Dev Genes Evol (1998) 208:547–557

ORIGINAL ARTICLE

Michael D. Peterson · Aleksandar Popadić
Thomas C. Kaufman

The expression of two *engrailed*-related genes in an apterygote insect and a phylogenetic analysis of insect *engrailed*-related genes

Received: 4 May 1998 / Accepted: 2 August 1998

Abstract Homologues of the *Drosophila* segment polarity gene *engrailed* have been cloned from many insect species, as well as other arthropods and non-arthropods. We have cloned partial cDNAs of two *engrailed* homologues, which we call *engrailed*-related genes, from the phylogenetically basal insect, *Thermobia domestica* (Order Thysanura) and possibly as many as four *engrailed*-related genes from the phylogenetically intermediate insect, *Oncopeltus fasciatus* (Order Hemiptera). Previous to our findings, only single *engrailed*-related homologues had been found in phylogenetically intermediate insect species (*Tribolium* and *Schistocerca*) and in the crustacean *Artemia*, while two *engrailed*-related homologues have been found in more derived orders (Hymenoptera and the *engrailed* and *invected* genes of lepidopterans and dipterans). Consequently, we performed a phylogenetic analysis of insect *engrailed*-related genes to determine whether insects ancestrally had one or two *engrailed*-related genes. We have found evidence of concerted evolution among *engrailed*-related paralogues, however, that masks the true phylogenetic history of these genes; the phylogeny may only be decipherable, therefore, by examining the presence or absence of *engrailed*-specific and *invected*-specific motifs, which will require cloning the full length cDNAs from more species. In addition, we examined the embryonic expression pattern of the two *Thermobia engrailed*-related genes; like *Drosophila engrailed* and *invected*, they are expressed in very similar patterns, but show one temporal difference in pregnathal segments that correlates with the tentative phylogenetic placement of the genes. *Thermobia engrailed*-related expression also confirms that the dorsal ridge is an ancient structure in insects.

Key words *engrailed* · *Thermobia domestica* · *Oncopeltus fasciatus* · Dorsal ridge

Edited by D. Tautz

M.D. Peterson · A. Popadić · T.C. Kaufman(✉)
Howard Hughes Medical Institute at Indiana University
at Bloomington, Bloomington, IN 47405, USA

Introduction

The *engrailed (en)* gene in *Drosophila melanogaster* is a segment polarity gene, a class of genes required for proper intrasegmental patterning during embryogenesis (Pankratz and Jäckle 1993). The En protein is a homeoprotein transcription factor (Jaynes and O'Farrell 1991; Han and Manley 1993) that interacts cooperatively with the homeoprotein cofactor Extradenticle to bind DNA target sequences (Peltenburg and Murre 1996). In its role as a segment polarity gene in *Drosophila* embryos, *en* is expressed in a lateral stripe in the posterior portion of each segment and is required for proper segmentation (DiNardo et al. 1985; Fjose et al. 1985; Kornberg et al. 1985). Reiterated expression of *en* homologues in the posterior region of each segment has also been observed in other insects (Patel et al. 1989a; Fleig 1990; Brown et al. 1994; Schmidt-Ott et al. 1994; Rogers and Kaufman 1996), other arthropod classes (Patel et al. 1989b; Manzanares et al. 1993; Scholtz et al. 1994; Scholtz 1995), an onychophoran (Wedeen et al. 1997) and an annelid (Wedeen and Weisblat 1991), suggesting that *en* may have an ancient role in protostome segmentation.

In *Drosophila*, there are two *engrailed*-class paralogues: *engrailed* and *invected* (*inv*; Coleman et al. 1987; Gustavson et al. 1996). They reside next to each other on the second chromosome and share a common enhancer region (Gustavson et al. 1996). For the most part, they have similar overlapping expression patterns and overlap in function, though *inv* mutations are not lethal, while *en* mutations are (Gustavson et al. 1996). One notable difference is that *en* is expressed earlier than *inv* in *Drosophila* embryos.

Likewise, in the moth, *Bombyx mori*, there are two *en*-class paralogues (Hui et al. 1992). By comparing the sequence of the two *Bombyx* genes with *Drosophila en* and *inv*, Hui et al. (1992) discovered two *en*-specific motifs in one *Bombyx* paralogue and two *invected*-specific motifs in the other, in addition to four conserved regions shared among all *en*-class genes. Due to the *en*-specific and *inv*-specific motifs and a similarity in exon-intron

structure among all four genes, they proposed that a duplication of a single *en*-class gene had occurred in an ancestor of flies and butterflies. Accordingly, they named the *Bombyx* genes *en* and *inv*, proposing them to be direct orthologues of the *Drosophila* genes.

Two *en*-class paralogues *(En-1 and En-2)* also exist in mice, chickens and human beings (Joyner and Martin 1987; Logan et al. 1992), but they do not have the *en*-specific and *inv*-specific domains of the higher insect genes. Single *en*-class homologues in sea urchins (Dolecki and Humphreys 1988), the amphioxus *Branchiostoma floridae*, a cephalochordate (Holland et al. 1997), the beetle *Tribolium castaneum* (Brown et al. 1994), the grasshopper *Schistocerca americana* (Patel et al. 1989b), and the brine shrimps, *Artemia franciscana* and *A. parthenogenetica* (Manzanares et al. 1993), have led to the hypothesis that ancestral metazoans had a single *en*-class homologue and that independent duplications have occurred in vertebrates and higher insects (Dolecki and Humphreys 1988; Manzanares et al. 1993; Brown et al. 1994).

If, indeed, there was only a single ancestral metazoan *en*-class homologue, then *en*-class genes have been prone to duplication among metazoans. In addition to the duplicate paralogues that exist in amniote vertebrate species and higher insects, three *en*-like genes, *eng-1*,*-2*, and *-3*, have been found in zebrafish (Ekker et al. 1992). Furthermore, at least two independent duplications of *en*-class genes have occurred among barnacles (Gibert et al. 1997) and two *en*-class homologues have been cloned from the honeybee, *Apis mellifera* (Walldorf et al. 1989). Because of the plethora of *en*-class genes that exist among metazoans, in this report we call all *en*-class homologues *"engrailed*-related*"* (*en*-r) genes and let *"engrailed"* and *"invected"* refer to the actual higher insect genes only.

Since the evidence that a single *en*-r gene existed in the insect ancestor rests on a small sampling of insects, we cloned *en*-r partial cDNAs from representatives of two insect orders not sampled before, Hemiptera (the milkweed bug, *Oncopeltus fasciatus*), an intermediate insect taxon, and Thysanura (the firebrat, *Thermobia domestica*), a basal apterygote lineage (Kristensen 1991). Unexpectedly, we recovered two different *en*-r partial cDNAs from the firebrat, which raised the question of whether these genes are direct orthologues of *en* and *inv*, respectively, or the result of independent duplications. Four highly similar partial cDNAs were recovered from the milkweed bug that vary mainly in a region encoded by a microexon in other insect species. A phylogenetic analysis of insect genes found evidence for concerted evolution between *en*-r paralogues, making it difficult to determine the phylogeny of *en*-r genes in insects. We also examined the embryonic expression patterns of the two *Thermobia en*-r genes via in situ hybridization to determine whether they have similar overlapping expression patterns like the *Drosophila* paralogues.

Materials and methods

Arthropod colonies, embryo collection and in situ hybridization

Care for laboratory populations and embryo collection of milkweed bugs, firebrats and millipedes has been described in Rogers and Kaufman (1996), Rogers et al. (1997) and Popadić et al. (1998), respectively. The in situ hybridization protocol used was described in Rogers et al. (1997). Protocols are available upon request.

RT-PCR cloning of *en*-r clones

RT-PCR (reverse transcription of RNA, followed by polymerase chain reaction) was used to amplify partial *en*-r cDNAs for cloning and sequencing. RNA was prepared using the TriZol reagent (Gibco), according the manufacturer's instructions. Reverse transcription was performed using the GeneAmp (Perkin Elmer Cetus) reagents, following its protocol, except that the incubation at 42°C was done for 90 min.

For the PCR, we used codon-degenerate primers targeting the highly conserved amino acid sequences WPAWVYC (forward primer) and MAQGLYN (reverse primer). WPAWVYC=5' TGG CCN GCN TGG GTN TAY TGY 3'; MAQGLYN=5' RTT RTA NAR NCC YTG NGC CAT 3', using the IUPAC symbols. Two rounds of PCR were performed to amplify the partial cDNAs for cloning, according the GeneAmp protocol with each primer at 1 μM. For the first five cycles, the hybridization temperature was 37°C, followed by a 1 min ramp to the extension temperature (72°C). The remaining cycles used a hybridization temperature of 50°C and had no extended ramp time. All extension times were 30 s. The primary PCR product was run out on a 4% agarose gel. The properly sized band was then touched with a toothpick, which was then touched to a secondary PCR mix, with conditions as specified in the GeneAmp protocol.

Thirty firebrat *en*-r clones were recovered and sequenced over two independent PCR trials. Of the 30 firebrat clones 28 yielded the same nucleotide sequence, with the exception of one silent polymorphism. This clone was named *Td-en*-r1. The two remaining clones were identical to each other and differed significantly from *Td-en*-r1 at both the nucleotide and inferred amino acid sequence level; this sequence variant was named *Td-en*-r2. Additional *Td-en*-r2 partial cDNA clones were recovered using exact primers on an independently generated firebrat cDNA pool. Thirty-three milkweed bug (*Oncopeltus fasciatus*) *en*-r clones were recovered and sequenced from two PCR trials. Four different types were found (Fig. 1B); the first two were named in accordance with the firebrat genes. Eleven identical *Oxidus en*-r clones were recovered and sequenced from a single PCR trial.

Phylogenetic analysis of *engrailed*-related genes

The sequences used in the phylogenetic analysis of *engrailed*-related genes were the highly conserved regions from domains II, III and the homeodomain that lie between the *en*-specific primers used to clone firebrat and milkweed bug *en*-r genes (Fig. 2). These conserved regions were labeled A, B, C, D and E, as shown in Fig. 2, and entered as a contiguous sequence into the phylogeny programs. Only two milkweed bug genes (*Of-en*-r1 and 2) were used in the analysis, as *Of-en*-r1, 3 and 4 are identical in all but region B. *Artemia en*-r was left out of the final analysis as *Artemia* sequences are known to be problematic in phylogenetic analyses (Field et al. 1988; Aguinaldo et al. 1997). Gap characters introduced by region B were specified using the default modes for each program and all characters were equally weighted. For the cDNA-based phylogenies, only the first two nucleotides of each codon were used in order to eliminate saturation effects. All sequences used for phylogenetic analysis (other than those cloned in this paper) were obtained from GenBank and aligned by eye according to the protein alignment of Fig. 2. For analysis of the phylogenetic

relationships of full-length insect *en*-r genes, protein sequences were used with all characters weighted equally. Phylogenetic analyses were performed using the neighbor joining (NJ) algorithm in the PHYLIP software package (Felsenstein 1993) and the maximum parsimony method using the Phylogenetic Analysis Using Parsimony (PAUP) software package (Swofford 1993). The PAUP trees were generated using the branch-and-bound search algorithm. Four hundred data sets were analyzed for the PAUP bootstrap analysis using the branch-and-bound algorithm. For bootstrap analysis of the NJ trees, the SEQBOOT program (in PHYLIP) was used to produce 100 bootstrapped data sets. Then, the DNADIST or PROTDIST, NEIGHBOR and CONSENSE programs (PHYLIP) were used in succession to produce the bootstrap values.

Results and discussion

Firebrat and milkweed bug *engrailed*-related genes

Two highly conserved amino acid motifs present in *en*-r homologues were used to design degenerate oligonucleotide primers for RT-PCR in order to amplify a partial *en*-r cDNA from the milkweed bug, *Oncopeltus fasciatus* (Hemiptera), and the firebrat, *Thermobia domestica* (Thysanura). PCR on both firebrat and milkweed bug embryonic cDNA with these primers yielded a strong

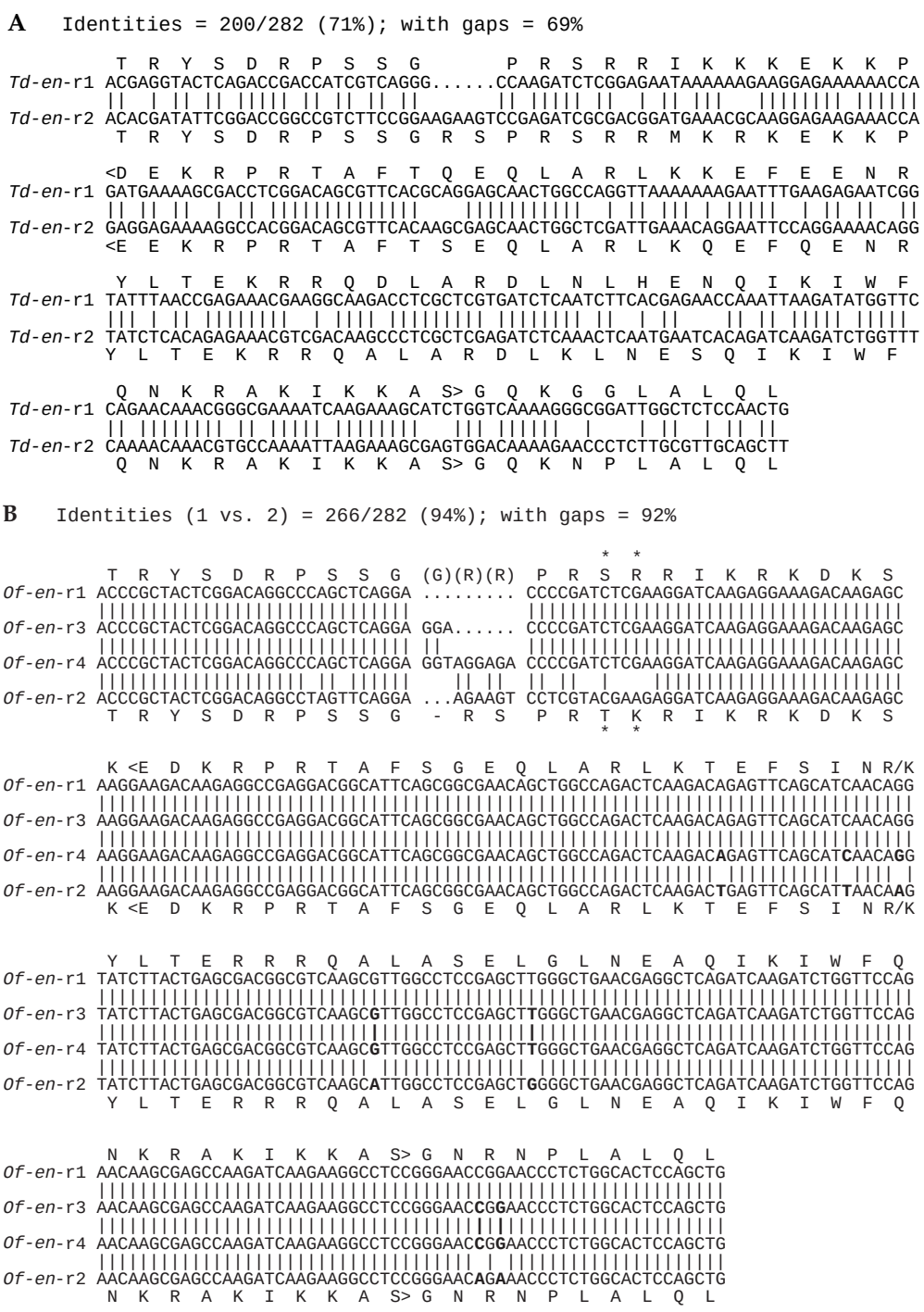

Fig. 1 **A** Nucleotide alignment of *Thermobia engrailed*-related (*en*-r) genes. The predicted protein sequence of each gene is shown *above* (*Td-en*-r1) or *below* (*Td-en*-r2) the nucleotide sequence. The homeodomain is marked by *right and left carrots*. Identical nucleotides are marked by *vertical lines*. Gaps for spacing are indicated by *dots*. The nucleotide percent identity between the two genes, including and excluding the gap, is shown. **B** Nucleotide alignment of *Oncopeltus en*-r genes. Display and symbols are presented as in **A**. In addition, protein residues that differ between *Of-en*-r2 and the other three genes are marked with an *asterisk*. Downstream of the start of the homeodomain, polymorphic sites among the clones are shown in *bold*. No polymorphisms were found among *Of-en*-r1 clones. Two sequence types were found in *Of-en*-r2 and *Of-en*-r4 clones, which differed by six silent polymorphisms and one amino-acid changing variant (R versus K). The first of these sequence types is the sequence shown for *Of-en*-r4; the second is the sequence shown for *Of-en*-r2. Both types were found for both clones, so that the only consistent difference between *Of-en*-r2 and the other genes are the differences upstream of the homeodomain. Lastly, *Of-en*-r3 had two sequence variants: the one shown (equal to *Of-en*-r4 variant shown) and a second that had the last four polymorphic differences of the *Of-en*-r2 sequence type shown. The nucleotide percent identity between the most different alleles of Of-en-r1 and Of-en-r2 is also shown

band at ~330 bp in each case (not shown). From firebrats, two different *en*-r clones were recovered that were 69% identical at the nucleotide level; they were named *Td-en*-r1 and -2 (Fig. 1A). A primary difference between the firebrat genes is the presence (*Td-en*-r2) or absence (*Td-en*-r1) of an arginine-serine (RS) dipeptide. As shown in the protein alignment of Fig. 2, the *Bombyx* and *Drosophila en* genes also lack the RS dipeptide sequence, whereas it is present in *Bombyx* and *Drosophila inv* (Coleman et al. 1987; Hui et al. 1992). Interestingly, the RS motif is encoded by a six-nucleotide microexon in the *inv* genes and also in the single *Tribolium en*-r gene (Brown et al. 1994).

Four nearly identical milkweed bug variants were recovered from 33 clones; they differed primarily in the sequence of the "RS region" (Fig. 1B). In accordance with the firebrat nomenclature, the milkweed bug genes lacking and possessing the RS dipeptide were named *Of-en*-r1 and –2, respectively. The other two variants have novel motifs in this region. *Of-en*-r3 encodes only a glycine and *Of-en*-r4 encodes a glycine-arginine-arginine (GRR) tripeptide. Aside from these differences, however, the milkweed bug clones were very similar at the nucleotide level (Fig. 1B).

In fact, the degree of sequence identity of the milkweed bug clones raises caution as to whether they originated from different genes. It is possible that these transcripts arose from the same gene, with alternative splicing accounting for the different sequences in the RS region. Our first *Of-en*-r2 and *Of-en*-r4 clones had 21 nucleotide differences (out of the 291), but subsequent clones revealed the (7) differences in the latter two-thirds of the partial cDNA to be the result of a polymorphic sequence variant that appears in both *Of-en*-r2 and -4. The seven polymorphic positions are shown in bold in Fig. 1B; one type is shown for *Of-en*-r2 and the other for *Of-en*-r4, but clones with both sequence variants were recovered for both genes. Many fewer *Of-en*-r1 and -3 clones were recovered (3 each out of 33 total clones); all *Of-en*-r1 clones were identical, while one *Of-en*-r3 clone showed differences at only four of the seven polymorphic positions seen in the *Of-en*-r2 and -4 allelic variants (Fig. 1B). Thus, *Of-en*-r1, -3 and -4 share one clone sequence type that is identical among all three clones, as shown in Fig. 1B. They may all come from the same locus, with the differences between them accounted for by alternative splicing and polymorphism. On the other hand, *Of-en*-r2 has some unique nucleotides – some of which encode for different amino acids – around the putative microexon region. If these clones arose from separate genes, there has been a high level of sequence homogenization between the loci. This question could be resolved by cloning and characterizing genomic copies of the milkweed bug *en*-r gene(s).

The conceptually translated sequences of these partial cDNAs are shown in an alignment with other *en-r* homologues in Fig. 2. The region encompassed by the PCR primers contains three conserved domains: domains II and III (after Hui et al. 1992) and the homeodomain. All residues in the homeodomain critical for DNA binding are conserved in all milkweed bug and firebrat genes (Kissinger et al. 1990).

Peltenburg and Murre (1996) demonstrated that domain II (of mouse En-2) is both necessary and sufficient for interaction with the Pbx proteins, which are homologues of the *Drosophila* Extradenticle (Exd) protein. The *exd* gene of *Drosophila* encodes a divergent homeodomain-containing protein that is necessary for modulating the specificity of binding of HOM-C proteins and En (van Dijk and Murre 1994). Specifically, they found that the interaction occurs through the N-terminal half of domain II, which they called domain EH2, approximately equivalent to region "A" in Fig. 2. In particular, the two tryptophan residues in this region were shown to be required for forming a Pbx-En-DNA cooperative binding complex (Peltenburg and Murre 1996). EH3 (from region "C" in Fig. 2 to the start of the homeodomain) was found to be important as well; variations in the length of the EH3 domain decreased the cooperative binding of En with Pbx and a DNA binding target.

The RS dipeptide motif lies within domain II; it is present in the *inv* genes, the *Tribolium en*-r gene, the single *en-r* homologues of the grasshopper *Schistocerca americana* (Patel et al. 1989a), and the crustacean brine shrimp *Artemia franciscana* (Manzanares et al. 1993). In *Artemia*, the intron position on the 3' side of the RS-encoding residues is conserved (Manzanares et al. 1993). These findings prompted Brown et al. (1994) to postulate that the RS dipeptide was present in a single ancestral insect *en*-r homologue and was subsequently lost in the (true) *en* gene of higher insects (lepidopterans and dipterans), while it was maintained in the *inv* genes.

The RS dipeptide is not present in chordates or other non-arthopod *en*-r genes (Dolecki and Humphreys 1988; Logan et al. 1992; Webster and Mansour 1992; Holland et al. 1997). Its conservation among arthropods implies that it has an important function, but this has not been tested. As the murine En proteins lack the RS motif in domain II, it is not necessary for the interaction of the Pbx proteins and En-2 that was demonstrated by Peltenburg and Murre (1996).

In order to determine the point of origin of this dipeptide motif and determine whether its presence is an ancestral feature of arthropod *en*-r genes, it will be necessary to clone domain II from other arthropods and non-arthropods. Other groups have cloned *en*-r gene fragments from other arthropod species (Gibert et al. 1997), molluscs (Wray et al. 1995), an onychophoran (Wedeen et al. 1997) and an annelid (Wedeen et al. 1991), but none included domain II. A *Caenorhabditis elegans en*-r homologue does not have the RS dipeptide (GenBank accession no. L14730) but, as with some other *C. elegans* sequences (Fitch et al. 1995; Aguinaldo et al. 1997), its sequence is highly divergent and, thus, may not be a good representative of a non-arthropod protostome. Consequently, we cloned an embryonically expressed *en*-r partial cDNA from the millipede *Oxidus gracilis*. We recovered only one *Oxidus en*-r gene with

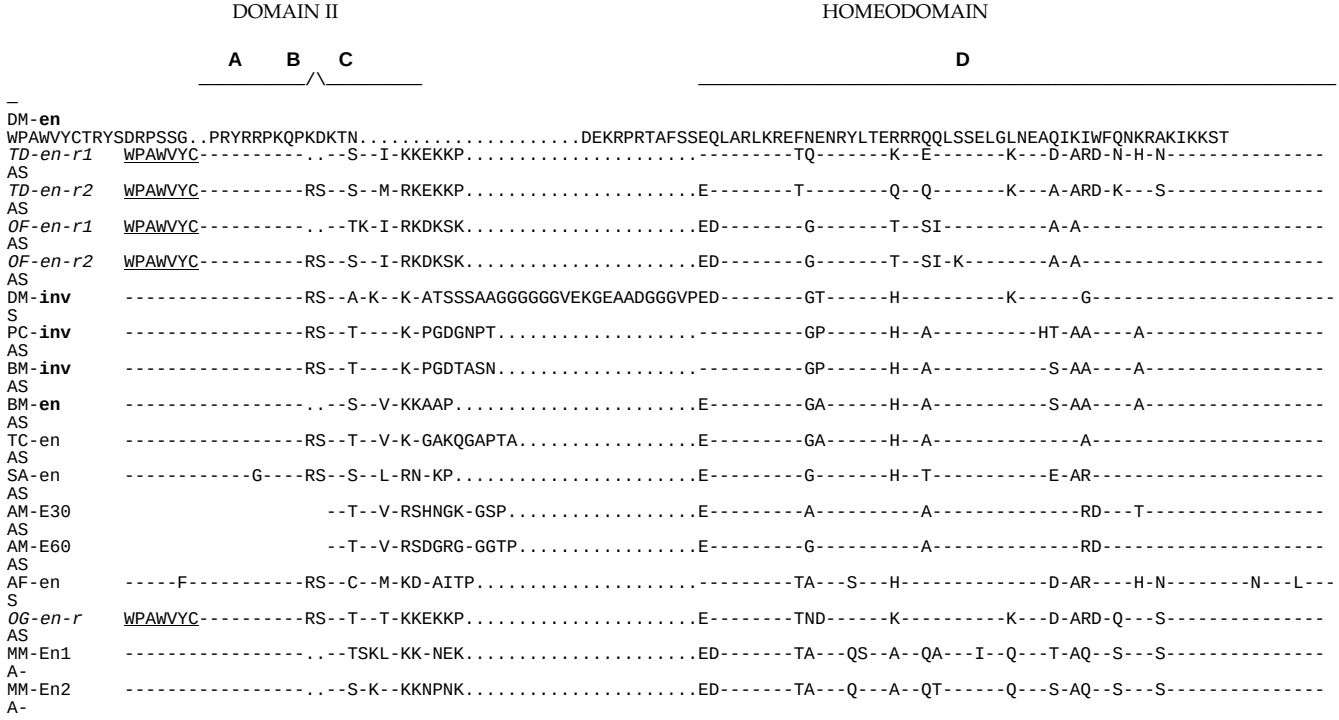

Fig. 2 The amino acid alignment of multiple *engrailed*-related homologues is shown. The *Thermobia, Oncopeltus* and *Oxidus engrailed*-related partial cDNAs were cloned using primers corresponding to the underlined sequences. These clones encompass Domains II and III and the homeodomain. *Dashes* represent amino acids identical to *Drosophila engrailed*. *Periods* are used as gaps for sequence alignment. The genes cloned for this report are *italicized* and the true "*engrailed*" and "*invected*" genes are in *boldface*. Conserved portions between the primers that can be aligned among all homologues were divided into five regions (*A,B,C,D,E*) for phylogenetic analysis. Only *Of-en*-r1 and –2 are shown, as *Of-en*-r3 and -4 are identical to *Of-en*-r1, except in region B [*DM Drosophila melanogaster* (Order Diptera), *TD Thermobia domestica*, firebrat (Thysanura), *OF Oncopeltus fasciatus*, milkweed bug (Hemiptera), *PC Precis coenia*, butterfly (Lepidoptera), *BM Bombyx mori*, moth (Lepidoptera), *TC Tribolium castaneum*, beetle (Coleoptera), *SA Schistocerca americana*, grasshopper (Orthoptera), *AM Apis mellifera*, honeybee (Hymenoptera), *OG Oxidus gracilis*, millipede (Class Myriapoda), *AF Artemia franciscana*, brine shrimp (Class Crustacea), *MM Mus musculus*, mouse]

RT-PCR; no further attempt was made to determine whether more *en*-r genes exist in *Oxidus*. Like *Artemia en*-r, it possesses the RS dipeptide in domain II (Fig. 2). Therefore, it appears very likely that the RS motif in domain II is an ancestral feature to insects and evolved before the divergence of myriapods, crustaceans and insects.

Phylogeny of insect *engrailed*-related genes

Prior to our findings, duplicate *en-r* genes among insects had been reported only in higher insects, the Lepidoptera, Hymenoptera and Diptera (Coleman et al. 1987; Walldorf et al. 1989; Hui et al. 1992). The finding of two paralogues in firebrats and, at the very least, RS+ and RS– variants in milkweed bugs, raises the question as to whether a single *en*-r duplication occurred before the radiation of the insects. To assay the gene phylogeny, we performed a phylogenetic analysis of insect *en*-r genes using maximum parsimony (PAUP; Swofford 1993) and distance methods (Neighbor Joining; Saitou and Nei 1987), as described in the Materials and methods.

The topology supported by majority-rule bootstrap analysis using the maximum parsimony (MP) algorithm is shown in Fig. 3A. Analysis with the Neighbor Joining (NJ) algorithm derived a similar topology (not shown). For both analyses, *Oxidus gracilis en*-r was defined as the outgroup. Bootstrap values were generated for each node to test their strengths. With the exception of the firebrat genes, the *en*-r paralogues group according to their host species, not by orthologue group. A literal interpretation of the tree postulates that ancestrally in insects there were two *en*-r genes, one of which gave rise

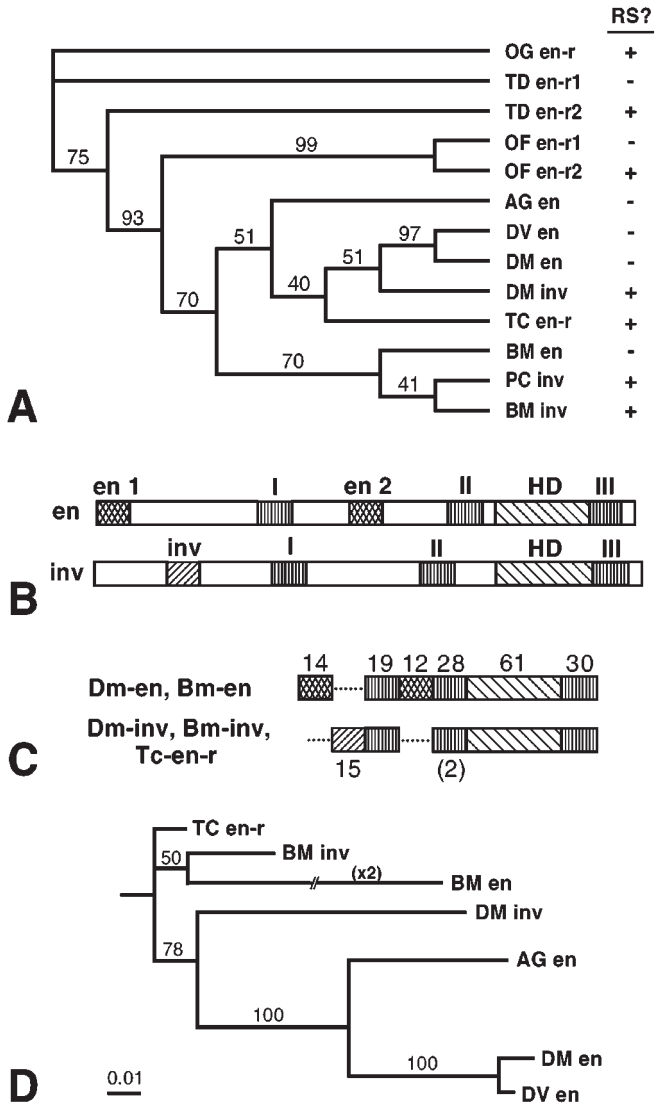

Fig. 3A–D Phylogeny of insect *en*-r genes. **A** Phylogeny of insect *engrailed*-related genes based upon the ABCDE partial cDNA alignment. The most parsimonious phylogeny of the insect *engrailed*-related genes is shown with bootstrap values for each node. Nodes with bootstrap values below 50 are shown, but should be considered unresolved. The presence (+) or absence (–) of the RS dipeptide of en-r domain II is shown next to each gene. **B** Diagram of the conserved domains within fly/butterfly *en* and *inrected* (*inv*) genes. Domains I, II, III and the HD (*homeodomain*) are present in all *en*-r genes. The *en*1, *en*2 and *inv* domains are orthologue-specific domains. **C** Diagram of the alignment of conserved domains used for phylogenetic analysis of full length proteins in **D**, using the same shading as in **B**. *Gaps* were put in for missing orthologue-specific domains. *Tribolium (Tc) en*-r has the *inv*-specific domain but neither *en*-specific domain. The size of the domain in amino acids is shown. **D** Neighbor joining phylogeny of full length En-r protein sequences with bootstrap values. Branch lengths between nodes and taxa are drawn to proportion, in accordance with the distance key shown. *TD Thermobia, OF Oncopeltus, DM Drosophila BM Bombyx, TC Tribolium castaneum, DV D. virilis, AG* the mosquito *Anopheles gambiae, PC* the butterfly *Precis coenia, OG* the millipede *Oxidus gracilis*

to *Td-en*-r1, and the other to *Td-en*-r2. After the separation of pterygotes and apterygotes, the *Td-en*-r1 ancestor was lost in the pterygote lineage. Subsequent to this, the pterygote *Td-en*-r2 orthologue was duplicated three times: in the lineage leading to milkweed bugs and separately in the lineages leading to dipterans and lepidopterans after the split of these groups.

Only one aspect of this evolutionary scenario, the monophyletic grouping of the milkweed bug genes, is strongly supported, however. The very high bootstrap value for this grouping strongly suggests, as does a simple visual comparison of the sequences, that these two genes are either the result of a recent duplication or a very high degree of sequence homogenization. The node that splits the firebrat genes and suggests that the *Td-en*-r2 ancestor gave rise to all pterygote *en*-r genes is not as strongly supported, having a bootstrap value of 75 by MP and 77 by NJ (not shown), which is only moderate support for splitting the firebrat genes. Also, trees of only one step longer in the parsimony analysis grouped the firebrat genes together. Thus, support for the braching order of the firebrat genes is weak.

The separate grouping of lepidopteran and dipteran genes contradicts the hypothesis of Hui et al. (1992), which is based on exon-intron structure and the presence or absence of four orthologue-specific motifs. Thus, to further test that hypothesis we did a phylogenetic analysis of the insect *en*-r genes for which an entire protein sequence is available in sequence databases, namely *Tribolium en*-r, *en* and *inv* from *Bombyx* and *D. melanogaster*, and *en* from *D. virilis* and *Anopheles*. The peptide sequences of all conserved domains of each gene were joined together to create an input file for the NJ program, as shown in Fig. 3B,C. Gaps were put in for missing orthologue-specific domains. However, using all conserved domains also failed to corroborate the single duplication hypothesis for dipteran and lepidopteran *en*-r genes (Fig. 3D). Even so, the additional domains resolved the branching order within the dipteran gene clade into a monophyly of the *en* genes.

If we accept the Hui et al. (1992) hypothesis as true on intuitive grounds, then there has been concerted evolution between *en*-r paralogues in both dipterans and lepidopterans that has resulted in sequence homogenization, such that the *en* and *inv* paralogues of a given species appear more similar in sequence to each other than they do to their true orthologues in a different species. Our results show that the strength of the concerted evolution is strong enough to be seen between insect orders, but is too weak to be seen within an order. Gene conversion is one possible cause of the homogenization effect (Dover 1986). While this cannot be ruled out as having occurred in either *Drosophila* or *Bombyx*, there are still significant differences between their *en* and *inv* genes. For instance, the size of the linker connecting domain II and the homeodomain differs dramatically between the *Drosophila* genes (Fig. 2). Another interesting possibility is that as long as both *en* and *inv* retain similar and at least partially overlapping functions, as they have in

Drosophila (Tabata et al. 1995; Gustavson et al. 1996), there will be selection for sequence covariation, because not only might they have to bind to identical DNA target sequences, but they will presumably interact with a common cofactor, Exd, for proper function (van Dijk and Murre 1994). Although the *Drosophila* Inv protein has not yet been shown to interact with Exd, this coevolutionary constraint hypothesis explains why we see homogenization in domain II (the putative Exd interaction domain) and the homeodomain (the DNA-binding domain). On the other hand, it does not explain the divergence in size of *Drosophila en* and *inv* in the region between domain II and the homeodomain, a difference that could affect interaction with Exd (Peltenburg and Murre 1996).

Sequence homogenization between paralogues, by whatever means, presents a barrier to computational analysis of the phylogeny of *en*-r genes. With only domains II, III and the homeodomain, it is impossible to distinguish between independent duplication and concerted evolution. The paraphyletic split between the firebrat genes may indicate that some resolution is possible, due perhaps to a lower rate of sequence homogenization in firebrats. Under the circumstances, however, it may be that resolution of the *en*-r phylogeny will come only from a qualitative analysis of the presence or absence of orthologue-specific motifs, which will only be possible if those domains evolved before the radiation of insects and are maintained in most insects. At present, we know that the N-terminal *inv*-specific motif is conserved in *Tribolium* (Brown et al. 1994), that the N-terminal-most *en*-specific motif is partially conserved in *Artemia* (Manzanares et al. 1993) and that the RS dipeptide motif is conserved in myriapods, crustaceans and insects. Using the RS motif, *Td-en*-r1 and *Of-en*-r1 appear to be *en* orthologues and *Td-en*-r2 and *Of-en*-r2 *inv* orthologues. Whether the longer orthologue-specific motifs will be in agreement with this preliminary assessment remains to be seen, particularly for the highly homogenized milkweed bug sequences. A problem with this preliminary assessment is that the RS dipeptide would be very easy to lose if it is encoded by a microexon in all insects; a single mutation in the splice acceptor site, for example, would eliminate it completely. In this light, it will be important to determine what differential function, if any, the RS dipeptide imposes on *en*-r genes.

In conclusion, with the present data set, it is not possible to determine whether there were one or two *en*-r genes in the insect ancestor. On either hypothesis, the evolutionary history of *en-r* genes in insects has been dynamic, as it has been among cirripede crustaceans (Gibert et al. 1997). Our findings of two *en*-r genes in both the firebrat, a phylogenetically basal insect, and perhaps more than two in the milkweed bug, raise the possibility that two *en-r* genes may have existed in the insect ancestor, not one as has been thought (Brown et al. 1994). Longer cDNAs from more taxa will be needed to determine the dynamics of *en*-r gene evolution in insects and other arthropods with greater clarity. The problem of concerted evolution of paralogues may reduce the analysis to determining whether the *inv*-specific or *en*-specific domains are present in order to decipher the lineage of these genes.

Embryonic expression of firebrat *en*-r genes

The monoclonal antibody 4D9, which binds a highly conserved epitope in the homeodomain of En-r proteins (Patel et al. 1989b), has been used on insects of numerous orders and on other arthropod classes to examine the expression of *en*-r genes (e.g., Patel et al. 1989b; Fleig 1990; Brown et al. 1994; Scholtz et al. 1994; Rogers and Kaufman 1996). Our clones reveal that the 4D9 epitope is present in all *Of-en*-r genes and, therefore, the accumulation pattern of En-r proteins in milkweed bug embryos described by Rogers and Kaufman (1996) is probably a composite pattern of all encoded proteins. Single amino acid differences in both *Td-en*-r1 and *Td-en*-r2 (a Gly to Asn change in *Td-en*-r1 and a Gly to Lys in *Td-en*-r2 at residue 40 of the homeodomain) have modified the 4D9 epitope such that neither is recognized by the antibody. Thus, we analyzed the expression of the *Td-en*-r genes in firebrat embryos via whole-mount in situ hybridization using the *Td-en*-r1 and the *Td-en*-r2 partial cDNAs.

Figure 4 shows the expression patterns of *Td-en*-r1 and *Td-en*-r2 in firebrat embryos at early germ band elongation and at the end of elongation, near the start of dorsal closure. All embryos are shown ventral side up, except in Fig. 4A,D. The ventral side of these embryos is attached to the chorion and they are the youngest embryos that can be recovered by manual dissection. At this stage, the embryos have recently undergone germ condensation from the blastoderm stage, a mesoderm layer has formed and the elongating embryos are three to four cell layers thick (Woodland 1957). The embryos in Fig. 4A,D share in common *en*-r expression in the posterior region of the antennal, mandibular, maxillary and labial segments. *en*-r1 (Fig. 4A) is also expressed in the intercalary segment, while *en*-r2 (Fig. 4D) transcripts show faint accumulation in the primordium of the ocular segment. This is the most obvious difference in expression between the two genes. The *en*-r2 ocular spots form before the thoracic stripes, whereas *en*-r1 ocular expression is not present until after the abdominal stripes begin to appear (Fig. 4B and C; Peterson 1998). Conversely, the *en*-r1 intercalary stripe forms before the *en*-r2 intercalary stripe (Fig. 4A,D).

In *Drosophila*, *Ctenocephalides* (flea), *Oncopeltus*, *Acheta* (cricket) *Tribolium* and *Schistocerca*, intercalary expression is not established until after the abdominal stripes begin to appear and only in *Schistocerca* do the ocular spots accumulate before the abdominal stripes (Patel et al. 1989a; Schmidt-Ott and Technau 1992; Brown et al. 1994; Rogers and Kaufman 1996). Furthermore, for all these species, ocular expression precedes intercalary expression. The firebrat *en*-r genes, on the

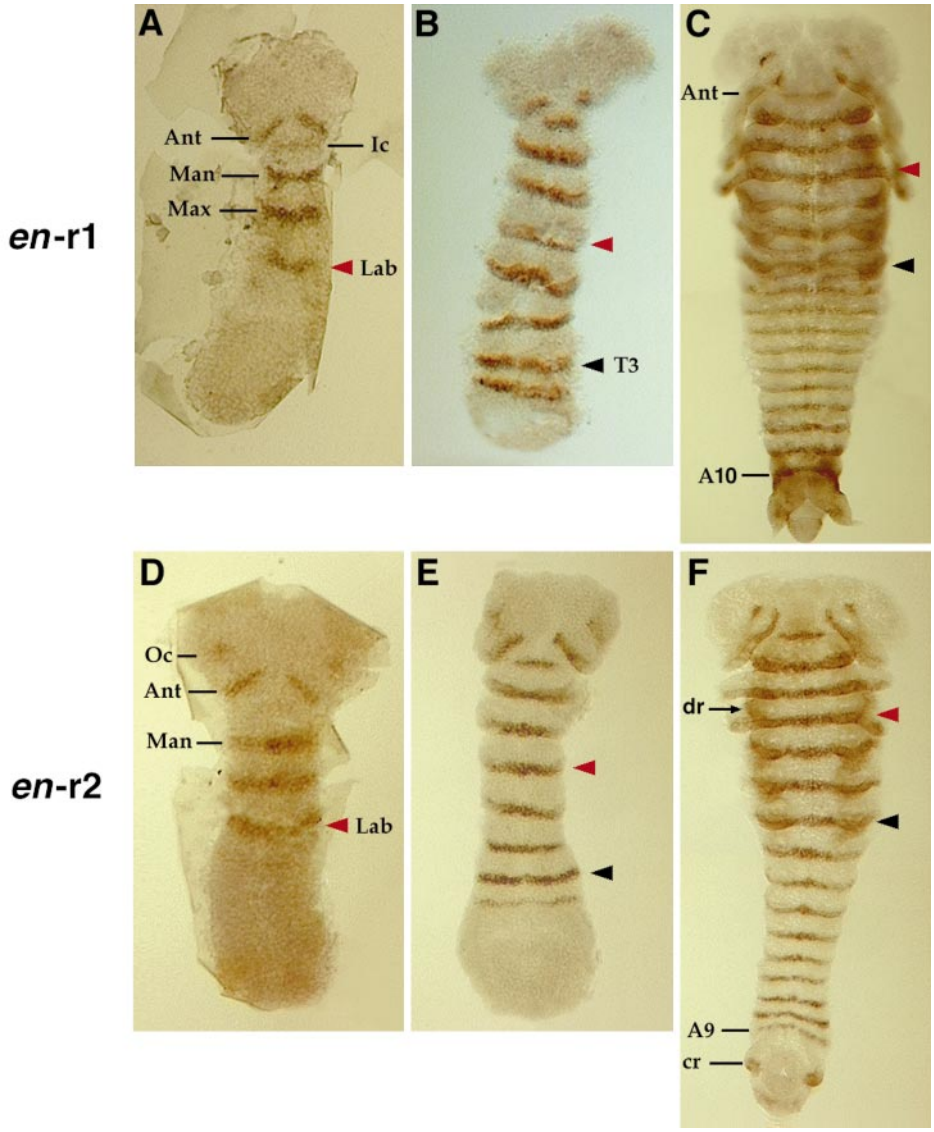

Fig. 4A–F Embryonic expression of *Td-en*-r1 (**A–C**) and *Td-en*-r2 (**D–F**). At the earliest stage at which firebrat embryos can be recovered, both *en*-r1 (**A**) and *en*-r-2 (**B**) are expressed in the antennal (*Ant*) and gnathal segments *Man* (mandibular, *Max* maxillary, *Lab* labial). They differ in that *en*-r1 is also expressed in the intercalary (*Ic*) segment, while *en*-r2 is expressed in the ocular (*Oc*) segment. Ocular *en*-r1 expression is not present until after the abdominal segments appear (**B,C**). Intercalary *en*-r2 expression, however, appears before the first abdominal segments (**E**). During late germ band elongation (**F**) and after (**C**), six head domains are present, *en*-r expression in the dorsal ridge (*dr*) is apparent and both genes are expressed in the cerci (*cr*). *Red arrowheads* mark the labial segment and *black arrowheads* mark the third thoracic segment (*T3*)

other hand, differ from this pattern and from each other. What can we make of these differences in a phylogenetic context? First, the earlier expression of the intercalary and ocular domains in firebrats may reflect a simpler, more ancestral pattern, one closer to a strict anterior-to-posterior delineation of stripes. Second, in malacostracan crustaceans, *en*-r expression in the second antennal segment, the probable homologue of the insect intercalary segment (Tamarelle 1984), precedes that of the ocular segment (Scholtz 1995), the opposite of what is seen in pterygote insects. Thus, it is interesting that *Td-en*-r2 expression is more similar to *en*-r in pterygote insects on this basis, while *Td-en*-r1 expression resembles that in malacostracan crustaceans, a similarity that mirrors the sister grouping of *Td-en*-r2 with pterygote *en*-r genes and the outgroup placement of *Td-en*-r1 in the phylogenetic analysis.

The order of initiation among the antennal and gnathal *Td-en*-r stripes is unknown, because of the difficulty of obtaining embryos younger than those in Fig. 4A,D (bleach dechorionation does not work). The thoracic and abdominal *en*-r stripes arise one after the other in an anterior to posterior fashion (Fig. 4; Peterson 1998), as seen in other short-germ insects. Metameric constrictions demarcating the segment boundaries become apparent in the gnathal segments just after the embryo detaches from the chorion. No parasegmental compartment grooves are ever observed. Nor is any splitting of expression into secondary domains observed, such as the secondary antennal or ocular head spots present in many dipterans, *Tribolium* (Schmidt-Ott et al. 1994) and milkweed bugs (Rogers and Kaufman 1996). Lastly, unlike some other insects, no expression is observed in the pre-oral clypeolabrum or in the hindgut primordium in the stages examined. However, the hindgut expression and secondary cephalic spots are not observed in *Tribolium* until dorsal closure (Schmidt-Ott et al. 1994), although secondary ocular spots appear in milkweed bugs before dorsal closure begins (Rogers and Kaufman 1996). As it is very difficult to recover good in situ hybridization expression

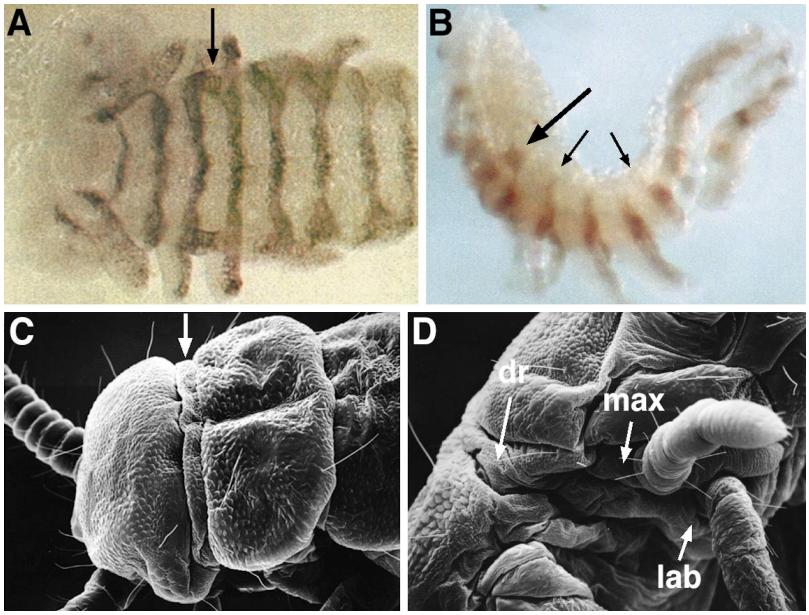

Fig. 5A–D The firebrat dorsal ridge. The firebrat *en*-r genes are expressed in the embryonic dorsal ridge. **A** The fully elongated embryo is shown dorsal side *up*, anterior to the *left*. The *arrow* points to the *en*-r-expressing dorsal ridge that lies between and dorsal to the maxillary and labial *en*-r stripes. **B** The embryo is shown on its side, anterior to the *left*. The *large arrow* shows *en*-r expression in the dorsal ridge. *Small arrows* point to dorsal portions of the thoracic and abdominal *en*-r stripes; these portions are absent in the head, other than the dorsal ridge. **C** Scanning electron micrographs of firebrat L1 hatchlings in dorsolateral view and **D** ventrolateral view. The dorsal ridge (*dr*; *arrow* in **C**) appears like a dorsal "neck" between the head and thorax. On its ventralmost extent, the dorsal ridge merges into the lateral folds of the maxillary (*max*) and labial (*lab*) segments (**D**), showing its composite structure

patterns from firebrat embryos after the onset of dorsal closure, it is unclear whether hindgut expression or cephalic secondary spots appear at later stages.

At the end of germ band elongation, there are six expression domains in the head, ten abdominal *en*-r stripes, expression in the cerci and possibly weak expression in the median caudal filament (Fig. 4C,F). Insofar as the expression of *en*-r reveals the number of segments within arthropod tagmata, firebrat *en*-r supports the hypothesis that there are six segments in the insect head (see Rogers and Kaufman 1996 for a thorough discussion). *en*-r expression in the firebrat abdomen is similar to the grasshopper (Patel et al. 1989a).

The similarity of expression of the firebrat genes mirrors the similarity of *Drosophila en* and *inv* embryonic expression, which also differ in relative timing (though of the whole pattern, not a specific part of it). Although function cannot be inferred from expression pattern, this parallel at least suggests that the firebrat *en*-r genes have overlapping functions like *Drosophila en* and *inv*. If so, then the firebrat genes may also have undergone concerted evolution due to the constraints of overlapping function, in which case the strength of support for a paraphyletic splitting of the firebrat genes may have been underestimated.

It is perhaps surprising that species with two *en*-r genes express those genes in nearly identical patterns, instead of deploying one paralogue in a new domain with new regulatory interactions. Even mouse *En-1* and *En-2* have similar expression patterns (Joyner and Martin 1987; Davis and Joyner 1988; Davis et al. 1988, 1991) and show functional redundancy, whereby En-2 can replace En-1 (Hanks et al. 1995). Perhaps the roles that *en*-r genes play in development find redundancy especially beneficial (Tautz 1992; Cooke et al. 1997). On the other hand, *Drosophila en* has a number of functional roles in development, including embryonic segmentation (Martinez Arias 1993) and neurogenesis (Goodman and Doe 1993; Bhat and Schedl 1997), a potential role in dorsal ridge and hindgut development, and imaginal disc development (Cohen 1993). Duboule and Wilkins (1998) have recently argued that as gene multifunctionality is accompanied by an increase of regulatory interactions, usable, non-lethal, variation decreases. There is undoubtedly some interplay of the constraints of multifunctionality and the freedom of redundancy occurring in the evolution of *en*-r genes, but knowing its exact nature requires a better understanding of the phylogeny of these genes and their functions in various representative species.

The dorsal ridge is an ancient structure in insects

As the posterior abdominal stripes appear, *en*-r expression is established at the dorsal edge of the labial segment in its anterior half (arrow in Fig. 4F). This region of dorsal *en*-r-expression is unique among the head segments. As germ band elongation is completed, this *en*-r expressing patch of cells lies dorsal to the maxillary and labial stripes, midway between them (Fig. 5A,B). At this stage, thoracic and abdominal *en*-r stripes extend to the

dorsal edge of the embryo (small arrows in Fig. 5B), but among the head segments, only this anterior labial stripe extends as far dorsally (large arrow in Fig. 5B).

This dorsal anterior labial *en*-r-expressing structure is present in other insects as well (Patel et al. 1989a; Diederich et al. 1991; Brown et al. 1994; Rogers and Kaufman 1996). Rogers and Kaufman (1996) proposed that this *en*-r-expressing entity is homologous to the dorsal ridge of the dipterans *Calliphora* and *Drosophila*, where it was first described (Schoeller 1964; Turner and Mahowald 1979). At the end of germ band retraction in *Drosophila*, it is composed of distinct paired lobes that lie dorsal to the gnathal segments at the boundary between the head and thorax. As dorsal closure begins, the dorsal ridge lobes fuse dorsally, forming a continuous ridge that looks like dorsal segment between the head and thorax (Turner and Mahowald 1979). At that stage, the *Drosophila* dorsal ridge appears nearly identical to the dorsal ridge of firebrat L1 hatchlings (Fig. 5C).

Interestingly, ectopic expression of *Ubx* in the *Drosophila* head revealed that the dorsal ridge is the anteriormost structure capable of producing dorsal cuticle (Rogers and Kaufman 1996). In the firebrat hatchling, the lateral portions of the maxillary and labial segments merge into the dorsal ridge (Fig. 5D), suggesting that these gnathal segments produce no dorsal structures other than their contribution to the dorsal ridge. Furthermore, Rogers and Kaufman (1996) proposed a model in which the dorsal ridge of insects is composed of two parts, one that expresses *en* and is formed from the dorsal portions of the labial and maxillary segments (Dr-I), and the other that expresses the *labial* gene and is formed from the dorsalmost cells of the pregnathal and mandibular segments (Dr-II). We have confirmed the existence of Dr-I in firebrats and it, therefore, appears to be an ancestral head structure in insects.

Acknowledgments We would like to thank Bryan Rogers and Arkhat Abzhanov for *Oncopeltus* embryonic cDNA, Douglas Rusch for a Perl script to remove third codon positions, Rudi Turner and Bryan Rogers for the SEMs, Patrick Keeling and Will Fischer for assistance with the phylogenetic analysis, and Dee Verostko for administrative assistance. This work was supported by the Howard Hughes Medical Institute (HHMI) and a National Science Foundation (NSF) Sloan Fellowship. M.D.P. is an HHMI Predoctoral Fellow, A.P. is a NSF/Sloan Postdoctoral Fellow and T.C.K. is an HHMI Investigator.

References

Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387: 489–493

Bhat KM, Schedl P (1997) Requirement for *engrailed* and *invected* genes reveals novel regulatory interactions between *engrailed/invected, patched, gooseberry* and *wingless* during *Drosophila* neurogenesis. Development 124: 1675–1688

Brown SJ, Patel NH, Denell RH (1994) Embryonic expression of the single *Tribolium engrailed* homolog. Dev Genet 15: 7–18

Cohen SM (1993) Imaginal disc development. In: Bate M, Martinez Arias A (eds) The development of *Drosophila melanogaster*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 517–608

Coleman KG, Poole SJ, Weir MP, Soeller WC, Kornberg TB (1987) The *invected* gene of *Drosophila*: sequence analysis and expression studies reveal a close kinship to the *engrailed* gene. Genes Dev 1: 19–28

Cooke J, Nowak MA, Boerlijst M, Maynard Smith J (1997) Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet 13: 360–364

Davis CA, Joyner AL (1988) Expression of the homeo box-containing genes *En-1* and *En-2* and the proto-oncogene *int-1* diverge during mouse development. Genes Dev 2: 1736–1744

Davis CA, Noble-Topham SE, Rossant J, Joyner AL (1988) Expression of the homeo box containing gene *En-2* delineates a specific region of the developing mouse brain. Genes Dev 2: 361–371

Davis CA, Holmyard DP, Millen KJ, Joyner AL (1991) Examining pattern formation in mouse, chicken and frog embryos with an En-specific antiserum. Development 111: 287–298

Diederich RJ, Pattatucci AM, Kaufman TC (1991) Developmental and evolutionary implications of *labial, Deformed* and *engrailed* expression in the *Drosophila* head. Development 113: 273–281

Dijk MA van, Murre C (1994) Extradenticle raises the DNA binding specificity of homeotic selector gene products. Cell 78: 617–624

DiNardo S, Kuner JM, Theis J, O'Farrell PH (1985) Development of embryonic pattern in *D. melanogaster* as revealed by accumulation of the nuclear *engrailed* protein. Cell 43: 59–69

Dolecki GJ, Humphreys T (1988) An *engrailed* class homeobox gene in sea urchins. Gene 64: 21–31

Dover GA (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet 2: 159–165

Duboule D, Wilkins AS (1998) The evolution of 'bricolage.' Trends Genet 14: 54–59

Ekker M, Wegner J, Akimenko MA, Westerfield M (1992) Coordinate embryonic expression of three zebrafish *engrailed* genes. Development 116: 1001–1010

Felsenstein J (1993) PHYLIP, Phylogeny Inference Package, version 3.5c University of Washington, Seattle

Field KG, Olsen GJ, Lane DJ, Giovannoni SJ, Ghiselen MT, Raff EC, Pace NR, Raff RA (1988) Molecular phylogeny of the animal kingdom. Science 239: 748–753

Fitch DH, Bugaj GB, Emmons SW (1995) 18S ribosomal RNA gene phylogeny for some Rhabditidea related to *Caenorhabditis*. Mol Biol Evol 12: 346–358

Fjose A, McGinnis W, Gehring WJ (1985) Isolation of a homeobox-containing gene from the *engrailed* region of *Drosophila* and the spatial distribution of its transcript. Nature 313: 284–289

Fleig R (1990) *Engrailed* expression and body segmentation in the honeybee *Apis mellifera*. Roux's Arch Dev Biol 198: 467–473

Gibert J-M, Mouchel-Viehl E, Deutsch JS (1997) *engrailed* duplication events during the evolution of barnacles. J Mol Evol 44: 585–594

Goodman CS, Doe CQ (1993) Embryonic development of the *Drosophila* central nervous system. In: Bate M, Martinez Arias A (eds) The development of *Drosophila melanogaster*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 517–608

Gustavson E, Goldsborough AS, Ali Z, Kornberg TB (1996) The *Drosophila engrailed* and *invected* genes: parameters in regulation, expression, and function. Genetics 142: 893–906

Han K, Manley JL (1993) Functional domains of the *Drosophila* engrailed protein. EMBO J 12: 2723–2733

Hanks M, Wurst W, Anson-Cartwright L, Auerbach AB, Joyner AL (1995) Rescue of the *En-1* mutant phenotype by replacement of *En-1* with *En-2*. Science 269: 679–682

Holland LZ, Kene M, Williams NA, Holland ND (1997) Sequence and embryonic expression of the amphioxus *engrailed* gene (*AmphiEn*): the metameric pattern of transcription resembles

that of its segment-polarity homolog in *Drosophila*. Development 124: 1723–1732
Hui C, Matsuno K, Ueno K, Suzuki Y (1992) Molecular characterization and silk gland expression of *Bombyx engrailed* and *invected* genes. Proc Natl Acad Sci USA 89: 167–171
Jaynes JB, O'Farrell PH (1991) Active repression of transcription by the engrailed homeodomain protein. EMBO J 10: 1427–1433
Joyner AL, Martin GR (1987) *En-1* and *En2*, two mouse genes with sequence homology to the *Drosophila engrailed* gene: expression during embryogenesis. Genes Dev 1: 29–38
Kissinger CR, Liu B, Martin-Blanco E, Kornberg T, Pabo C (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8Å resolution: a framework for understanding homeodomain-DNA interactions. Cell 63: 579–590
Kornberg T, Siden I, O'Farrell PH, Simon M (1985) The *engrailed* locus of *Drosophila*: in situ localization of transcripts reveals compartment-specific expression. Cell 40: 45–53
Kristensen NP (1991) Phylogeny of extant hexapods. In: Division of Entomology CSIRO (ed) The Insects of Australia. Cornell Univ Press, Ithaca, NY, pp 125–140
Logan C, Hanks MC, Noble-Topham S, Nallainathan D, Provart NJ, Joyner AL (1992) Cloning and sequence comparison of the mouse, human, and chicken *engrailed* genes reveal potential functional domains and regulatory regions. Dev Genet 13: 345–358
Manzanares M, Marco R, Garesse R (1993) Genomic organization and developmental pattern of expression of the *engrailed* gene from the brine shrimp *Artemia*. Development 118: 1209–1219
Martinez Arias A (1993) Development and patterning of the larval epidermis of *Drosophila*. In: Bate M, Martinez Arias A (eds) The development of *Drosophila melanogaster*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 517–608
Pankratz MJ, Jäckle H (1993) Blastoderm segmentation. In: Bate M, Martinez Arias A (eds) The development of *Drosophila melanogaster*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 467–516
Patel NH, Kornberg TB, Goodman CS (1989a) Expression of *engrailed* during segmentation in grasshopper and crayfish. Development 107: 201–212
Patel NH, Martin-Blanco E, Coleman KG, Poole SJ, Ellis MC, Kornberg TB, Goodman CS (1989b) Expression of *engrailed* proteins in arthropods, annelids, and chordates. Cell 58: 955–968
Peltenburg LTC, Murre C (1996) Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. EMBO J 15: 3385–3393
Peterson MD (1998) Analysis of the sequence and expression patterns of *engrailed* and homeotic genes in the primitively wingless insect, *Thermobia domestica*. Ph.D. Thesis, Indiana University, Bloomington
Popadić A, Panganiban G, Shear WS, Rusch D, Kaufman TC (1998) Molecular evidence for the gnathobasic derivation of arthropod mandibles and for the appendicular origin of the labrum and other structures. Dev Gen Evol 208: 142–150
Rogers BT, Kaufman TC (1996) Structure of the insect head as revealed by the EN protein pattern in developing embryos. Development 122: 3419–3432
Rogers BT, Peterson MD, Kaufman TC (1997) Evolution of the insect body plan as revealed by the *Sex combs reduced* expression pattern. Development 124: 149–157
Saitou N, Nei M (1987) The neighbor joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425
Schmidt-Ott U, Technau GM (1992) Expression of *en* and *wg* in the embryonic head and brain of *Drosophila* indicates a refolded band of seven segment remnants. Development 116: 111–125
Schmidt-Ott U, Sander K, Technau GM (1994) Expression of *engrailed* in embryos of a beetle and five dipteran species with special reference to the terminal regions. Roux's Arch Dev Biol 203: 298–303
Schoeller J (1964) Recherces descriptives et experimentales sur la cephalogenese de *Calliphora erythrocephala* (Meigen), au cours des developpements embryonnaire et postembryonnaire. Arch Zool Exp Gen 103: 1–216
Scholtz G (1995) Head segmentation in Crustacea– an immunocytochemical study. Zool Anz 98: 104–114
Scholtz G, Patel NH, Dohle W (1994) Serially homologous *engrailed* stripes are generated via different cell lineages in the germ band of amphipod crustaceans (Malacostraca, Peracarida). Int J Dev Biol 38: 471–478
Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL
Tabata T, Schwartz C, Gustavson E, Ali Z, Kornberg TB (1995) Creating a *Drosophila* wing de novo, the role of *engrailed*, and the compartment border hypothesis. Development 121: 3359–3369
Tamarelle M (1984) Concrete evidence for transient rudiments of 2nd antennae on the "intercalary" segment of insect embryos: comparative survey by scanning electron microscopy in *Anurida maritima* Guer (Collembola: Arthropleona) and *Hyphandria cunea* Drury (Lepidoptera: Arctiidae). Int J Insect Morphol Embryol 13: 331–336
Tautz D (1992) Redundancies, development and the flow of information. Bioessays 14: 263–266
Turner FR, Mahowald MP (1979) Scanning electron microscopy of *Drosophila melanogaster* embryogenesis. III. Formation of the head and caudal segments. Dev Biol 68: 96–109
Walldorf U, Fleig R, Gehring WJ (1989) Comparison of homeobox-containing genes of the honeybee and *Drosophila*. Proc Natl Acad Sci USA 86: 9971–9975
Webster PJ, Mansour TE (1992) Conserved classes of homeodomains in *Schistosoma mansoni*, an early bilateral metazoan. Mech Dev 38: 25–32
Wedeen CJ, Weisblat DA (1991) Segmental expression of an *engrailed*-class gene during early development and neurogenesis in an annelid. Development 113: 805–814
Wedeen CJ, Price DJ, Weisblat DA (1991) Cloning and sequencing of a leech homolog to the *Drosophila engrailed* gene. FEBS Lett 279: 300–302
Wedeen CJ, Kostriken RG, Leach D, Whitington P (1997) Segmentally iterated expression of an *engrailed*-class gene in the embryo of an australian onychophoran. Dev Genes Evol 270: 282–286
Woodland JT (1957) A contribution to our knowledge of lepismatid development. J Morphol 101: 523–577
Wray CG, Jacobs DK, Kostriken R, Vogler AP, Baker R, DeSalle R (1995) Homologues of the *engrailed* gene from five molluscan classes. FEBS Lett 365: 71–74