Available online at www.sciencedirect.com
SCIENCE DIRECT
Organisms, Diversity & Evolution 6 (2006) 127–139
ORGANISMS DIVERSITY & EVOLUTION
www.elsevier.de/ode

The phylogenetic characteristics of three different 28S rRNA gene regions in *Encarsia* (Insecta, Hymenoptera, Aphelinidae)

Stefan Schmidt[a,*], Felice Driver[b], Paul De Barro[c]

[a]*Zoologische Staatssammlung München, Münchhausenstr. 21, 81247 München, Germany*
[b]*CSIRO Entomology, GPO Box 1700, Canberra 2601, Australia*
[c]*CSIRO Entomology, 120 Meiers Road, Indooroopilly, Brisbane 4068, Australia*

Received 31 May 2004; accepted 8 July 2005

Abstract

Two expansion segments of the large ribosomal subunit (28S-D2 and 28S-D3) and the ribosomal first internal transcribed spacer (ITS-1) were sequenced to investigate the potential of each region for defining species limits and for inferring phylogenetic relationships within the aphelinid genus *Encarsia* (Hymenoptera, Chalcidoidea). Alignment of the ITS-1 region was complicated by the presence of high levels of length polymorphism. Secondary-structure models of the ITS-1 molecule are proposed which led to an improved alignment and the inclusion of some highly polymorphic sequences. The additional information obtained from the secondary-structure models can provide a justification for the exclusion of hypervariable regions from the phylogenetic analysis. The D3 region is the most conserved, ITS-1 the most variable of the three gene regions investigated. Phylogenetic analyses suggest the D2 region to be most suitable not only for inferring relationships, but also for taxonomic and diagnostic purposes at species level. The ITS-1 region is best suited for inferring relationships between closely related species or among populations of the same species. The combined analysis using all three regions led to a better tree and a reduced number of equally most parsimonious trees.

Keywords: Parasitoids; Molecular phylogeny; 28S ribosomal RNA; Internal transcribed spacer 1; Alignment; Secondary structure
See also **Electronic Supplement** at: http://www.senckenberg.de/odes/06-07.htm

Introduction

Encarsia is a large genus of minute parasitic wasps of the chalcidoid family Aphelinidae, with currently about 275 described species (Heraty and Woolley 2002; Noyes 2002). The majority of species attack whiteflies (Hemiptera: Aleyrodidae) and armoured scale insects (Hemiptera: Diaspididae). Several *Encarsia* species are important biological control agents (Huang and Polaszek 1998), a few are parasitoids of aphids (Hemiptera: Aphididae) (Evans et al. 1995) or of Lepidoptera eggs (Polaszek 1991). Since *Encarsia* is a taxonomically difficult genus, attempts have been made to divide it into smaller, manageable subunits. Currently, about 29 species-groups are recognised (Manzari et al. 2002). However, only a few of these species-groups are defined unambiguously on the basis of morphological characters alone (Manzari et al. 2002). Furthermore, the number of species-groups and the assignment of species to a particular group vary depending on the respective author's opinion (e.g. Viggiani and Mazzone 1979; Hayat 1998; Heraty and Polaszek 2000).

*Corresponding author.
E-mail address: schmidt@zsm.mwn.de (S. Schmidt).

1439-6092/$ - see front matter
doi:10.1016/j.ode.2005.07.002

Phylogenetic relationships within the genus still are largely unresolved; only recently have attempts been made to use molecular data to underpin the taxonomy based on morphological characters, and to resolve phylogenetic relationships (Babcock et al. 2001; Schmidt et al. 2001; Manzari et al. 2002). All molecular studies conducted so far have used the D2 expansion region of the 28S ribosomal RNA; there has been little information about the suitability of other gene regions to inferring phylogenetic relationships or to defining species limits.

So far, molecular data are available for only a small percentage of all known *Encarsia* species, but rather than providing another, more comprehensive molecular phylogeny of *Encarsia*, the aim of our study is to investigate two 28S gene regions that have not been used for phylogenetic purposes, and to compare them with the 28S-D2 region. These gene regions are the third expansion region of the large ribosomal subunit (28S-D3), and the first internal transcribed spacer region 1 (ITS-1).

Material and methods

Surveys

The study was based largely on material collected over a period of 3 years from the Pacific Islands (1996–1997) and Australia (1996–1998). All parasitoids were reared from whiteflies (Sternorrhyncha: Aleyrodidae) collected in the field, most of them from either *Bemisia tabaci* (Gennadius) or *Trialeurodes vaporariorum* (Westwood), some from other species of *Bemisia* Quaintance & Baker, from *Aleurocanthus* Quaintance & Baker spp., or from *Aleurodicus dispersus* Russell. Each sample was given a unique code number, and the host plant, host whitefly species, date, location and collector were noted. Nymphs of parasitised hosts were kept in emergence chambers, emerging parasitoids transferred to gelatine capsules or 94% ethanol, where they remained at room temperature until further examination or DNA analysis. Whiteflies were identified to species level using the fourth-instar pupal case from which the parasitoid had emerged (Martin 1987, 1999). *B. tabaci* biotypes were identified using adults collected along with the parasitised nymphs according to the method described in De Barro and Driver (1997).

Sample data set

A subset of 22 populations belonging to 17 *Encarsia* species was selected from the data set used in Babcock et al. (2001) which included representatives of all major species-groups. DNA extractions of the same specimens used in the earlier study were used to obtain sequences of the 28S-D3 and IST-1 regions. In addition, *Coccophagoides fuscipennis* (Girault) was included as the outgroup species for tree routing purposes. Preliminary results of our ongoing efforts to resolve phylogenetic relationships within the aphelinid subfamily Coccophaginae indicate that *Encarsiella* Hayat is polyphyletic and that several species presently placed in that genus will have to be transferred to *Encarsia*. The second species included for outgroup comparison in the earlier analysis; *Encarsiella noyesi* Hayat, came out within the genus *Encarsia* instead (Babcock et al. 2001), therefore was not used for that purpose again in the current study. Generic limits of *Encarsia* and closely related genera (including *Encarsiella*) are in need of redefinition, but clarification of higher-level relationships of these genera is beyond the scope of the present study.

DNA extraction, amplification, and sequencing

DNA was extracted from single, whole specimens in the manner described by De Barro and Driver (1997). The crude lysate was boiled for 5 min; samples were stored at −20 °C until required. Sequences of the domains two and three (D2 and D3) of the 28S rRNA gene and of the ribosomal ITS-1 gene region were obtained from 22 populations of 17 *Encarsia* species (see Appendix) as well as from the outgroup species, *C. fuscipennis*, each with 2–3 replicates (specimens). The polymerase chain reaction (PCR) was used to amplify the gene regions. Primer sequences are given in Table 1. All reaction volumes were 50 µl, containing 20 pM of each primer, 200 µM each of dGTP, dATP, dCTP and dTTP, 1.5–2.5 mM $MgCl_2$, 2 µl DNA lysate, 1 × supplied buffer and 2.5 U Taq polymerase (Bresatec, Australia). PCR amplifications were done in a Hybaid thermocycler, included a pre-cycle denaturation step for 5 min at 94 °C, followed by the addition of Taq polymerase, and a final post-cycle extension step at 72 °C for 5 min (Table 1).

The D2 and D3 amplicons were purified and prepared for sequencing by electrophoresis in 0.8% TAE agarose gels containing 10 µg ml^{-1} ethidium bromide (Sambrook et al. 1989). Fragments were excised and transferred to a microfuge tube. The agarose slices were mashed in 30 µL sterile distilled water using a toothpick, then incubated at 50 °C for 1 h. Samples were left at room temperature overnight to allow the DNA to elute from the gel, and subsequently stored at −20 °C until required. ITS-1 amplicons were selectively precipitated and ligated into the pPCR-Script Amp SK(+) cloning vector from Stratagene (La Jolla, California, USA), according to the manufacturer's protocol.

Five microlitres of the eluted PCR amplicons, or approximately 250 ng cloned plasmid DNA, and the

Table 1. Primer sequences and cycling conditions

	Primer sequence	Cycling conditions			
		Denaturation	Annealing	Extension	Cycles
28S-D2 (Dr. B. Campbell, USDA-ARS, W. Regional Res. Ctr.)					
D2F	5′ CGTGTTGCTTGATAGTGCAGC 3′	94 °C	55 °C	72 °C	35
D2R	5′ TTGGTCCGTGTTTCAAGACGG 3′	(1 min)	(1 min)	(1.5 min)	
28S-D3 (Nunn et al. 1996)					
D3A	5′ GACCCGTCTTGAAACACACGGA 3′	94 °C	55 °C	72 °C	30
D3B	5′ TCGGAAGGAACCAGCTACTA 3′	(1 min)	(1 min)	(1.5 min)	
ITS1 (Brust et al. 1998)					
TW81	5′ GTTTCCGTAGGTGAACCTGC 3′	94 °C	55 °C	72 °C	30
Aed5.8R	5′ GAGAACAGCAGGAACACAGAAC 3′	(1 min)	(1 min)	(1.5 min)	

appropriate PCR primers were used for sequencing according to the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit Manual (PE-Applied BioSystems). Both strands of each fragment were sequenced; reactions were loaded onto an ABI Model 373A Sequencer.

Sequences have been deposited in the GenBank database under Accession nos. AY615734–AY615785, with the exception of those 28S-D2 sequences that had been submitted as part of an earlier study (Babcock et al. 2001). The alignment is available as an Organisms Diversity and Evolution Electronic Supplement (http://www.senckenberg.de/odes/06-07.htm).

Analysis of DNA sequences

The D2, D3, and ITS-1 gene sequences were aligned using Clustal W (Thompson et al. 1994). Secondary-structure models were calculated and drawn using mfold, version 3.1 (Mathews et al. 1999; Zuker 2003; http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi). The constraint option was used to force bases to be double-stranded in cases where the predicted foldings and manual inspection of the sequence suggested possible base pairings that were not realised by the program.

Phylogenetic analysis by maximum parsimony was performed using PAUP* version 4.0b10 (Swofford 2002). Each heuristic search included 100 replicates, each with a random stepwise-addition sequence. Optimisation was accelerated transformation (ACCTRAN), branch swapping was by tree-bisection and reconnection (TBR) with a single tree held at each step. All characters were assigned equal weight and treated as unordered. Initially, alignment gaps were treated as missing data. Since the way how gaps are treated can have a considerable effect on the phylogenetic reconstruction (Giribet and Wheeler 1999), all analyses were repeated with gaps coded as an additional character state. Bootstrap support values were calculated with 100 heuristic search replicates and 10 random-addition sequences each.

Results

D3 expansion segment of the 28S rRNA

The alignment of the D3 sequences was unambiguous; only a single insertion/deletion had to be postulated. Of the 314 positions in the alignment, 50 positions vary, all of which are located between positions 19 and 173. Uncorrected pairwise divergences (*p*-distances) between ingroup taxa ranged from no differences to 8.0%, with an average of 4.0%. The base composition is shown in Table 2. Individuals of the same species from different geographical regions usually have identical sequences, e.g. those of *E. lutea* (Masi) from Australia and from the Pacific Islands, as well as of *E. pergandiella* Howard and *E. formosa* Gahan from Australia and from the United States, respectively (see Appendix). However, sequences of *E. sophia* (Girault & Dodd) from Australia and from the Pacific Islands differ in two positions, and individuals of the Bundaberg population of *E. bimaculata* Heraty & Polaszek differ from all other Australian populations of that species by a single position, indicating the presence of cryptic species. Hence, the Bundaberg population of *E. bimaculata* might represent a distinct, morphologically very similar species (Schmidt et al. 2001). In some cases the level of variation was very low even between morphologically distinct species: individuals of *E. cibcensis* Lopez-Avila and *E. pergandiella* differ in only two positions, as do *E. smithi* (Silvestri) and *E. pergandiella*. The Bundaberg population of *E. bimaculata* differs from *E. protransvena* Viggiani and *E. oakeyensis* Schmidt & Naumann in

Table 2. Mean nucleotide frequencies

	A	G	C	T
D2	21.1	29.7	24.2	25.0
D3	25.8	31.6	23.5	19.1
ITS-1	29.6	18.5	18.3	33.6

Table 3. Tree statistics of the separate and combined analyses of the three gene regions

Gene region	Number of most parsimonious trees	Tree length	CI	RI
28S-D2	3	436	0.596	0.760
28S-D3	159	81	0.716	0.828
ITS-1	1	2158	0.648	0.608
Combined analysis	1	2698	0.636	0.650

only a single position; sequences of *E. accenta* Schmidt & Naumann and *E. azimi*, both members of the *E. inaron* species-group, are identical.

The D3 region is the most conserved and, with an alignment length of 314 positions, the shortest of all three gene regions sequenced, comprising 36 parsimony-informative sites (11.5%), whereas 14 of the 50 variable characters are parsimony-uninformative. A heuristic search yielded 159 equally most parsimonious trees of 81 steps, with a consistency index of 0.716 and a retention index of 0.828 (Table 3). The 50% majority rule consensus and the strict consensus trees are identical, except that in the former *E. inaron* is the sister group of clade containing the species *E. accenta*, *E. adusta*, *E. ustulata*, and *E. dispersa*, which is part of the basal polytomy. The strict consensus tree is given in Fig. 1, with bootstrap support values larger than 50% shown above branches. The D3 trees have a higher consistency index than trees of the other two gene regions, indicating a low level of homoplasy (Table 3). The number of changes at variable sites of the D3 region is very low, ranging from one to four changes (Fig. 2). Most branches of the few resolved relationships on the tree are at least moderately supported by bootstrap values, although in some cases support is weak (Fig. 1). A χ^2 test of homogeneity of base frequencies across taxa did not indicate significant differences ($p = 1.0$).

D2 expansion segment of the 28S rRNA

Alignment of the D2 region resulted in 621 characters of which 201 positions are variable. The level of variation between ingroup taxa ranged from no differences between allopatric populations of the same species to 87 positions (14.9%) between distantly related species, with an average of 10.4%. The base composition is given in Table 2. Some closely related species show very few differences in this gene region, although they are morphologically distinct. For instance, *E. protransvena* and *E.* nr *citri* differ by 1.7%, and differences among species of the *E. inaron* species-group (*E. inaron* (Walker), *E. accenta*, *E. adusta* Schmidt & Naumann, *E. azimi*) range from 0.9% to 2.7%.

The 23 unique D2 sequences contain 158 parsimony-informative sites (25.4%); a heuristic search resulted in three equally most parsimonious trees of length 436, with a consistency index of 0.596 and a retention index of 0.760 (Table 3). The 50% majority rule consensus and the strict consensus trees are identical, except that in the former *E. nigricephala* Dozier is the sister group to the species of the *E. luteola* group clade (*E. dispersa* Polaszek, *E. formosa*, *E. guadeloupae* Viggiani). The strict consensus tree is given in Fig. 1, with bootstrap support values shown above branches. Variable sites of the D2 region show a low number of multiple changes, ranging from one to six in the corresponding tree (Fig. 2). A χ^2 test of homogeneity of base frequencies across taxa did not indicate significant differences ($p = 1.0$).

Ribosomal ITS-1

Alignment

Alignment of the ITS-1 was complicated by a high level of sequence-length variation ranging from 456 base pairs (bp) for the shortest (*E. protransvena*) to 645 bp for the longest sequence (*E. smithi*). However, an initial alignment indicated the presence of several highly conserved sections, which alternate with hypervariable parts in closely related species-groups. Furthermore, manual inspection of the conserved regions revealed that all were able to form helices of variable length, but with the same secondary-structure configuration. Minimum free energy foldings were obtained from a number of species, including taxa with very long and others with very short sequences. This was done for two reasons: (1) to identify structural similarities between sequences of different species, which would help to improve the alignment; and (2) to justify the exclusion of hypervariable regions from the alignment where consistent secondary-structural features could not be predicted that were otherwise identified and characterised by lying between conserved sequence motifs in the alignment. Based on these folding patterns as well as on manual inspection and adjustment of the sequence alignment, conserved sequence elements were elucidated in all species sequences; the arrangement of the consensus alignment is shown in Fig. 3. In addition, several

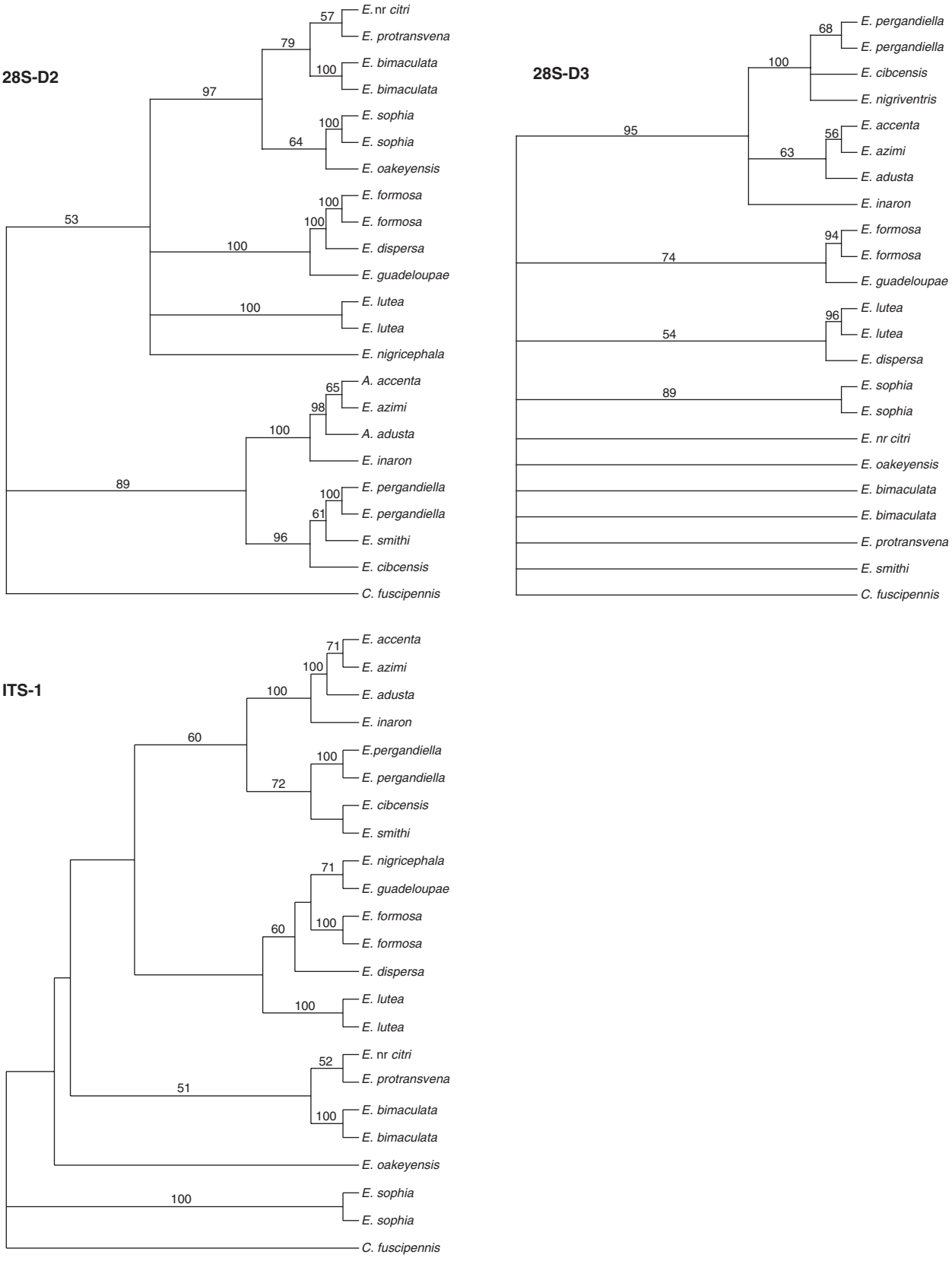

Fig. 1. Separate analyses of the three gene regions. 28S-D2: strict consensus tree of three equally most parsimonious trees (MPTs); 28S-D3: strict consensus tree of 159 MPTs; ITS-1: single MPT. Numbers above branches indicate bootstrap support values (> 50%) for 100 iterations.

putative structural motifs could be predicted that were common to all species included in the analysis; these are denoted in Figs. 5–7. Some highly conserved sequence elements in the ITS-1 region have the potential to form long-range based pairing stem regions, for instance I′ + I″, II′ + II″, and III′ + III″ in Figs. 3 and 5–7. These motifs have the potential to stabilise the long axis of the putative 'cruciform' structures.

Since the additional information obtained from the secondary-structure models aided in aligning the complete ITS-1 section, we did not regard exclusion of hypervariable regions as justifiable. However, to investigate their effect on the phylogenetic results, the data set was analysed also with hypervariable regions excluded.

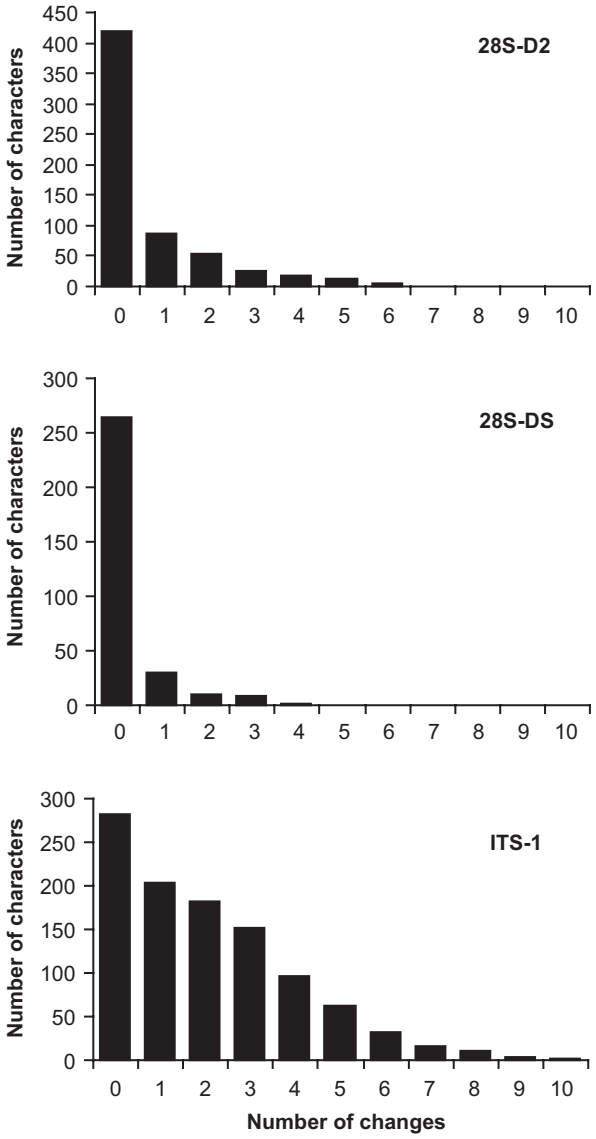

Fig. 2. Frequency distribution of the number of nucleotide changes per character for the three gene regions 28S-D2, 28S-D3, and ITS-1, calculated on the most parsimonious trees.

Phylogenetic analysis

The final alignment of 22 ITS-1 sequences of all ingroup taxa contained 1044 characters, with 545 positions parsimony-informative (52.2%). A heuristic search produced a single most parsimonious trees of 2158 steps, with a consistency index of 0.648 and a retention index of 0.608 (Table 3). The 50% majority rule consensus and the strict consensus trees differ in that the latter basal relationships are better resolved. In the 50% majority rule consensus tree several species are part of a basal polytomy, e.g. *E. lutea* and *E. oakeyensis*.

Furthermore, the *E. luteola* species-group shows a higher level of resolution in the strict consensus tree, which is given in Fig. 1 with bootstrap support values shown above branches. Variable sites of the ITS-1 region exhibit many multiple changes on the ITS-1 tree, ranging from one to ten changes, indicating a high level of saturation of the more variable sites (Fig. 2). In contrast to the other two gene regions, a χ^2 test of homogeneity of base frequencies across taxa resulted in significant differences ($p < 0.01$). This bias is caused by the increased A-T content of the ITS-1 region, although it is not as extreme as in hymenopteran mtDNA (Dowton and Austin 1997; Whitfield and Cameron 1998).

In the cladogram based on ITS-1 sequences, grouping of species into morphologically defined species-groups is similar to the results from the D2 region, and supported by mostly moderate or high bootstrap values (Fig. 1). The exclusion of hypervariable regions did not result in different groupings, except for a few differences in the arrangement of basal relationships. However, these were poorly supported in both analyses, i.e. including and excluding hypervariable regions (cf. Fig. 1). Based on these results, inclusion of the whole ITS-1 sequence in

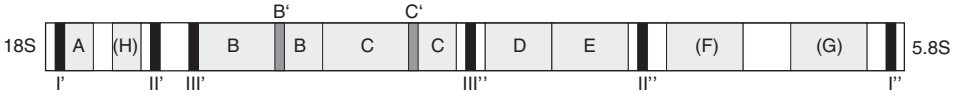

Fig. 3. Consensus map of the ITS-1 sequence structural elements (not to scale). Explanation of gene region indices: A, B, C, D, E: self-complementary double-strand helices; B′, C′: conserved internal sequences, elements or structural motifs in helices B and C, respectively; I′ + I″, II′ + II″, III′ + III″: highly conserved long-range single-stranded pairing elements; (F), (G), (H): partially conserved self-complementary double-strand helices; white regions: single-stranded loops or long-range pairing elements that make up the double-stranded cruciform structure.

the combined analysis of all three gene regions seemed appropriate.

Combined analysis of D2, D3 and ITS-1

Sequences of all 22 ingroup taxa were tested for incongruence among data sets. The incongruence length difference (ILD) test with 100 replicates did not indicate significant incongruence among data sets ($p = 0.17$). It has been shown that the ILD test may overestimate the level of incongruence, unless uninformative characters are excluded (Lee 2001). The three data sets differ considerably in their percentage of informative sites, ranging from 11.5% for the D3 to 52.2% for the ITS-1 partition. Therefore parsimony-uninformative (constant and autapomorphic) characters were excluded prior to the analysis. The aligned data set of all three gene regions comprised 1979 characters, with 1013 variable positions and 739 parsimony-informative sites. The parsimony analysis yielded a single most parsimonious tree of 2698 steps, with a consistency index of 0.636 and a retention index of 0.650 (Table 3). The resulting tree (Fig. 4) exhibits a higher level of resolution than any of the trees resulting from the separate analyses, and all branches except one received bootstrap support values >50%. The composition of species-groups is similar between the separate and combined analyses. An exception is *E. nigricephala* which is a species of uncertain placement. It is the only representative of

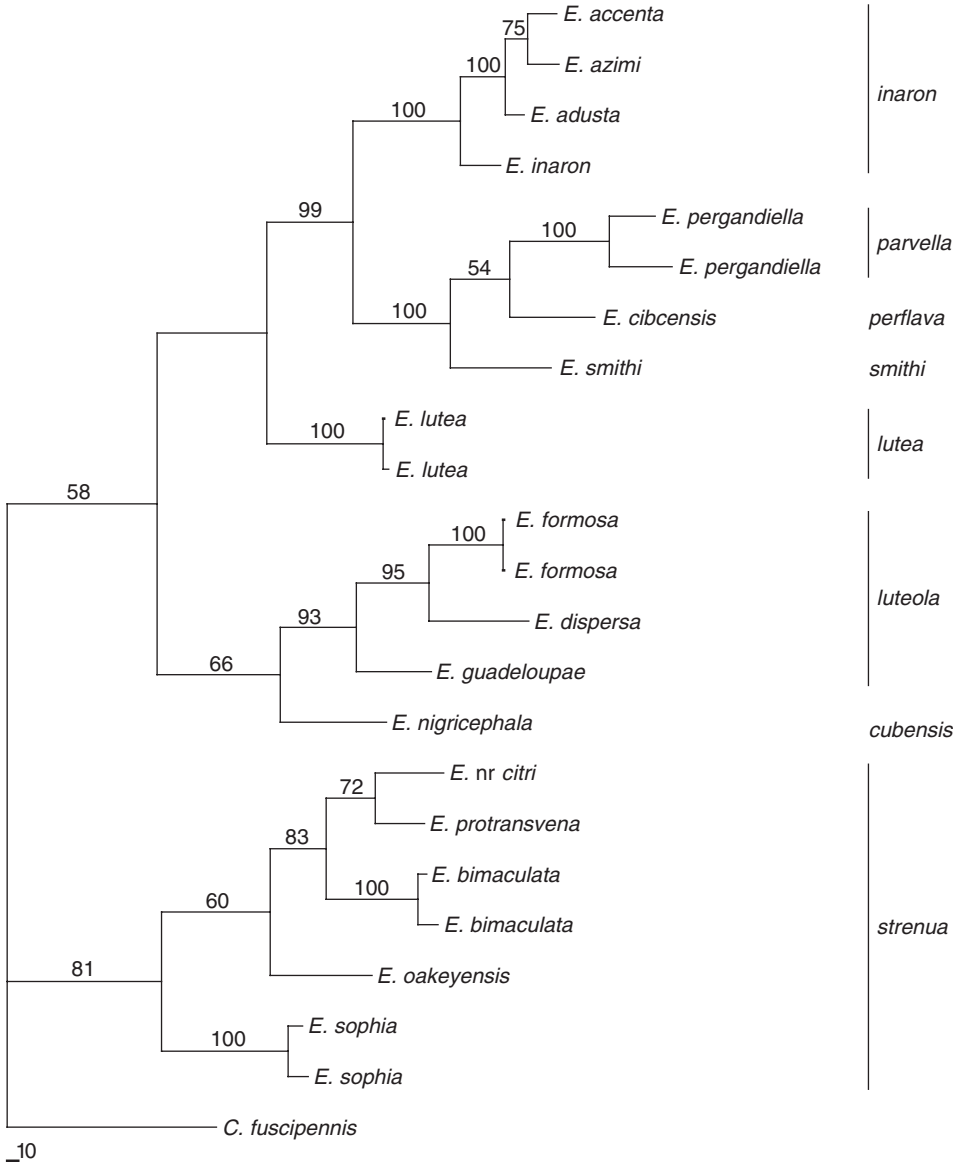

Fig. 4. Single most parsimonious tree of the combined analysis of the three gene regions 28S-D2, 28S-D3, and ITS-1. Numbers above branches indicate bootstrap support values (>50%) for 100 iterations.

the *E. cubensis* species-group, and the only one with an 11 bp insertion in the D2 sequence between positions 565 and 576 of the alignment.

Besides the better resolution, bootstrap support was higher or at least comparable in the combined analysis, e.g. 81% for the *E. strenua* group, 100% for the *E. inaron* group, and 93% for the *E. luteola* group. There are a few branches with lower support in the combined analysis compared to the corresponding branch in the separate analyses, but in those cases the difference is low or the bootstrap value is higher in only one of the three separate analyses. The tree resulting from the combined analysis most closely resembles the D2 tree, although the former suggests different hypotheses about phylogenetic relationships at basal parts of the tree. For instance, the clade containing the *E. strenua*, *E. luteola*, *E. lutea* and *E. cubensis* groups does not exist in the combined analysis where the *E. strenua* group represents one of two major clades, the second clade containing all remaining species.

The analyses of each of the three gene regions and the combined data set with gaps treated as additional character states resulted in identical topologies with similar or identical bootstrap values that did not differ by more than 10% from the values obtained in the analyses with gaps treated as missing characters.

Discussion

The comparative analysis of the second and third expansion regions of the large ribosomal subunit (28S-D2 and D3) and the ITS-1 revealed considerable differences in their suitability to defining species limits on a molecular level and to inferring phylogenetic relationships. The 28S-D3, with an alignment length of 314 positions, is the shortest and, with an average variation of 4.0%, the most conserved of all three gene regions. *Encarsia* species with identical sequences limit the suitability of the D3 region for diagnostic purposes at species level. The low level of variation resulted, as expected, in a poorly resolved tree, although the few resolved relationships are in agreement with traditionally accepted groupings (Fig. 1). An exception is *E. dispersa* which is closer to *E. lutea* than to members of the *E. luteola* species-group (*E. formosa* and *E. guadeloupae*) in which it is usually placed on the basis of morphological and molecular evidence (cf. Appendix; Polaszek et al. 2004).

The alignment of the D2 expansion region, with 621 characters, is about twice as long as the D3 alignment, and at 10.4% variation is considerably more variable than the D3 region. It does not contain morphologically separate species with identical sequences, although interspecific variation between some closely related species is very low (see Results section). The phylogenetic analysis confirms all species-groups traditionally defined on a morphological basis, and most of them are supported by very high bootstrap support levels, e.g. 98% for the *E. strenua* group, and 100% for both the *E. luteola* and the *E. inaron* groups (Fig. 1). Basal relationships vary between the three equally parsimonious reconstructions of the D2 data, leading to less resolution near the base of the strict consensus tree. In particular, the variable positions of *E. lutea* and *E. nigricephala* in each of the three most parsimonious D2 trees led to the polytomy at the base of the clade containing the *E. strenua* and the *E. luteola* species-groups. Compared to the other two gene regions, the D2 expansion region appears to be the one most suitable for species diagnosis on a molecular level.

Considerable length variation between sequences of the ITS-1 region caused complications when applying the alignment algorithms of Clustal W. However, the presence of conserved regions was taken as an indication that functional constraints exist within the region, possibly related to a secondary-structure configuration that is required for further processing of the transcript (Venkateswarlu and Nazar 1991). The secondary-structure models obtained by folding the molecule using energy minimisation assisted in identifying and visualising how the length-variable sections of the molecule are interspersed between conserved sequence motifs (cf. Figs. 5–7), and thereby assisted in improving the alignment. Figs. 5–7 show secondary-structure models for two *E. protransvena* (Fig. 5) and *E. sophia* (Fig. 6) as well as for the outgroup species, *C. fuscipennis* (Fig. 7), and illustrate the considerable interspecific differences in sequence length. For example, helices of domains B and E are distinctly longer in *E. sophia* (Fig. 6) than in the corresponding motifs in *E. protransvena* (Fig. 5). The length of a particular domain mostly correlates with the overall length of a gene region. The B domain of the ITS-1 gene region, for instance, is generally longer in species with a long ITS-1 sequence than in species with a short one, e.g. *E. protransvena* (Fig. 5) which has the shortest ITS-1 region (456 bp). However, this is not always the case: in *E. sophia* from Australia, despite having a comparatively long ITS-1 region (579 bp), the D domain is distinctly shorter than the corresponding domain in the relatively short sequence of *E. protransvena* (Fig. 5). Similarly, the lengths of the hypervariable sections of the ITS-1 molecule (H and G in Figs. 5–7) apparently do not correspond to the overall length of the molecule. Both H and G are the most variable sections of the ITS-1 sequence. These regions may be short in some species, and do not always form helices (cf. H in Fig. 5 compared to H in Figs. 6 and 7), but rather contribute to stabilising the long axis of the secondary structure. Our observations indicate that the functional integrity of the ITS-1 region does not depend

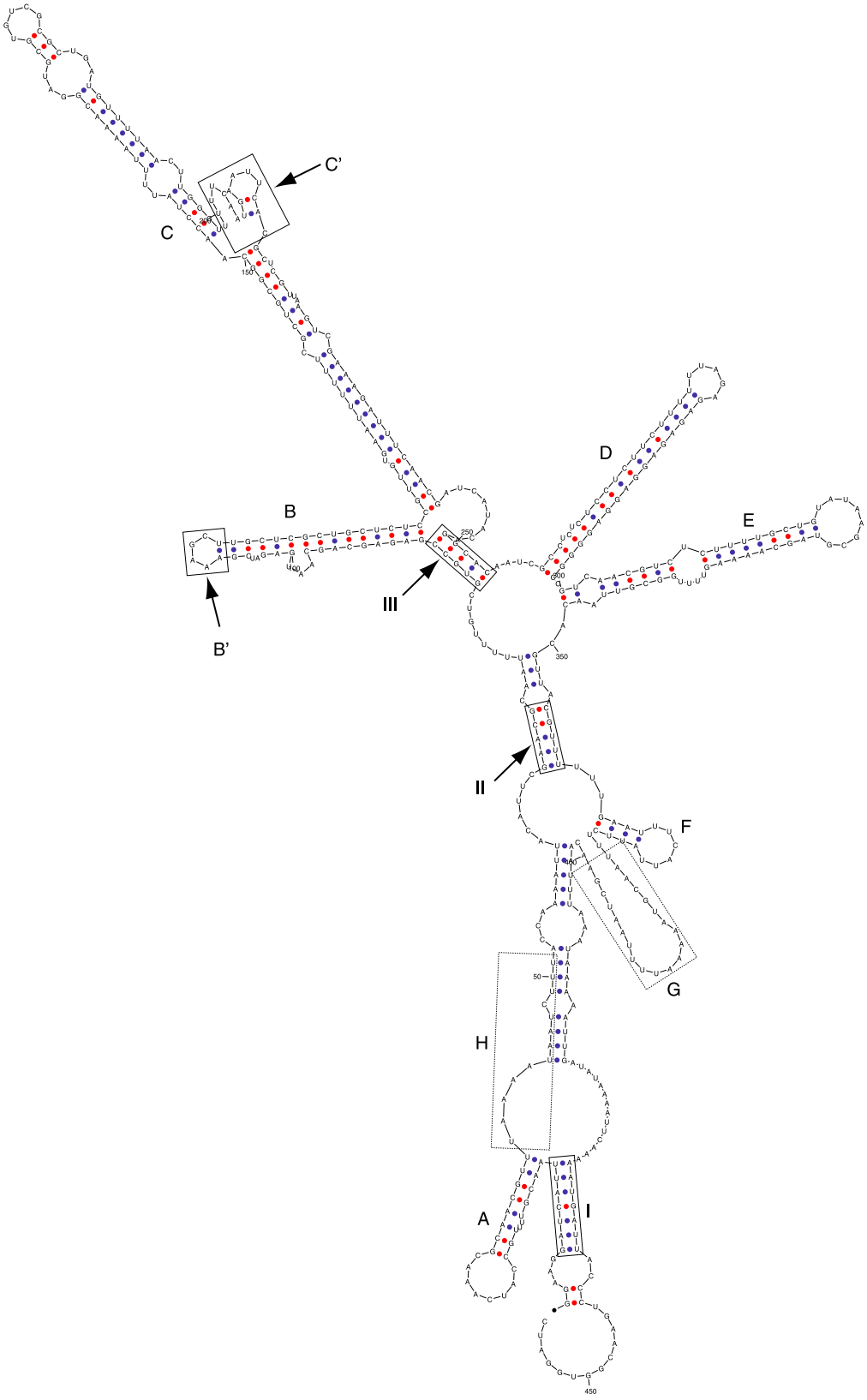

Fig. 5. ITS-1 secondary-structure model for *Encarsia protransvena*. Solid boxes indicate highly conserved regions (I, II, and III) or conserved internal sequences (B′ and C′), dotted lines show approximate positions of hypervariable regions. For explanation of indices see Fig. 3 and text.

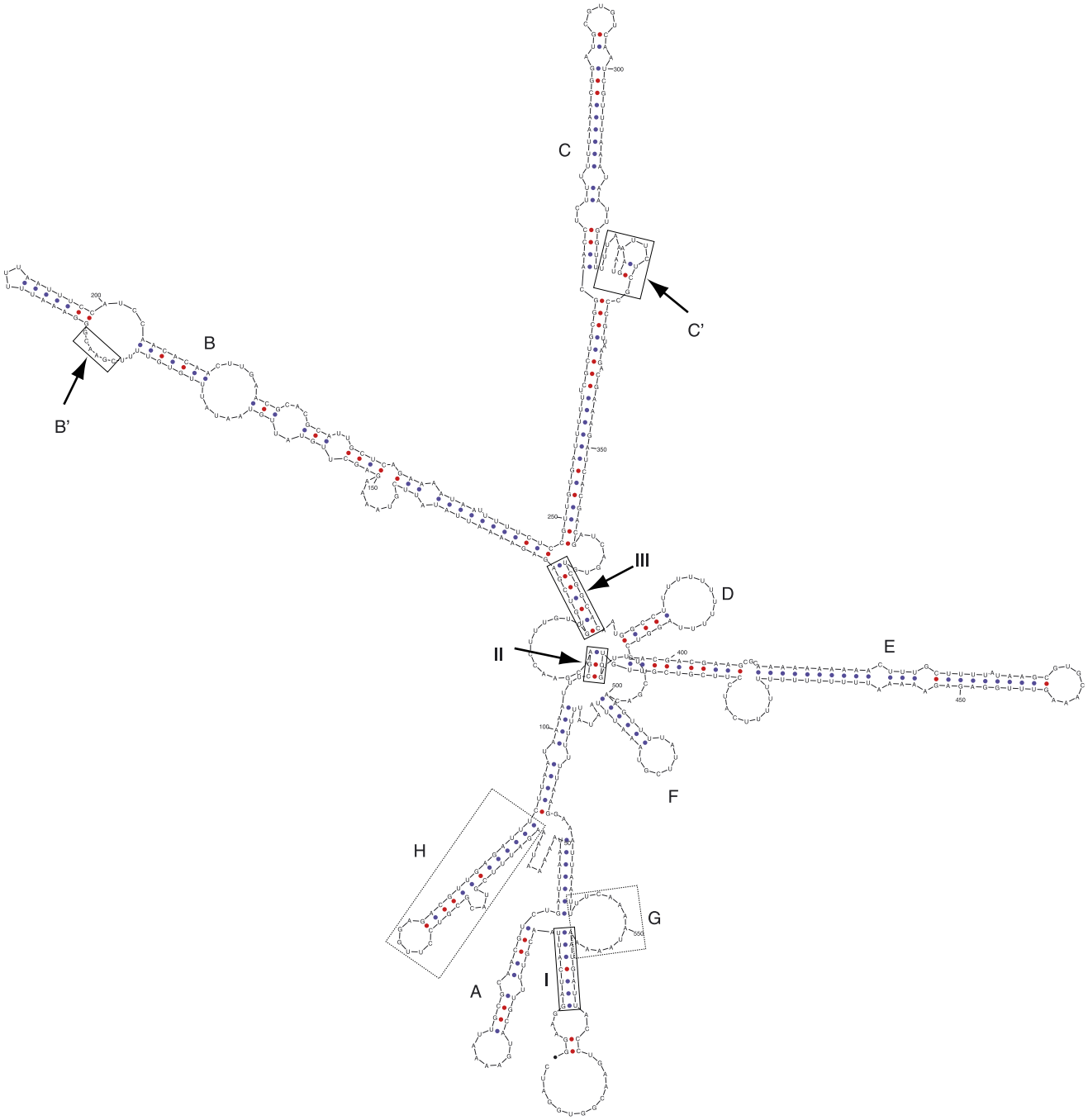

Fig. 6. ITS-1 secondary-structure model for *Encarsia sophia*. Solid boxes indicate highly conserved regions (I, II, and III) or conserved internal sequences (B′ and C′), dotted lines show approximate positions of hypervariable regions. For explanation of indices see Fig. 3 and text.

strictly on structural constraints related to the length proportions of domains, and explains to some extent the strong interspecific variation in ITS-1 region length.

The presence of numerous alignment gaps, as observed in the ITS-1 region, inevitably leads to the uncertainty of positional homologies. However, the cladogram resulting from the analysis of the ITS-1 region mostly recovered the same clades (species-groups) as the D2 region which has been shown to support hypotheses about phylogenetic relationships within *Encarsia* that are in agreement with hypotheses based on morphological characters. Some groupings receive more support than in previous phylogenetic analyses (e.g. Babcock et al. 2001), e.g. the placement of

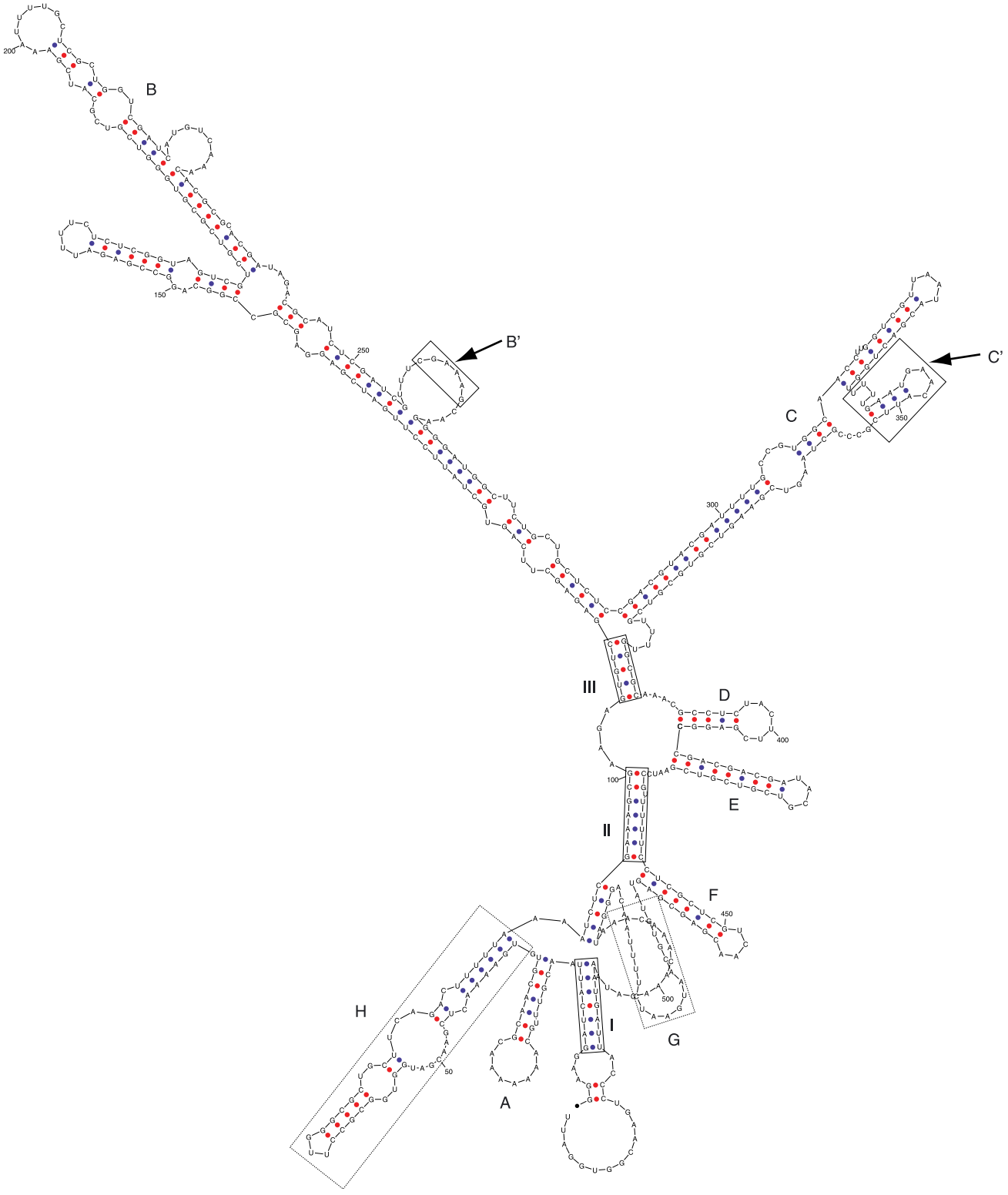

Fig. 7. ITS-1 secondary-structure model for *Coccophagoides fuscipennis*. Solid boxes indicate highly conserved regions (I, II, and III) or conserved internal sequences (B′ and C′), dotted lines show approximate positions of hypervariable regions. For explanation of indices see Fig. 3 and text.

E. nigricephala within the *E. luteola* group (Fig. 1). *Encarsia. nigricephala* belongs to the *cubensis* species-group but, apart from members of the *luteola* group, is the only species with 4-segmented midtarsi, a putative synapomorphy for this clade (Babcock et al. 2001). In the ITS-1 analysis the sister species clade

guadeloupae + *nigricephala* receives a bootstrap support of 71% (Fig. 1), whereas in the combined analysis the clade containing the *luteola* group + *nigricephala* is supported at 66% (Fig. 4).

The trees resulting from separate analysis of the D2, D3, and ITS-1 sequences, respectively, reflect the characteristics of each of the three gene regions. The D2 region shows a level of variation that makes it suitable for both diagnostic and phylogenetic purposes, whereas the D3 region is very conservative and does not allow differentiation between several closely related species. The ITS-1 region, though highly variable, leads to results very similar to those from the D2 region; it could be used for diagnostic purposes on the population level, or in cases where the D2 sequences of closely related species are identical. An example are Italian populations of *Encarsia asterobemisiae* Viggiani & Mazzone and *E. lutea*, which apparently have identical D2 sequences (authors' unpublished data) but are morphologically distinct and have different numbers of chromosomes (Baldanza et al. 1999).

Considering the large number of species in the genus *Encarsia*, there are still relatively few sequence data available. However, the D2 region is promising because it supports most of the traditionally defined groupings within the genus. Our study shows that the ITS-1 region has similar properties, leading to higher bootstrap support values for closely related species pairs; for example, *E. accenta* + *E. azimi* received 71% support in the ITS-1 cladogram versus 64% in the D2 tree. Furthermore, the tree resulting from the combined data set shows a higher level of resolution and similar or even higher support values compared to the trees from the separate analyses, which demonstrates that combining sequence data for congruent gene regions with different levels of variability may lead to improved phylogenetic reconstructions.

Appendix

Encarsia species used in this study; species-group placements according to Hayat (1989, 1998), Polaszek et al. (1992), Huang and Polaszek (1998), Schmidt et al. (2001).

Species-group	Species	Code	Locality
Cubensis	*E. nigricephala*	nig218a	French Polynesia, Tahiti
Inaron	*E. accenta*	P15364a	Australia, South Australia, Renmark
Inaron	*E. adusta*	P18166a	Australia, Northern Territory, Darwin
Inaron	*E. azimi*	P17259a	Australia, Queensland, Ayr
Inaron	*E. inaron*	EinaronNZa	New Zealand
Lutea	*E. lutea*	P06220a	Cook Islands, Rarotonga
Lutea	*E. lutea*	P16175a	Australia, Western Australia, Wanneroo
Luteola	*E. formosa*	P07157a	Australia, ACT, Canberra
Luteola	*E. formosa*	P07M92030a	USA
Luteola	*E. guadeloupae*	P22868a	Tonga
Luteola	*E. dispersa*	hait105a	Nauru
Parvella	*E. pergandiella*	P10873a	Australia, Queensland, Dalby
Parvella	*E. pergandiella*	P10US396a	USA
Perflava	*E. cibcensis*	P04210a	Cook Islands, Rarotonga
Smithi	*E. smithi*	P14291a	Micronesia, Truk
Strenua	*E. bimaculata*	P12162a	Australia, Northern Territory, Darwin
Strenua	*E. bimaculata*	P12Buna	Australia, Queensland, Bundaberg
Strenua	*E.* nr *citri*	P13706a	Australia, Queensland, Mundubbera
Strenua	*E. oakeyensis*	P12Aoakey	Australia, Queensland, Oakey
Strenua	*E. protransvena*	P05219a	French Polynesia, Tahiti
Strenua	*E. sophia*	P11261a	Australia, Queensland, Ayr
Strenua	*E. sophia*	P21223a	French Polynesia, Tahiti

References

Babcock, C.S., Heraty, J.M., De Barro, P.J., Driver, F., Schmidt, S., 2001. Preliminary phylogeny of *Encarsia* Förster (Hymenoptera: Aphelinidae) based on morphology and 28S rDNA. Mol. Phylog. Evol. 18, 306–323.

Baldanza, F., Gaudio, L., Viggiani, G., 1999. Cytotaxonomic studies of *Encarsia* Förster (Hymenoptera: Aphelinidae). Bull. Entomol. Res. 89, 209–215.

Brust, R.A., Ballard, J.W.O., Driver, F., Hartley, D.M., Galway, N.J., Curran, J., 1998. Molecular systematics and hybrid crossing identify a third taxon, *Aedes* (*Halaedes*) *wardangensis* sp.n., of the *Aedes* (*Halaedes*) *australis* species group (Diptera: Culicidae). Can. J. Zool. 76, 1236–1246.

De Barro, P.J., Driver, F., 1997. Use of RAPD PCR to distinguish the B biotype from other biotypes of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). Aust. J. Entomol. 36, 149–152.

Dowton, M., Austin, A.D., 1997. Evidence for AT-transversion bias in wasp (Hymenoptera: Symphyta) mitochondrial genes and its implications for the origin of parasitism. J. Mol. Evol. 44, 398–405.

Evans, G.A., Polaszek, A., Bennet, F.D., 1995. The taxonomy of the *Encarsia flavoscutellum* species-group (Hymenoptera: Aphelinidae), parasitoids of Hormaphididae (Homoptera: Aphidoidea). Oriental Insects 29, 33–45.

Giribet, G., Wheeler, W.C., 1999. On Gaps. Mol. Phylog. Evol. 13, 132–143.

Hayat, M., 1989. A revision of the species of *Encarsia* Förster (Hymenoptera: Aphelinidae) from India and the adjacent countries. Oriental Insects 23, 1–131.

Hayat, M., 1998. Aphelinidae of India (Hymenoptera: Chalcidoidea): a taxonomic revision. Mem. Entomol. Int. 13, 1–416.

Heraty, J.M., Polaszek, A., 2000. Morphometric analysis and descriptions of selected species in the *Encarsia strenua* group (Hymenoptera: Aphelinidae). J. Hymenoptera Res. 9, 142–169.

Heraty, J.M., Woolley, J.B., 2002. *Encarsia* Species of the World, a Searchable Database. http://chalcidoids.tamu.edu/ENCARSIA/encarsia.htm

Huang, J., Polaszek, A., 1998. A revision of the Chinese species of *Encarsia* Förster (Hymenoptera: Aphelinidae): parasitoids of whiteflies, scale insects and aphids (Hemiptera: Aleyrodidae, Diaspididae, Aphidoidea). J. Nat. Hist. 32, 1825–1966.

Lee, M.S.Y., 2001. Uninformative characters and apparent conflict between molecules and morphology. Mol. Biol. Evol. 18, 676–680.

Manzari, S., Polaszek, A., Belshaw, R., Quicke, D.L.J., 2002. Morphometric and molecular analysis of the *Encarsia inaron* species-group (Hymenoptera: Aphelinidae), parasitoids of whiteflies (Hemiptera: Aleyrodidae). Bull. Entomol. Res. 92, 165–176.

Martin, J.H., 1987. An identification guide to common whitefly pests of the world (Homoptera: Aleyrodidae). Trop. Pest Manage. 33, 298–322.

Martin, J.H., 1999. The whitefly fauna of Australia (Sternorrhyncha: Aleyrodidae). A taxonomic account and identification guide. Commonwealth Scientific and Industrial Research Organization, Technical Paper 38, 197pp.

Mathews, D.H., Sabina, J., Zuker, M., Turner, D.H., 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911–940.

Noyes, J.S., 2002. Interactive Catalogue of World Chalcidoidea, second ed. Taxapad and The Natural History Museum, London.

Nunn, G.B., Theisen, B.F., Christensen, B., Arctander, P., 1996. Simplicity-correlated size growth of the nuclear 28S ribosomal RNA D3 expansion segment in the crustacean order Isopoda. J. Mol. Evol. 42, 211–223.

Polaszek, A., 1991. Egg parasitism in Aphelinidae (Hymenoptera: Chalcidoidea) with special reference to *Centrodora* and *Encarsia* species. Bull. Entomol. Res. 81, 97–106.

Polaszek, A., Evans, G.A., Bennett, F.D., 1992. *Encarsia* parasitoids of *Bemisia tabaci* (Gennadius) (Hymenoptera: Aphelinidae, Homoptera: Aleyrodidae): a preliminary guide to identification. Bull. Entomol. Res. 82, 375–392.

Polaszek, A., Manzari, S., Quicke, D.L.J., 2004. Morphological and molecular taxonomic analysis of the *Encarsia meritoria* species-complex (Hymenoptera, Aphelinidae), parasitoids of whiteflies (Hemiptera, Aleyrodidae) of economic importance. Zool. Scripta 33, 403–421.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: a Laboratory Manual, second ed. Cold Spring Harbour Laboratory Press, NY.

Schmidt, S., Naumann, I.D., De Barro, P.J., 2001. *Encarsia* species (Hymenoptera: Aphelinidae) of Australia and the Pacific Islands attacking *Bemisia tabaci* and *Trialeurodes vaporariorum* (Hemiptera: Aleyrodidae)—a pictorial key and descriptions of four new species. Bull. Entomol. Res. 91, 369–387.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MS.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choices. Nucl. Acids Res. 22, 4673–4680.

Venkateswarlu, K., Nazar, R., 1991. A conserved core structure in the 18–25S rRNA intergenic region from tobacco, *Nicotiana rustica*. Plant Mol. Biol. 17, 189–194.

Viggiani, G., Mazzone, P., 1979. Contributi alla conoscenza morfo-biologica delle specie del complesso *Encarsia* Förster—*Prospaltella* Ashmead (Hym. Aphelinidae). 1. Un commento sull'attuale stato, con proposte sinonimiche e descrizione di *Encarsia silvestrii* n.sp., parasita di *Bemisia citricola* Gom. Men. (Hom. Aleyrodidae). Boll. Lab. Entomol. Agrar. 'Filippo Silvestri', Portici 36, 42–50.

Whitfield, J.B., Cameron, S.A., 1998. Hierarchical analysis of variation in the mitochondrial 16S rRNA gene among Hymenoptera. Mol. Biol. Evol. 15, 1728–1743.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucl. Acids Res. 31, 3406–3415.