & Crespi, 1995). It is known that fast evolving sequences can be artefactually grouped by molecular phylogenetic analyses (Felsenstein, 1978), and recent work suggests that strepsipteran and dipteran rDNA branch lengths are long enough to attract in this manner (Huelsenbeck, 1997; Huelsenbeck, 1998). Therefore, current molecular data do not definitively resolve the phylogenetic placement of Strepsiptera within the holometabolous insects.

When the resolving power of primary sequence data is reduced, rare large-scale molecular changes provide an independent source of phylogenetic information, largely immune from rate effects (Holland & Garcia-Fernàndez, 1996). The principal difficulty with this approach is identifying rare mutations unique to clades of interest. Examples of rare mutations that have been used for phylogenetic inference include mitochondrial gene order, retroposon integrations, gene duplication and protein domain evolution (Holland & Garcia-Fernàndez, 1996; Balavoine, 1997). Here we propose the use of intron insertions as shared derived characters, and examine the potential of a recently evolved intron in the *engrailed* (*en*) homeobox gene as a diagnostic character for resolving the phylogenetic affinities of Strepsiptera.

Results and Discussion

We first compared intron-exon organization of all *en* class homeobox genes currently available on GenBank. We note that an intron insertion within the homeobox is located at an identical site in the two *en* class genes (*en* and *inv*) of the dipteran *Drosophila melanogaster*. The *en* gene of *Anopheles gambiae* (Diptera), the *en* gene of *D. virilis* (Diptera) and the *en* and *inv* genes of *Bombyx mori* (Lepidoptera) all have an intron present at the same site within the homeobox. These represent all the *en* class genes characterized at the genomic level from Diptera and Lepidoptera. In contrast to these two orders, an intron at this position in the homeobox is absent from *en* class genes characterized from representatives of Coleoptera (*Tribolium castaneum*), Hymenoptera (*Apis mellifera*) and all outgroup taxa (Crustacea, Chelicerata, Onycophora, Annelida, Mollusca, Brachiopoda, Echinodermata, Chordata; see Experimental procedures). Clearly, absence of the intron is the primitive state, and presence of the intron in Diptera and Lepidoptera is the derived condition. Because intron insertion is rare, and intron position is likely to be selectively neutral, these types of mutation are effectively immune from convergent evolution. Hence, the intron considered here is a shared derived character of Diptera and Lepidoptera.

If Strepsiptera are indeed the sister group to Diptera, and evolved by homeosis, we predict that they are likely to possess this intron, at precisely the same insertion site as in flies and moth; absence is predicted if Strepsiptera are allied to Coleoptera or to any other order of insects. In an attempt to resolve between these alternatives, we used degenerate PCR to clone the *en* homeobox from the Papua New Guinea strepsipteran *Stichotrema dallatorreanum* Hofeneder, an endoparasite of a tettigonid grasshopper, *Segestidea novaeguineae* (Brancsik).

One set of PCR primers was designed to flank the diagnostic intron (JM36b and AROUT); a second set targets a region adjacent to the intron to control for potential amplification bias resulting from intron size (AR3 and MAQGLY; Fig. 1). After DNA amplification with each primer set, bands were cloned and multiple independent clones sequenced. Two distinct *en* class genes were isolated from the strepsipteran sample. One sequence has high identity to the *en* gene of the locust *Schistocerca gregaria*, suggesting it is host-derived; this was confirmed by cloning of this (and no other) *en* gene from an unparasitized specimen of the host. The second sequence is derived from the strepsipteran genome. This was verified by designing a gene-specific primer (STREN) which, when used with the 5′ degenerate primer (JM36b), amplified the strepsipteran gene from two additional strepsipteran specimens (*Stichotrema* sp.). One specimen yielded a sequence identical over seventy-eight nucleotides, whereas the other yielded clones with one synonymous mismatch (G to A) at position 72. Because all four clones sequenced from this animal shared this transition, it is not a PCR error; it is likely to be an allelic variant of the same gene (Fig. 2). These primers did not amplify from control host DNA, confirming strepsipteran origin.

Importantly, we found that the strepsipteran *en* gene does not possess an intron in the homeobox (nor does the grasshopper gene). This contrasts to all four previously reported dipteran *en* class genes (*Drosophila melanogaster*, *D. virilis*, *Anopheles en*; *D. melanogaster inv*) and both lepidopteran *en* class genes (*Bombyx en* and *inv*).

(A) *Stichotrema dallatorreanum en*

```
AGTAGTGCTCAACTGGCAAGACTGAAGCACGAGTTTGCGGAAAATCGTTACCTAACCGAA
 S  S  A  Q  L  A  R  L  K  H  E  F  A  E  N  R  Y  L  T  E
AAACGGAGACAGCAATTGAGCAATGAGTTGGGACTCAACGAGGCGCAGATCAAGATCTGG
 K  R  R  Q  Q  L  S  N  E  L  G  L  N  E  A  Q  I  K  I  W
TTTCAAAATAAAAGGGCAAAAATTAAAAAATCTTC
 F  Q  N  K  R  A  K  I  K  K  S  S
```

(B) *Segestidea novaeguineae en*

```
AGCGGAGAGCAACTGGCGAGGCTGAAGCACGAGTTCGCGGAGAACCGCTACCTGACGGAG
 S  G  E  Q  L  A  R  L  K  H  E  F  A  E  N  R  Y  L  T  E
CGAAGACGCCAAGAGTTGGCCCGGGAGCTGGGCCTCAACGAGGCGCAGATCAAGATCTGG
 R  R  R  Q  E  L  A  R  E  L  G  L  N  E  A  Q  I  K  I  W
TTCCAGAACAAACGTGCAAAGATCAAGAAGGCGAGTGGTCAGAAG
 F  Q  N  K  R  A  K  I  K  K  A  S  G  Q  K
```

Figure 1. Generalized structure of an *en* class homeobox showing position of the diagnostic intron and the PCR primers used for intron detection. The translation shown is that of the *D. melanogaster* engrailed homeodomain.

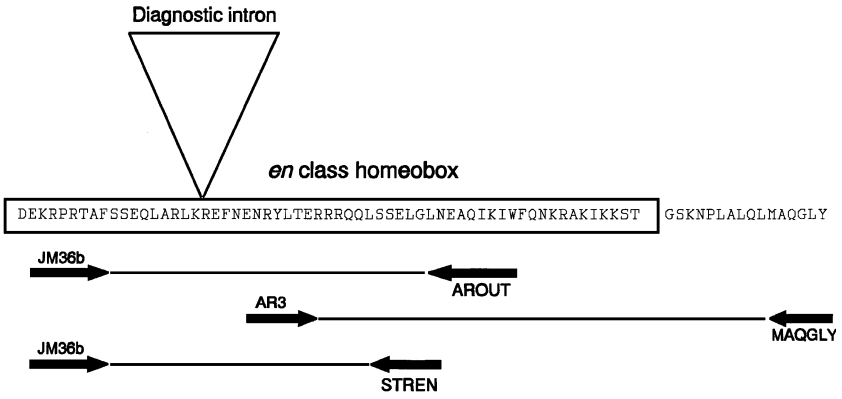

Figure 2. DNA and deduced protein sequences of the (A) strepsipteran and (B) orthopteran *en* class homeoboxes (excluding primer sequences). One of three strepsipteran specimens had a synonymous G to A substitution at position 72. GenBank accession numbers: AF130851–2.

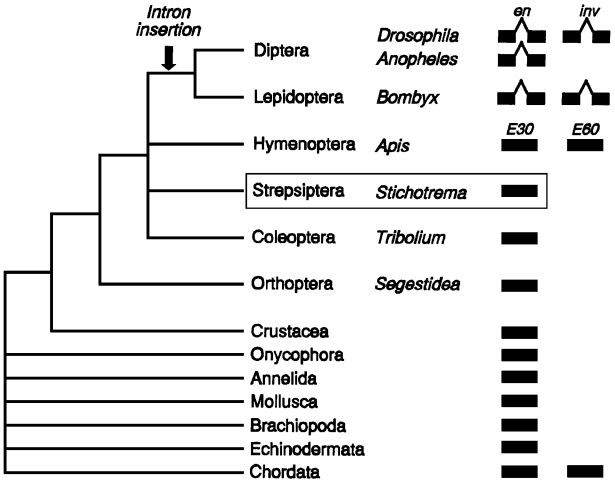

Figure 3. Phylogenetic tree indicating a shared intron insertion in *en* class homeoboxes of Diptera and Lepidoptera but not Strepsiptera, Coleoptera, Hymenoptera or outgroup taxa. Black boxes denote *en* class homeoboxes, with or without the diagnostic intron.

Lack of the intron is shared with Coleoptera, more basal insect orders and outgroups. The finding that the *en* homeobox of Strepsiptera lacks the diagnostic intron implies that the *en* gene adds no support to the suggestion that Strepsiptera are closely allied to Diptera. We suggest it is more likely that Strepsiptera occupy a more basal phylogenetic position within the Holometabola along with the Coleoptera, Hymenoptera and other non-Panorpoid insects (Fig. 3). If our interpretation is correct, then Strepsiptera are not homeotic Diptera. Instead, the common ancestor of Diptera and Strepsiptera did not possess halteres and these organs evolved from wings independently in the two orders.

The robustness of our interpretation depends on the reliability of intron insertion as a phylogenetic indicator. Notwithstanding the debate about the evolutionary origins of introns, it is clear that new introns have been inserted into genes during the course of animal evolution (O'Neill *et al*., 1998; Tarrio *et al*., 1998). Because insertion into a given gene is very rare, and intron position is likely to be selectively neutral, we argue that intron insertion events are effectively immune from convergent evolution. Identical intron positions must reflect common ancestry. The probability of intron loss is more difficult to estimate; intron loss has not yet been documented in the *en* class genes and we note it is rare in other gene families such as calmodulin (Côrte-Real *et al*., 1994). However, possible loss of the *en* intron in the strepsipteran lineage cannot be totally excluded. Further sampling for the presence or absence of this intron within Strepsiptera and other insect orders will test our argument that the halteres of Strepsiptera and Diptera represent a remarkable case of convergent evolution and not a rare case of natural homeotic transformation.

Experimental procedures

Adult female specimens of the strepsipteran *Stichotrema dallatorreanum* Hofeneder and *Stichotrema* sp. (family Myrmecolacidae) were collected from Papua New Guinea, as were unparasitized specimens of their insect host, the long-horned grasshopper *Segestidea novaguineae* (Brancsik) (family Tettigonidae). All specimens were stored in 95% ethanol prior to DNA extraction. Total DNA was extracted using the QIAamp Blood Kit (QIAGEN cat. no. 29104; QIAGEN, Crawley, U.K.) following the supplier's protocol for insect DNA extraction. Oligonucleotide primers JM36b and AROUT were designed to amplify across the diagnostic intron site in *en* class homeobox genes:

JM36b (5′–3′) GAGAAGCGNCCACGNACNGCNTT
AROUT (5′–3′) GATCTTGATCTGCGCCTCGTTGAG

To control for potential amplification bias from intron insertion, primers AR3 and MAQGLY were also designed; these target a region adjacent to the diagnostic intron:

AR3 (5′–3′) GA(G,A)AA(C,T)CGNTATCTGACNGAG
MAQGLY (5′–3′) GTGGTTGTACAGNCC(C,T)TGNGCCAT

PCR reactions were performed using standard reaction concentrations (Holland, 1993) with the following cycling parameters: 94 °C for 2 min, followed by thirty-five cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min, and a final extension stage

of 10 min at 72 °C. PCR fragments were purified from agarose gels, cloned into plasmid vectors, and multiple recombinant clones sequenced for each amplified band. After sequencing of the strepsipteran *en* class homeobox, a strepsipteran-specific primer (STREN) was designed to enable a test of clone authenticity.

STREN (5′–3′) TTCAGTCCCAACTCATT(G,A)CT

DNA sequences obtained in this study are available on GenBank (*Stichotrema dallatorreanum en* AF130851, *Segestidea novaguineae en* AF130852). Accession numbers for other *en* class genes to which comparison was made are: Diptera (*Drosophila melanogaster* K03055–8, *D. virilis* X04727, *Anopheles gambiae* U42214), Lepidoptera (*Bombyx mori* M64335–6), Hymenoptera (*Apis mellifera* M29489–90), Coleoptera (*Tribolium castaneum* S73225), Crustacea (U69098–105), Chelicerata (AF071404), Onychophora (AKY10771), Annelida (X58692, U26639), Mollusca (U23153–4, U23212–4, U23431–3, U21675, U21857), Brachiopoda (X62688), Echinodermata (U58775, M19709), Chordata (X59120–6, U82487).

Acknowledgements

We thank Rebecca Saff and Natalie Halbert for DNA from the two supplementary specimens of *Stichotrema* sp. Antonis Rokas was partially funded by the Greek Foundation of State Scholarships, through the SOCRATES link between the Universities of Crete and Reading. P.W.H.H. acknowledges a Royal Society grant. Specimens were collected by J.K. and staff of the Entomology Division, Papua New Guinea Oil Palm Research Association Inc., with special thanks to the Director, Mr Ian Orrell.

References

Balavoine, G. (1997) The early emergence of platyhelminths is contradicted by the agreement between 18S rRNA and Hox gene data. *Comptes Rendus de l'Academie des Sciences serie III de la Vie – Life Sci* **320**: 83–94.

Carmean, D. and Crespi, B.J. (1995) Do long branches attract flies? *Nature* **373**: 666.

Chalwatzis, N., Hauf, J., van de Peer, Y., Kinzelbach, R. and Zimmermann, R.K. (1996) 18S ribosomal RNA genes of insects: Primary structure of the genes and molecular phylogeny of the Holometabola. *Ann Entomol Soc Am* **89**: 788–903.

Côrte-Real, H.B.S.M., Dixon, D.R. and Holland, P.W.H. (1994) Intron-targeted PCR: a new approach to survey neutral DNA polymorphism in bivalve populations. *Mar Biol* **120**: 407–413.

Crowson, R.A. (1960) The phylogeny of Coleoptera. *Ann Rev Entomol* **5**: 111–134.

Felsenstein, J. (1978) Cases in which parsimony and compatibility methods will be positively misleading. *Syst Zool* **27**: 401–410.

Holland, P.W.H. and Garcia-Fernàndez, J. (1996) *Hox* genes and chordate evolution. *Dev Biol* **173**: 382–395.

Holland, P.W.H. (1993) Cloning genes using PCR. *Essential Developmental Biology: a Practical Approach* (Stern, C.D. and Holland, P.W.H., eds), pp. 243–255. IRL Press at Oxford University Press.

Huelsenbeck, J.P. (1997) Is the Felsenstein zone a fly trap? *Syst Biol* **46**: 69–74.

Huelsenbeck, J.P. (1998) Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Syst Biol* **47**: 519–537.

Hwang, U.W., Kim, W., Tautz, D. and Friedrich, M. (1998) Molecular phylogenetics at the Felsenstein zone: approaching the Strepsiptera problem using 5.8S and 28S rDNA sequences. *Mol Phylogen Evol* **9**: 470–480.

Kathirithamby, J. (1989) Review of the order Strepsiptera. *Syst Entomol* **14**: 41–92.

Kristensen, N.P. (1991) Phylogeny of extant hexapods. *Insects of Australia: a Textbook for Students and Research Workers* (Naumann, I.D., Carne, P.B., Lawrence, J.F., Nielsen, E.S., Spradberry, J.P., Taylor, R.W., Whitten, M.J. and Littlejohn, M.J., eds), pp. 125–140. 2nd edn. CSIRO and Melbourne University Press, Melbourne.

Kukalova-Peck, J. and Lawrence, J.F. (1993) Evolution of the hind wing in Coleoptera. *Can Entomol* **125**: 181–258.

Lewis, E.B. (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* **276**: 565–570.

O'Neill, R.J.W., Brennan, F.E., Delbridge, M.L., Crozier, R.H. and Graves, J.A.M. (1998) *De novo* insertion of an intron into the mammalian sex determining gene, SRY. *Proc Natl Acad Sci USA* **95**: 1653–1657.

Pix, W., Nalbach, G. and Zeil, J. (1993) Strepsipteran forewings are haltere-like organs of equlibrium. *Naturwissenschaften* **80**: 371–374.

Raff, R.A.(1996). *The Shape of Life: Genes, Development and the Evolution of Animal Form.* University of Chicago Press, Chicago.

Tarrio, R., Rodriguez-Trelles, F. and Ayala, F.J. (1998) New *Drosophila* introns originate by duplication. *Proc Natl Acad Sci USA* **95**: 1658–1662.

Whiting, M.F. and Wheeler, W.C. (1994) Insect homeotic transformation. *Nature* **368**: 696.

Whiting, M.F., Carpenter, J.C., Wheeler, Q.D. and Wheeler, W.C. (1997) The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from the 18S and 28S ribosomal sequences and morphology. *Syst Biol* **46**: 1–68.