Insect Molecular Biology (1995) 4(3), 203–215

Phylogeny of the Coleoptera based on mitochondrial cytochrome oxidase I sequence data

D. E. Howland and G. M. Hewitt
School of Biological Sciences, University of East Anglia, Norwich

Abstract

A 400 base pair region of the mitochondrial cytochrome oxidase I gene (COI) was sequenced for thirty-seven species of beetle, representing fifteen families. The sequence was found to be highly variable, with 12–18% divergence within families, and up to 27% divergence between families. Phylogenetic analysis using the neighbour-joining method shows the Carabidae (Adephaga) as a distinct clade, but also shows that all other beetles (Polyphaga) diverged soon after the Adephaga/Polyphaga split. Whilst some species are grouped within their respective families, others are too diverged for easy resolution. The main disagreement with trees constructed using comparative morphology and the fossil record is the position of the Coccinellidae. Whilst COI may be an informative gene for molecular systematics at lower taxonomic levels, or in other insects, improved resolution of this particular phylogeny will require a more highly conserved sequence.

Keywords: molecular phylogeny, mitochondrial DNA, cytochrome oxidase I, Coleoptera.

Introduction

The Coleoptera is an ancient group of insects, which has evolved into a large and highly successful order, with an estimated 370,000 described species, and many more awaiting discovery. They have an enormous diversity of life-style and behaviour, and can be found in every terrestrial ecosystem.

The evolutionary history of the Coleoptera is based on a fairly rich fossil record. Earliest records of 'Protocoleoptera' are from the lower Permian, ~280 million years ago (Crowson, 1981). These species had many similarities to recent Archostemata, an ancient suborder, but had some features which differed from all modern Coleoptera. By the Upper Permian, features became more like modern beetles, and by the Triassic there were definite suborders in existence. The enormous radiation of the Coleoptera, and of other insects, during the Jurassic (~136–190 million years ago) is clearly seen in the fossil record (Labandeira & Sepkoski, 1993). Many of these fossil Coleoptera can be attributed to extant families, and it is thought that all modern superfamilies were established as distinct lineages well before the end of the Jurassic period.

Despite the wealth of fossil data, the phylogeny of the Coleoptera depends to a great extent on inferences from comparative studies of modern beetle morphology. The aim of our study was to look at phylogenetic relationships based on molecular data. As the Coleoptera has been evolving for a considerable time, it was necessary to study a highly conserved coding sequence. There is a substantial body of literature on the molecular phylogeny of mitochondrial genes, and these have often been used to resolve phylogeny at and above species level (Avise *et al.*, 1987; Moritz *et al.*, 1987; Smith & Patton, 1993). For this study we chose the mitochondrial cytochrome oxidase I gene (COI). A number of primers were already available to us, and from the literature and from personal communication we had reason to believe that this region would have a useful level of variation for our purposes. A 400 base pair region of the COI gene was sequenced for thirty-seven species of beetle, representing fifteen of the major beetle families (Table 1).

Received 6 January 1995; accepted 29 June 1995. Correspondence: Dr D. E. Howland, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH.

Results

Cytochrome oxidase I sequence

Figure 1 shows the aligned nucleotide sequences of thirty-seven species of Coleoptera for a 400 bp region of the COI gene. No insertions or deletions were found in this region. Variation within and between families was high. Sequence divergence ranged from 12% to 18% within families, and up to 27% between families. All of the third codon positions were variable, whilst first and second positions were more highly conserved, with only 42% of first and 18% of second

Table 1. List of beetles sampled. One species from each genus was sequenced.

Family	Species
Cantharidae	*Cantharis* sp.
	Rhagonycha fulva
Carabidae	*Agonum* sp.
	Carabus violaceous
	Chlaenius sp.
	Harpalus aureus
	Laemostenus sp.
	Notiophilus sp.
	Pterostichus sp.
Cerambycidae	*Agapanthia* sp.
	Leptura sp.
	Rhagium sp.
Chrysomelidae	*Chrysolina* sp.
	Chrysomela sp.
	Lachnaia sp.
	Oulema sp.
	Timarcha sp.
Cicindelidae	*Cicindela* sp.
Coccinellidae	*Adalia bipunctata*
	Calvia quatuordecimpunctata
	Coccinella septempunctata
	Exochomus sp.
Cucujidae	*Oryzaephilus surinamensis*
Curculionidae	*Otiorhynchus* sp.
	Phyllobius sp.
Dermestidae	*Attagenus pellio*
Hydrophilidae	*Cercyon* sp.
Lagriidae	*Lagria* sp.
Lucanidae	*Dorcus parallelopipedus*
	Lucanus cervus
Scarabaeidae	*Copris lunaris*
	Melolontha melolontha
	Scarabaeus sp.
Staphylinidae	*Tachyporus* sp
	Xantholinus sp.
Tenebrionidae	*Tenebrio molitor*
	Tribolium sp.

positions varying. Such a pattern of variation is typical of a sequence under strong functional constraints, where most of the substitutions accumulate at third codon position and usually do not affect amino acid composition.

Highly biased base composition was seen in these sequences (Table 2). As with other insect mtDNA genes (Clary & Wolstenholme, 1985; Crozier & Crozier, 1993), there was a high AT content (80–85%). This was due mainly to bias at the third codon position which was almost exclusively composed of adenine and thymine. The bias

Table 2. Base composition of 400 bp sequence of cytochrome oxidase I at first, second and third codon positions, based on consensus sequence of thirty-seven species of Coleoptera.

Codon position	A (%)	C (%)	G (%)	T (%)
First	28.6	11.3	30.8	29.3
Second	18.0	24.1	13.5	44.4
Third	51.5	0.7	0.0	47.8

against guanines, seen at the third position, is also found in other animal mtDNA genes (Brown, 1985). There is also a clear bias in favour of thymine at the second position (also noted by Irwin *et al.*, 1991, in mammalian cytochrome b), and against cytosine at the first position.

Variability along the gene

The inferred amino acid sequence (Fig. 2) shows that the polypeptide is in fact highly conserved. However, there are two segments which are variable: peptides 49–69 and 109–133. When a topographical model of the COI protein within the mitochondrial membrane was consulted (Saraste, 1990), it could be seen that these variable regions corresponded exactly to the segments on the matrix side of the mitochondrial membrane (Fig. 3). In contrast, regions which spanned the membrane, and loops emerging on the cytoplasmic side of the membrane, were highly conserved. Figure 4 shows the amount of variability across the sequence, and clearly identifies the two inner membrane loops as more variable. Cytochrome oxidase is an enzyme involved in electron transport across the mitochondrial membrane, and consists of three subunits coded by the mitochondrial genome, and up to ten coded by the nuclear genome. DNA sequence data for COI is available from more than seventy-five different organisms, and has shown that a number of amino acid residues are totally or highly conserved. Site-directed mutagenesis on these residues has revealed their roles in binding to metal prosthetic groups within the subunit (Calhoun *et al.*, 1994). Such findings illustrate the extent of functional constraints on a DNA sequence, and also indicate how important good sequence data is for constructing new primers for amplifying sequences from a range of species.

Phylogenetic analysis

The neighbour-joining method of phylogenetic analysis was chosen in this study. The main alternative, parsimony, is only appropriate if evolutionary rates vary by no more than around 2-fold in different branches (Felsenstein, 1985), and so was not thought to be appropriate in this case. Neighbour-joining is less sensitive to this problem (Nei, 1991).

A number of analyses were carried out, using various weighting methods. By giving different substitution rates to each codon position in the analysis, we obtained sufficient resolution to produce a tree where many families can be distinguished (Fig. 5). Without this method, i.e. giving no weighting at all, the saturation at the third codon position resulted in nonsense trees. Neighbour-joining trees based on second codon position only, produced a broadly similar tree to that shown in Fig. 5, but did not improve the resolution. Protein sequences were far too conserved to

```
Adalia        AGCTATTGGTTTACTAGGATTTGTTGTATGAGCACATCATATATTTACTGTTGGAATAGACGTCGATACACGAGCTTACTTTACTTCGGCTACTATAATT
Agapanthia    ...A.....A...T..........A........C..C...........A.....T.....T..T..............T.....C..A..A.........
Agonum        .........AC..T..........C..T.....T..............X..A........T..T.....T.....A..T........A..A..A......
Attagenus     .........AC.T..T......A.............C...........A..A...........A..C........C...........C..A.........
Calvia        .........GC..T....C..............T..............A...........T..A.....T........T.....C..T.....A......
Cantharis     ...A.....A.....C.................C........G........A..G..G..T..T..C.................A..A..A.........
Carabus       ...A.........T.............T.......................A........T..T..C..T.....C........A..T...........C
Cercyon       .........A...T....T...A..........C..............A......T....T..T.....T........T.....A..A..A.........
Chlaenius     ..........C..T....T..............T..............A..A.......X...T.....T........T.....A..A.....A......
Chrysolina    ...A.....A.....C......A..........C..C...........CA.C........T..T.....C...........C..A..A..A.........
Chrysomela    ...A.....A...T........A..........T...............A..........T..A..............T........A..A.........
Cicindela     ...C.....AC.CT..........A........T.................A........T..A...........C..T..C..C..A.....G......
Coccinella    ...C.....GC.T...........A........T..C...........G..A........T..A...........A...........A............
Copris        ...A.....G...T........A.......G..T.....C...............T....T..T.....C........T.....A..C..A.....T...
Dorcus        ...A......C.CT....T..CA.C..T.....T.....C.....C..C..............T.....C.....C...........A.....A......
Exochomus     .........CC....T..T........T.....T.................A...........A.....T........T........A............
Harpalus      T............T....T................................A........T..A..............T........A..A..A......
Lachnaia      ...C...C.G.CT.....C..C..A....................C.....A..T.....T..T..............T..C..A..T..C..A......
Laemosenus    ...C.....GC.TT..........A........T..............XA..........T..A.....T..GX....T........A..A....XXX..
Lagria        G.........C.TT..........A..T.....C.................G..C..G..T..T........T..C..T.....A..A..A..A......
Leptura       ...A.....A............A..........C..............A..A........T..T.....C.....A..T........C.....A......
Lucanus       G..A.....CC.T..T.....CA..........T..........................T..G.....C.....C.....C..A..A..A........C
Melolontha    ......................A.......G..C..C........C..A.....G........T.....T..T..C........C..A.....A......
Notiophilus   T..A.....A.....T...........C.....T..C..............A........T..A.....T..T.....T........A..A..A......
Oryzaephilus  ...A.....CC.TT....G...A....C.....T.....C........A..A........T..A..C..T.....A..T........T..A..A......
Otiorhynchus  .A.A.....G.....T........A..T........C..C.....C..A..A........T..A.....C.....C..T........A..A..C......
Oulema        ...A..C..G..TT........XXX..T.XXXXC..C..............A........T..G..C..T.................A..A..A......
Phyllobius    ...A.....GC..T.......C....A.....................A..A........T..A.....T........T........A............
Pterostichus  .........A...T....T.............................A...........T..A.....T........T........A.....A......
Rhagium       ...A.....AC.C..T........A..T....XXX.C...........A...........T..A.....T........T........A..A.........
Rhagonycha    ...A.....GC.TT.........XX..T...........C..G........A...........A..............T........A..C.........
Scarabaeus    .A.A.....A...T....T...A....GX...................A..A........T..T.....T........T........A..A..A......
Tachyporus    ...A.....A..G..T.................T..........................T..T..C..T.....A..T.....A..A.....A......
Tenebrio      ...A.....AC..........C..A..T.....T..C........C..C...........T..A..C..T..C..A.....C..A..A..C..A......
Timarcha      .........A...T.....A...G...T..T..T..............A..A........T..A...........A...........A.....A...G..
Tribolium     ...A.....A........G..C..............C..C.....C.....A........T..T..C..T.....C..T..C..A..A..A.........
Xantholinus   ...A.........T........A..........T..............A...........T..T.....T..T.....T.....A..T............
```

Figure 1. Sequence of 400 bp region of mitochondrial cytochrome oxidase I from thirty-seven species of Coleoptera. Dots indicate identity to sequence of *Adalia*. X indicates ambiguous sites. GenBank/EMBL accession numbers for these sequences are X88933–X88969.

```
Adalia        ATTGCAGTACCAACAGGAATCAAAATTTTTTCATGACTAGCACACTTACATGGAGTTCAATTTATTTTTACTCCTTCATTATTATGAACCTTAGGGTTTT
Agapanthia    .....T.....T..T..T..T...G.A...AG....T.....ACT..C.....CTCA...A...C..AC.GA...GTT.CTA.....G.GC........G
Agonum        .....T..T...........T...............T.....ACTC.T.....T.C....A.ATC..A..G..........C.T...G.T.....A...G
Attagenus     ........TT.C..C.....T.........AG....T.....ACTC.TA......CC...A...A......A..AC.TG.........TAC....T...C
Calvia        .....T..............T..........T....T.....ACA...........A.......AA.....A..C...A.................A....
Cantharis     .....T.....T.....T..T.........AGT...T....TACTC.T.....CTCAA..A...A..GA......CA.A.........G.AACT..A...A
Carabus       ...........C........T...........G..G..T...ACTC.T......ACA...A..TC..AC.GC.........C.....G.T.....A...G
Cercyon       .....T.....T........T...............T.....ACTC.T......ACC.....A.AC.A.T.A...............GT.A.....A...G
Chlaenius     .....T..T..T........T.....................ACTC.T.......CC...A.ATC..A..G..........C.T...G.T.....A...G
Chrysolina    ...........C........T...G.A..CAG....T.....AC...C..CX..AC.XX.A.........AC..AGT.ACT......G.A.....A...G
Chrysomela    .....T..........T..T..T...G.....AG....T....TACT..T.....TACAATGA...C.....A...AGT.ACTC.T..GG.T.G...A...G
Cicindela     .....T.....T........T.................T..TACTC.T......TC....A...CA.A..GA..A..CC.TC.....G...........G
Coccinella    .....T.....T........T..G.........T..GT.....AC...G......A.........A...A.....C......C.......T.....A..CC
Copris        ..GCG...G.....T.....T.........AAG.AGA.T...ACTC.C......ACA.......A..ACT.A..A......A.T...G.TC....A...G
Dorcus        ...........C..T.....T........CAGG...T....TACTC.C......ACA...A.C.A..A.T.A..C...A....G..GG.......C..CG
Exochomus     .....T..T...........T.........CGT...T.....ACT.........AC.........A..A..AC.....T......G....T.....A....
Harpalus      .....T..T...........T...............T.....ACT..........CC.....ATC..A..GA...............TG.T.....A...G
Lachnaia      ........C..T........T..GG.A..CAGG...T.....AC..........ACAA....A....A.T.....G.TACTC.....G.T.....C...G
Laemosenus    ........T..T........T........C..T...T.....ACTC.X.....G.C....C.ATC..AC.GA...G..C....G...G.T..G..A...G
Lagria        .....T..T..T........T.........AG....T....TACT.....C............A.AC.ACT.A..G..T.....G...G.T.....C...G
Leptura       .....T...........T..T...G.....AG....A....TAC.C.C.....CAC....A...C..AC.GA..AGT.ACT.......TA.....A...G
Lucanus       .....T..T..T..T...................AGT.....T..TAC.C.T.....TAC....A.A.A..ACT.C..C...A.........G...........G
Melolontha    ........C.....T.....T.........AGT.........AC.C.C..C..CACC...C...AC.A.T....C..CC.T.....GG.A.....C..CG
Notiophilus   .....T..C.....T.....T........C..T...T....TACT..........C......A.CA.A..GA.....TC.TA.....G.T.....A...G
Oryzaephilus  ...........C..T..T.....G.....CAG....A....TACA..C..C...ACAA..A.C.AAAA..AG..AAG.ACTA.....T.TC....A...G
Otiorhynchus  .....T..T...........T........CAG...GA.G...ACT.AC......ACATT.G..TCC....A...AAG..C.C.C...T.AC....A...A
Oulema        ......A.T..T........T.........AG....A.....ACA.........A.AA..A...A..A..GA..AA.TA.CC.T...G.T.....A...A
Phyllobius    .....T..T..T..T.....T........CAG....A.....ACA.AC..C..T.A.TT.A.CTCA..C.AC..AAG..CTC.C...T.A.....A...A
Pterostichus  .....T..T...........T...............T.....ACTC.C..C..T.C......ATC..A..GC.........C.C...G.......A...G
Rhagium       .....T.....T.....T..T...G.....AG....A.....ACTC.T......ACA...A..T.A.AC.GA..AGT.ACT.......TA.....A...G
Rhagonycha    .................T..T.........AGT...T.....ACTC.T.....GTCCA..A..T.A.GA..C...CA.A..C.....G.AACT..A..CA
Scarabaeus    .....T..............T.........AGC...T.....ACA..C.......C....C....A.C.CA...A..TA..A.T...G.AC....A...G
Tachyporus    .....TA.T..T........T.........AG...GT.....ACAC.T......ACA...A...A..ACT....C..TA.........G.TC.T..A...G
Tenebrio      ........C.....................AG....T.....ACAC........ACA...A.A.A..A..GA..A...A.........G..C.T.AT...G
Timarcha      .....T.....C........T...GA...AGA....T....TACAA.......TAA.TTTA.A....AC.AA..AGT.ACT.......TA.....A...G
Tribolium     ........T..C..T.....T.........AG.........CAC.C.T..C...AC......A.A..A..GA..C...C..A.T...G.TC....A...G
Xantholinus   .....TA.T.....T.....T.........AGT...T.....ACAC.T.....TACA...A...A..A.T........A.........G.T.....T...G
```

Figure 1 (*Continued*)

```
Adalia        TATTTCTATTTACAATTGGG GGGCTTACAGGTGTAATTTTAGCTAATTCATCAATAGATATTATTTTACACGATACTTATTATGTTGTAGCCCATTTTCA
Agapanthia    .G..C..............T..AT.A..G..A..TG.A.....A.....C.....T........C....XT..............A..T..T........
Agonum        .....T........G.A..A..AT.A.....AC.TG.C.................T......G..C.T..T.....A...........T..T.....C..
Attagenus     ..............G....A..AT.A.....AA.GG.CC....A..C..C.....T.........C.G..T.....C........A.....T........
Calvia        .......C..C...........TT.A.....A.................T..T............C.T..T..............G.....T........
Cantharis     .T...T....C...G....A..AT.A.....A..T........A..C..T.....T.........C.T..T.....A........A..T..A........
Carabus       .....T....C..CG.A..A..AT.A..G..G...G.A........C........T......G..........C...........A..C..T........
Cercyon       .....T........G....T..A..A.................A...........T.........C.T..T...........C........T.....C..
Chlaenius     .....T.......TG.A..T..TT.A.....A......C................T.....AG.A.....T.........G..........T.....C..
Chrysolina    ....CT........G.A..C..AT.A.....A...G.......C.....T..T..T.........C.C..T.................T..A........
Chrysomela    .....T.......TG.A..A..AT.A.........G.......A..C........T........C.....T.....A........G..C..T........
Cicindela     .T...T........G.A.....AT.A.........G.......A...........T.........C.T..T.....A...........T.....C.....
Coccinella    .............T.....A..AT.A..........................T...........CC.T.................A...........C..
Copris        ....CT........G.A..T..AT.A.....A....................CC........G.CC....T.....C..C..C..A.....T........
Dorcus        .C..C.........G....A..TT.A..T..A...G.AC.T..A..C........T......G..C.C..T........C..C.....T........C..
Exochomus     ...................A..A..A..T..A....................T.................T.....C........A..T..T........
Harpalus      .....T........G.A..A..AT.A..T..A..TG.............T...C.T......G.A.....T..............A..T..A........
Lachnaia      .C..CT.............A..TT.A.....A..TG..C.G.....C..T.....T..C...G.AC.T..T........C.....A........C.....
Laemosenus    .....T....C...G.A..A..AT.A..T..A...G.A.....C.....T.....T..C...G..C.C..T........C.....A.....T.....C..
Lagria        .G...T........G.A..A..AT.A.....G...........A...........T........AC.T..T..............A..............
Leptura       ....CT........G.A.....TT.A..C..G......C....C........T..T..............T.....C........A..T..T........
Lucanus       .T..C..T.....TG.A.....A..A.....G..T...C.G..C........T..T..C..C...C.T..T..C.....C..C..A.....T..C..C..
Melolontha    ....CT........G.G..A..AT.A........TG.G........C..T.....T.........C.T..T.....C..C.....A..............
Notiophilus   .....T....C..TG....A..AT.A.....A..TG.A........C..T..T..T..C...G..C.T..T.....A...........T..T.....C..
Oryzaephilus  ....C..C..C...G.A..A..TT.A.....A..............C.....T..C.....C...C.T..T........C..............C..C..
Otiorhynchus  .T..CT..........G.....C..A..T..G.....CC.T.....C..T..T..T.....C..CC.T..T..............A..T..A........
Oulema        .T...T........G....A..CT.A.....A.................T..T..T..C..C...C.T..T.....C..C.....A..T..T.....C..
Phyllobius    .T...T....C.....A..A..AT.A.....G...G..C....A.....T..T..T.........C.T..T........C...........T..C.....
Pterostichus  .....T........G.......A..A.....A...G...................T.........C....T.................T..T........
Rhagium       ......................AT.A..T..A...G...................T..............T.....A.....C..A..T..T........
Rhagonycha    .T..CT........G.A..A..AT.A..G..G......C.......C..T.....T.........C.T.....C..C..C...........A........
Scarabaeus    .C............G.......A..A.....A..TG.A.....A..C.....T..T......G.......T.....A........A........C.....
Tachyporus    ....CT.......TG.......AT.A..T..A...G.AC.......C..T.....T..............T.............................
Tenebrio      ....CT....C..CG.A..TCCCT.C..C..A......C.G..C.....T.....C........A........C.....C..C..A.....G.....C..
Timarcha      .T.....G.....TG.A..A..AT.A..T..G..TG.............C.....T..............T..............A.....A.....C..
Tribolium     ....CT........G.A..A..T............G.A.....C..C........C........AC.T..T..C....................C..C..
Xantholinus   .....T........G.G..T...T.G..T..A.....CC................C....................A..C..C..A.....T........
```

Figure 1 (Continued)

```
Adalia       TTATGTTCTTTCTATAGGAGCTGTATTTGC TATTATAGCAGGATTTATTCATTGATTTCCATTAATTACAGGATTTAATATAAATAATAAAATATTAAAA
Agapanthia   C.....C..A..G..................................A.....A.....C..C...T..........A.CAT.....G.A...TG.......
Agonum       ......AT.A............A.........A.........T.............A.....C..T...T.......T..A.CA...........GTT..C.....
Attagenus    .........A..A.....T.......................A....C..A.........T...T.......G..AGCCC.T...G.A...T........
Calvia       .......T.A..A.........A..T.............................C...............A..T.........T............T..C..TT.
Cantharis    C......T.A..A.....T..A..T.............AG...T..AG..........C..CC.TG....T...G.A.CAC.T..CC.A..T.AT......
Carabus      ......AT.A.........T..A.........A...T....G...........A.....C...C..T....T.....A.CAT....C.......T........
Cercyon      .......T.A..A.....T..A..G.....A.......GG..TC..G....C.....C..T...T...........A.C........C.C.....T......
Chlaenius    .G....AA.A..A........................T..T........A.....C......T....T..G..A.G......C.G...GC.T......
Chrysolina   .................G..A........C...........GA.......A.....C......T....T.....A.CCT.......C...TAT......
Chrysomela   .......T.A........G..A.................T...A....C..A.....C..T...T.C.....T..A.C.T............TAT......
Cicindela    C.....AT.A..A.........G.........A......T.G..G..C.....A.....C...C..T...........A.C..........C.G.T..C.T...
Coccinella   ...C..A.....C............................G.A...........T.....A.......................C..C.....
Copris       .......T.A..A.........A.................G......G.A........C..CC.GT.C.........A.C...............T.T......
Dorcus       ......C.........G..C..A.....C...............T...G.............TC.TT....T...C..TCGT....C.......C........
Exochomus    .......T.A...............G.....C...............A..........A...TA.TC.......T.A...........T.....C.T......
Harpalus     .......T.A........G..A................................A.........T...T....T.....ATG.T............T........
Lachnaia     ......G.....C..G.............C...............T..C..C..A.........TC..T....T.....A.CAT....C....CCC.GGCT...
Laemosenus   ......A..A..A............T...............C..T........A.........C...T....T.....GXGA...........GTT..C.....
Lagria       ............A.........A..T.....A........G............A......A...T...T.......G..GTC.T....CG.A...T...GT...
Leptura      C......T.A..A..G..T..A..G.....A..C.....T..TA........A.....C..T...T....T..C..A.C.T....C..........C.T...
Lucanus      ......A.....A.....T..A..T..............G...G...G..........C..C...T.......TC...C.CT....C.GA..GT.C......
Melolontha   C.....A..A..G.........A.........................G.C...........G...T....C.....A.CCC....CCC....TAC.....G
Notiophilus  ......AT.A..A.........A..................................A.........T...T.C.....T..AGT................T.....C..
Oryzaephilus C.....A..A..C.........A......................AG....A.....C...C.......T..T..A.CCC.TC.CG.CCT..CC.C..T.
Otiorhynchus ...C...T.A.........C..A..T..........T.........A.CG....A.........CC..T.......T..A.C.T............TATC.....
Oulema       .......T.A.........G.......................G...TA.C..C..A.........C...T....T.....A.C.T.........T.TT.C......
Phyllobius   .................C...........A.....T......A.AG.A..A.......C..CC..T...........A.CCT...............T.T......
Pterostichus .........A..A.........A..........................T..............A.........T...T...........A.G.T...........TC.T......
Rhagium      ...C...T.A..A.........A..C.......................A.........A.....C..T...T....T..T..A.C.T..........TT..C......
Rhagonycha   ......C.................G.....A......AG...T..A.........G.....TC..G.C.....TG.A.C.T....CC.A.CTCAT......
Scarabaeus   ............A.....T..A.........A.........T......G..........C...C..T...........A.CA......TC...GC.C......
Tachyporus   C...........A......................A.........T.....A..A..A.............T....T..T..A.TA......G.A...T.C......
Tenebrio     ...C..AT.A..A.........A..G..C..C..C.....C...C.AG.............TC.TT.C.....CC.ATCAT....CCC.........GT...
Timarcha     C.....C.................A.......X...............A.AG.C..A.....C......T...........A.CCT.......C...TAC......
Tribolium    ...C.....A..C..G...............G..............C.AG....C.....C..CC..T....T...G.AGCA.....CCCA...C...GT...
Xantholinus  .......T.A.................T........................AG.A.........C......T....T.....AGC.T.....G.....T.T......
```

Figure 1 (*Continued*)

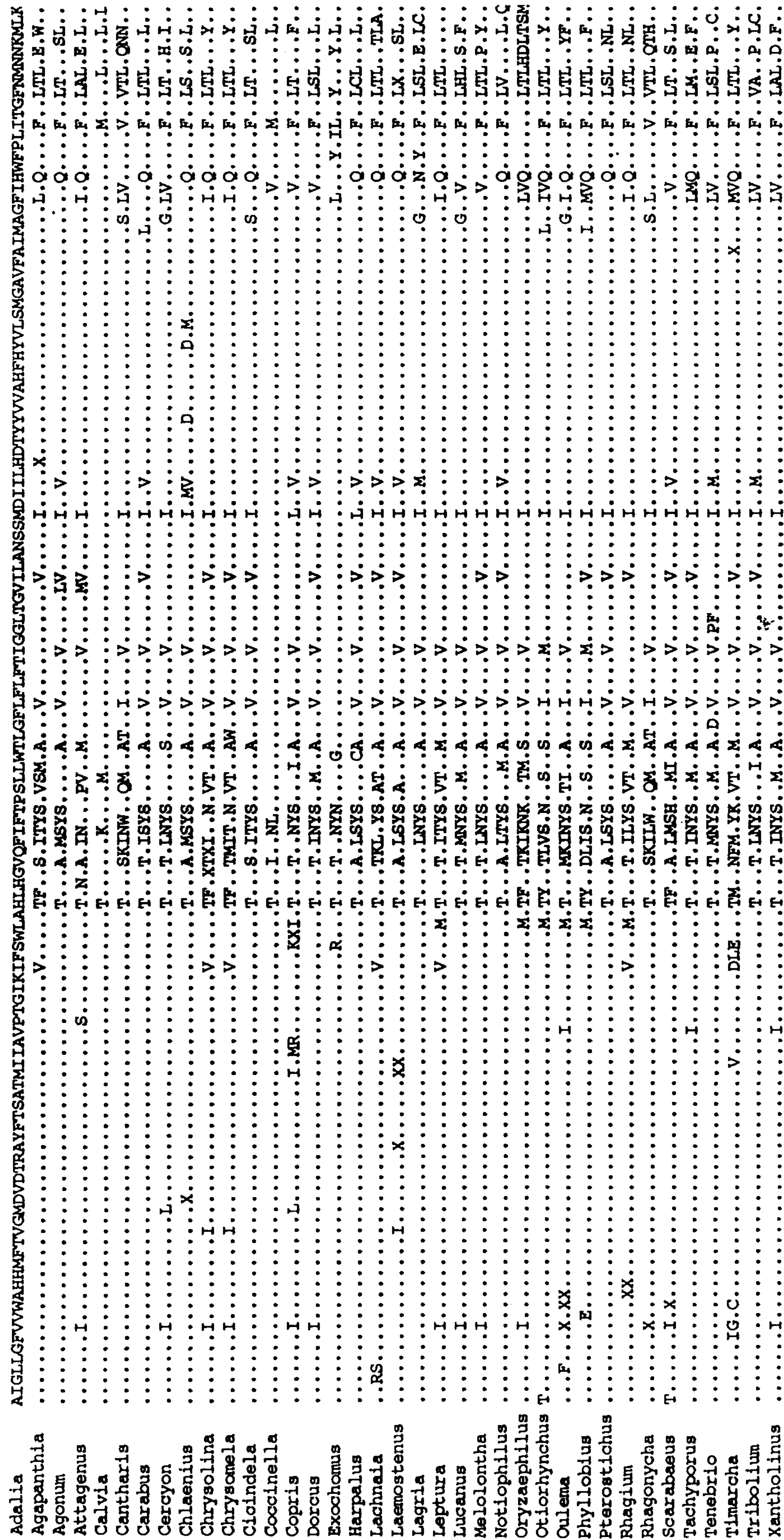

Figure 2. Inferred amino acid sequence of 400 bp region of cytochrome oxidase I for thirty-seven species of Coleoptera. Dots indicate identity to sequence of *Adalia*. X indicates ambiguous sites.

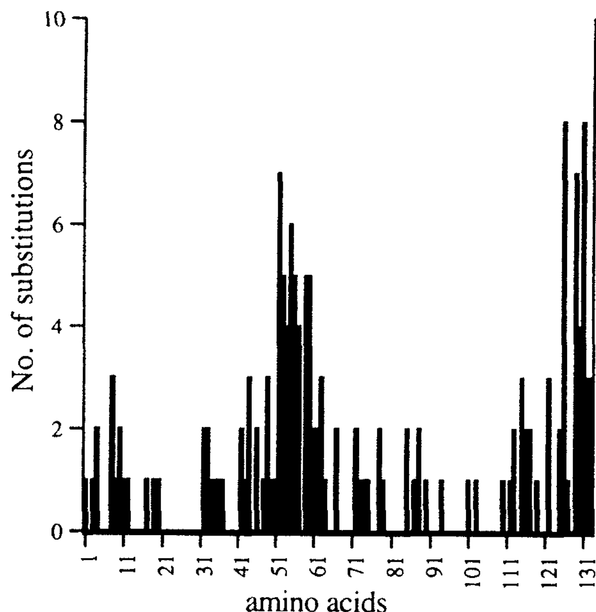

Figure 3. Variation across a 400 bp region of the cytochrome oxidase I gene. Variation is shown as the number of substitutions occurring at each site in thirty-six species compared to a reference species (*Adalia*).

provide any useful data for phylogenetic analysis – a much longer sequence would be needed for this method. Other researchers working on highly variable sequences have constructed trees based either on first and second codon positions only (Edwards *et al.*, 1991), or on first and second positions together with transversions at third codon position (Irwin *et al.*, 1991).

Figure 5 shows an unrooted neighbour-joining tree based on our sequence data. No outgroup was included in the study, hence the lack of rooting. The *Apis* COI sequence (Crozier *et al.*, 1989) was included in several of the analyses but rooted the phylogeny at *Scarabaeus*. Similarly, *Drosophila* (Clary & Wolstenholme, 1985) was also misplaced within the tree. Finding an outgroup in such a broad study is clearly problematical.

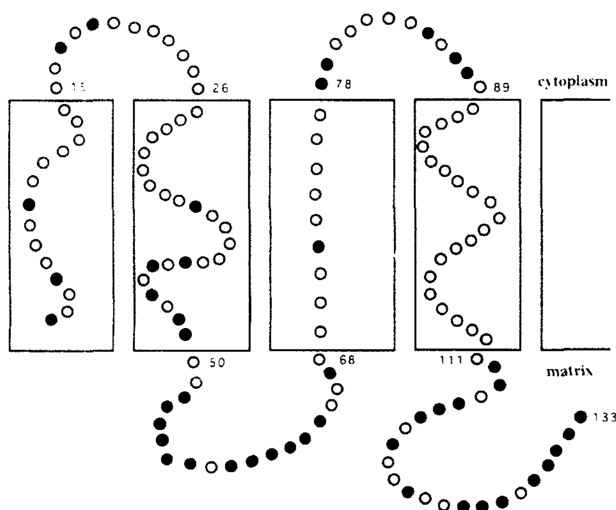

Figure 4 Variation in 133 amino acid residues of cytochrome oxidase I gene in the Coleoptera. ○, conserved residues; ●, residues varying in more than one species (adapted from Saraste, 1990).

It will be noted that bootstrap values are not given in Fig. 5. The validity of bootstrapping is questionable when analysing functional DNA sequences. The main assumptions underlying bootstrapping are that variation at the sites to be resampled are independent, and identically distributed (Felsenstein, 1985). It seems unlikely that either of these assumptions are correct when tertiary structure of a protein has to be maintained. The routine practice of bootstrapping in phylogenetic studies such as this is unsound, whilst the effects of functional constraints and non-random mutation have yet to be assessed (Brown, 1994).

The tree in Fig. 5 shows branch lengths which are proportional to the degree of dis-similarity, i.e. the longer the branch length, the more diverged the species is from neighbouring species. The main feature of this tree is the very short internal branches, indicating that the major groups diverged over a short time-scale. Most of the evolution in the Coleoptera has occurred since the divergence of these groups. Despite this, many of the families have been resolved.

The Carabidae form a distinct clade together with *Cicindela*, which is often regarded as belonging to a subfamily of the Carabidae. This group belongs to a separate sub-order (Adephaga) to all the other beetles (Polyphaga) and is regarded by taxonomists as a clear monophyletic group (Crowson, 1981). Within this suborder, all but *Carabus*, *Cicindela* and *Notiophilus* belong to a morphologically well-defined monophyletic group, suggesting that in our phylogeny *Notiophilus* is out of place.

Shortly after the Carabidae diverged there was apparently a great divergence of the Polyphaga. Within the Polyphaga, the Coccinellids form a very distinct monophyletic clade. *Calvia*, *Adalia* and *Coccinella* fall close together, and are all members of the subfamily Coccinellini, whilst *Exochomus* belongs to a distinct subgroup of Coccinellids, the Chilocorini. The Cantharidae, Cerambycidae, Curculionidae and Tenebrionidae also form distinct clades. Most of the other families with more than one species have been split. The Lucanidae, Scarabaeidae and Staphylinidae have not been fully resolved. The Chrysomelidae form one cluster of three species, all belonging to the Chrysomelinae (*Chrysomela*, *Chrysolina* and *Timarcha*), but two other species which belong to other subfamilies are split from this group. Above the family level, some superfamilies have resolved and others have not. The Cerambycidae and Chrysomelidae are closely related, both belonging to the Chrysomeloidea, and this is reflected in the tree. The Lucanidae and Scarabaeidae belong to the Scarabaeoidea, and the Cucujidae, Coccinellidae and Tenebrionidae

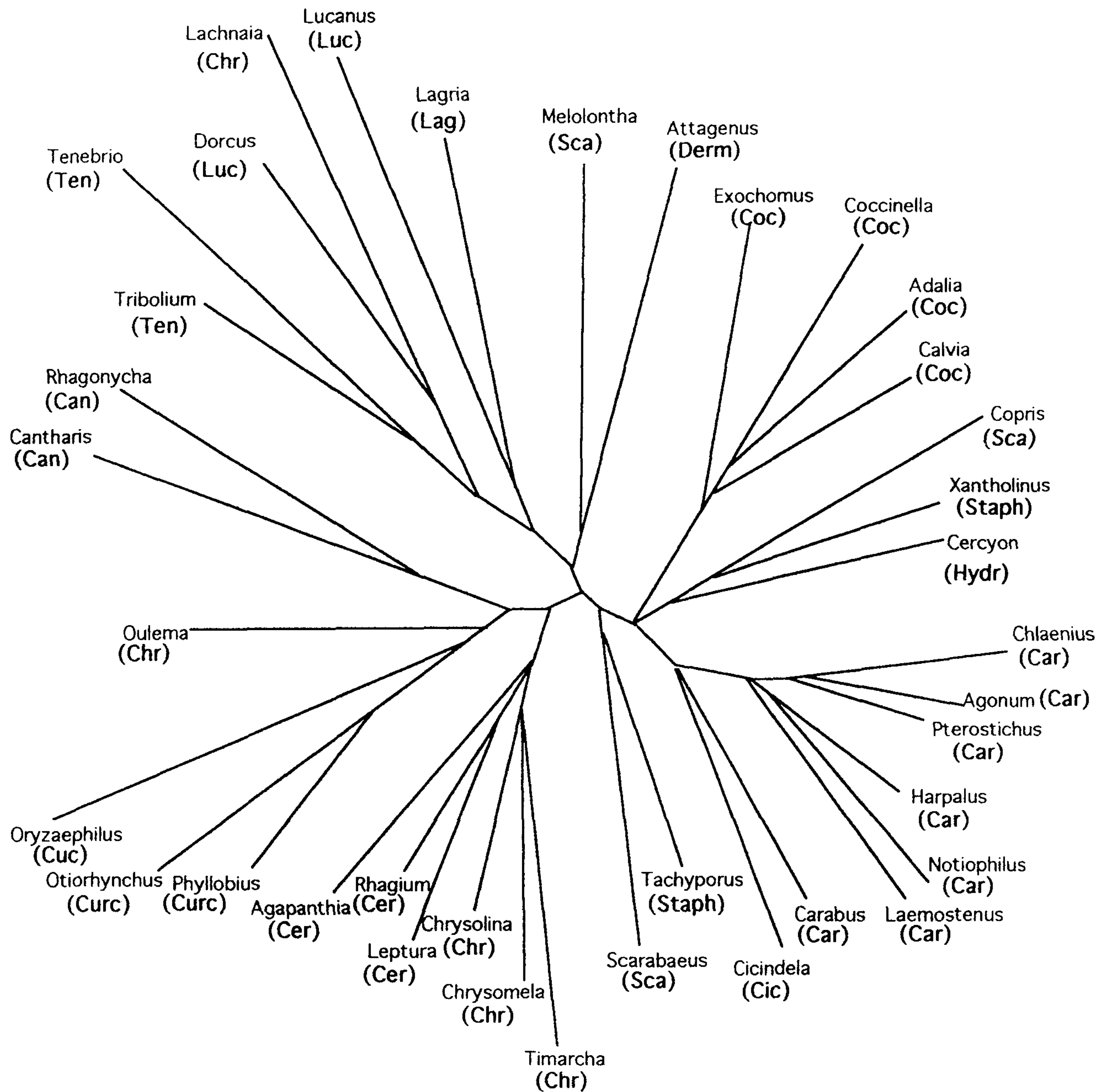

Figure 5. Unrooted neighbour-joining tree based on a 400 bp sequence of cytochrome oxidase I. Can, Cantharidae; Car, Carabidae; Chr, Chrysomelidae; Cer, Cerambycidae; Cic, Cicindelidae; Coc, Coccinellidae; Cuc, Cucujidae; Curc, Curculionidae; Derm, Dermestidae; Hydr, Hydrophilidae; Lag, Lagriidae; Luc, Lucanidae; Sca, Scarabaeidae; Staph, Staphylinidae; Ten, Tenebrionidae.

belong to the Cucujoidea. Neither of these two superfamilies have been resolved.

Perhaps the most striking difference between this tree and the classic taxonomy based on comparative morphology and the fossil record, is the position of the Coccinellidae. This family is clearly well-defined, but resolves close to the Carabidae in all trees that we have produced, rather than with other members of the Cucujoidea.

The variability of the COI sequence has clearly limited the powers of resolution above the family level. In some cases it is also too variable to place genera in their respective families. In these families, such as the Scarabaeidae and Tenebrionidae, the branch lengths for the species are very long, indicating that members of these families are in fact highly diverged from each other. This could explain some of the confusions encountered here.

We compared the neighbour-joining tree to a simple tree based on morphology and the fossil record (Crowson, 1981). If the tree is redrawn (ignoring branch lengths) to make the two trees comparable (Fig. 6), there seem to be

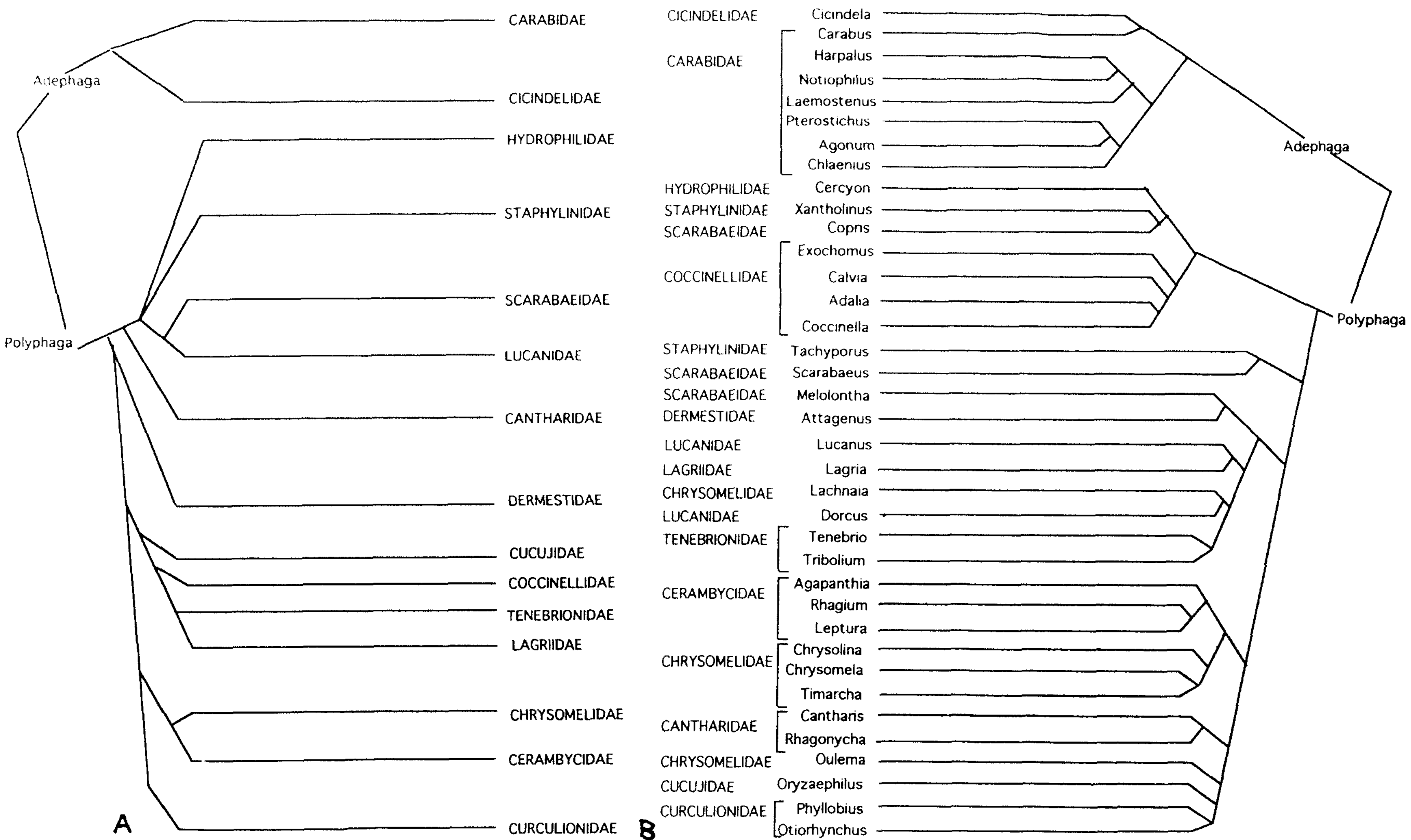

Figure 6. (A) Phylogeny of the Coleoptera (adapted from Crowson, 1981), showing families analysed in this study. (B). Cytochrome oxidase I phylogeny of species sequenced in this study. Family names are included on left-hand side of species names for comparison with Crowson's phylogeny.

many broad similarities. This suggests that some of the deepest branches in the molecular tree and morphological tree concur, and may in fact be correct. However, the positioning of the Cantharidae, and in particular the Coccinellidae, are clearly problematical.

Discussion

The COI sequence which we chose to study was more variable in the Coleoptera than we had anticipated. Initial work on a small number of samples had suggested lower levels of variation. Sequence variation between genera of one family was in some cases as great as variation between members of different families. This inevitably lowers the level of resolution of phylogenetic trees. Studies of other insect taxa using DNA sequence data have also encountered problems with resolving higher level phylogenies. Liu & Beckenbach (1992) found the COII gene to be too variable to unambiguously resolve ten holometabolous insect orders. In other studies of holometabolous insects, based on 18s rDNA, substitution rates in the sequence have been too *low* to resolve most relationships (Carmean *et al.*, 1992; Pashley *et al.*, 1993). At the generic level, COI and COII genes have produced a phylogenetic tree of the lepidoptera *Greya*, which is largely in agreement with that produced from morphological data (Brown *et al.*, 1994). Clearly, it is important to choose gene sequences which have suitable rates of substitution for the level of analysis required. Simon *et al.* (1994) point out that generalizations regarding the usefulness of particular genes for particular taxa are impossible to make. However, increased sequence data will give us a better understanding of how genes evolve, and which are appropriate to use in phylogenetic analyses.

Our particular phylogeny holds together at different taxonomic levels for different groups of Coleoptera. Some families and superfamilies are clearly defined, whilst others are scattered throughout the tree. It is inevitable that data from one sequence will be insufficient to resolve phylogenetic relationships at all levels within such a large group as the Coleoptera. Molecular data from one gene sequence reflects the evolution of that gene only, and this may be very different to phylogenies based on other characters. However, traditional taxonomic methods can also be misleading, as morphological characters are subject to convergent evolution. Brower (1994), for example, found many discrepancies between molecular and morphological phylogenies in *Heliconius* butterflies, where mimicry has resulted in wing patterns which are diverse amongst races of a single species, but can be almost identical between species.

Despite the high level of variation in COI, we have produced a neighbour-joining tree from our data which shows the Carabidae as a distinct clade, and also shows that the Polyphaga diverged, to a considerable extent, in a short time. Such findings are consistent with fossil records, where extensive adaptive radiation can be seen throughout the Jurassic. For an improved resolution in the molecular phylogeny of the Coleoptera, several molecules will need to be examined from many more samples. Acquiring a representative sample of sequences from all families or even superfamilies will take some time in such a large group. Our work here is comparable, although on a smaller scale, to the seed-plant phylogeny produced by Chase *et al.* (1993), using chloroplast rbcL sequences. Their phylogeny shows many discrepancies with traditional views. However, it is not considered to be authoritative, but is an initial hypothesis which can be analysed further using more molecular or morphological characters.

We have shown that the COI sequence *can* be informative at some levels, and it is anticipated that below the family level, and perhaps even at the species level, this sequence will be extremely useful in examining phylogenetic relationships, both in the Coleoptera and in other insects.

Experimental procedures

Samples

Samples were collected locally, or were sent to us preserved in ethanol by colleagues. A single individual of each species was sampled. Genera, and where possible, specific names are listed in Table 1.

DNA extraction

Beetles were ground in liquid nitrogen with sand and incubated on ice for 30 min in 15 ml of extraction buffer (50 mM Tris pH 8.0, 300 mM EDTA). Samples were spun down, and the pellet resuspended in 5 ml of sarkosyl buffer (50 mM Tris pH 8.0, 1 mM EDTA, 1% (w/v) N-laurylsarcosine (Sigma)). After a 30 min incubation on ice, samples were extracted twice with phenol/chloroform/iso-amylalcohol (25:24:1), and the DNA precipitated with 0.1 vol. 3 M NaAc and 2/3 vol. propan-2-ol.

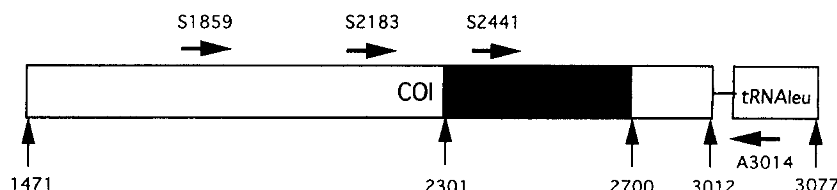

Figure 7. Diagram of cytochrome oxidase I gene and flanking tRNA, showing position of PCR and sequencing primers. Shading indicates 400 bp region sequenced. Numbering is according to the *Drosophila yakuba* mitochondrial genome (Clary & Wolstenholme, 1985).

PCR amplification and sequencing

Primer sequences were obtained from the Richard Harrison Laboratory at Cornell University, and numbered according to the *Drosophila yakuba* mtDNA sequence (Clary & Wolstenholme, 1985). 'S' and 'A' refer to sense and antisense strands, and numbers refer to the position of the 3′ end. Primer sequences are given in Table 3.

Table 3. Primers used in amplification and sequencing of COI. Primers numbered from 3′ end, according to *Drosophila yakuba* sequence (Clary & Wolstenholme, 1985).

Primer	Sequence
Sense strand	
S1859	GGA ACI GGA TGA ACT GTT TAC CCI CC (I = Inosine)
S2183	CAA CAT TTA TTT TGA TTT TTT GG
S2441	CCT ACA GGA ATT AAA GTT TTT AGA TGA TTA GC
Antisense strand	
A2507	CCT GTT AAT CCT CCA ACT GTA AAT A
A3014	TCC AAT GCA CTA ATC TGC CAT ATT A

Double-stranded PCR amplification was carried out in a Perkin Elmer Cetus 480 thermocycler using a 5 min denaturation at 95°C followed by thirty-five cycles of 1 min at 50°C, 1 min 30 s at 72°C, 30 s at 95°C and a final 10 min extension at 72°C. Reactions were 100 μl in volume, containing buffer at 4 mM Mg^{++} concentration, 200 μM dNTPs, 150 nM of each primer, 2 μg/ml BSA, and 1 unit of Taq polymerase (Promega), overlayed with mineral oil.

Samples were purified on Centricon-30 columns (Amicon). Single-stranded reamplification using a single primer (225 nM) was carried out using 25 PCR cycles with annealing and extension times reduced by 30 s. Single-stranded DNA was also purified through Centricon-30 columns. Sequencing of single-stranded products was carried out using Sequenase version 2.0 sequencing kits (USB), using standard protocols.

Double-stranded products were obtained using primer pairs S1859/A3014 or S2183/A3014. Single-stranded products were amplified with primer A3014, and sequenced with primers S2183 and S2441 (Fig. 7).

Data analysis

Sequences were aligned using the Lasergene software by DNASTAR. Distance matrices based on nucleotide sequence were produced with the DNADIST program with the PHYLIP package (Felsenstein, 1989), using the maximum likelihood method. Different categories of substitution rate were incorporated into this program, using the 'Categories' option of DNADIST, which allowed for different values for first, second and third codon positions. In the more conserved regions of the sequence, values of 2, 1 and 10 were given for these positions, respectively, reflecting the higher substitution levels at first and third codon position. The sequences corresponding to the matrix side of the membrane were considered to be approximately twice as variable as the conserved regions, and so values for substitution rates were doubled, reflecting the higher substitution rates throughout these regions. Trees were constructed by the neighbour-joining method (Saitou & Nei, 1987), using the NEIGHBOR program of PHYLIP.

Acknowledgements

We thank E. Petit-Pierre, C. Juan, C. Ferris and M. Majerus for supplying specimens, James Brown for his help with the phylogenetic analysis, and Peter Hammond for his helpful comments on the taxonomy of the Coleoptera. This work was funded by the Natural Environment Research Council (Grant GST/02/501).

References

Avise, J.C., Arnold, J. and Ball, R.M. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* **18**: 489–522.

Brower, A.V.Z. (1994) Phylogeny of *Heliconius* butterflies inferred from mitochondrial DNA sequences (Lepidoptera: Nymphalidae). *Molec Phylogen Evol* **3**: 159–174.

Brown, J.K.M. (1994) Bootstrap hypothesis tests for evolutionary trees and other dendrograms. *Proc Nat Acad Sci USA* **91**: 12293–12297.

Brown, J.M., Pellmyr, O., Thompson, J.N. and Harrison, R.G. (1994) Phylogeny of *Greya* (Lepidoptera: Prodoxidae), based on nucleotide sequence variation in mitochondrial cytochrome oxidase I and II: congruence with morphological data. *Molec Biol Evol* **11**: 128–141.

Brown, W.M. (1985) The mitochondrial genome of animals. *Molecular Evolutionary Genetics* (MacIntyre, R.J., ed.), pp. 95–130. Plenum Press, New York.

Calhoun, M.W., Thomas, J.W. and Gennis, R.B. (1994) The cytochrome-oxidase superfamily of redox-driven proton pumps. *Trends Biochem Sci* **19**: 325–331.

Carmean, D., Kimsey, L.S. and Berbee, M.L. (1992) 18S rDNA sequences and the Holometabolous insects. *Molec Phylogen Evol* **1**: 270–278.

Chase, M.W., Soltis, D.E., *et al.* (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Ann Missouri Bot Gard* **80**: 528–580.

Clary, D.O. and Wolstenholme, D.R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J Molec Evol* **22**: 252–271.

Crowson, R.A. (1981) *The Biology of the Coleoptera.* Academic Press, London.

Crozier, R.H., Crozier, Y.C. and Mackinlay, A.G. (1989) The COI and COII region of honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates. *Molec Biol Evol* **6**: 399–411.

Crozier, R.H. and Crozier, Y.C. (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133**: 97–117.

Edwards, S.V., Arctander, P. and Wilson, A.C. (1991) Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc Roy Soc* B **243**: 99–107.

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

Felsenstein, J. (1989) PHYLIP: Phylogeny Inference Package (Version 3.2). *Cladistics* **5**: 164–166.

Irwin, D.M., Kocher, T.D. and Wilson, A.C. (1991) Evolution of the cytochrome B gene of mammals. *J Molec Evol* **32**: 128–144.

Labandeira, C.C. and Sepkoski, J.J., Jr (1993) Insect diversity and the fossil record. *Science* **261**: 310–315.

Liu, H. and Beckenbach, A.T. (1992) Evolution of the mitochondrial cytochrome oxidase II gene among ten orders of insects. *Molec Phylogenet Evol* **1**: 41–52.

Moritz, C., Dowling, T.E. and Brown, W.M. (1987) Evolution of

animal mitochondrial DNA: relevance for population biology and systematics. *Annu Rev Ecol Syst* **18**: 269–292.

Nei, M. (1991) Relative efficiencies of different tree making methods for molecular data. *Phylogenetic Analysis of DNA Sequences* (Miyamoto, M.M. and Cracraft, J., eds), pp. 90–128. Oxford University Press, New York.

Pashley, D.P., McPheron, B.A. and Zimmer, E.A. (1993) Systematics of Holometabolous insect orders based on 18S ribosomal RNA. *Molec Phylogenet Evol* **2**: 132–142.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molec Biol Evol* **4**:406–425.

Saraste, M. (1990) Structural features of cytochrome oxidase. *Quart Rev Biophys* **23**: 331–366.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. and Flook, P. (1994) Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann Ent Soc Amer* **87**: 651–701.

Smith, M.F. and Patton, J.L. (1993) The diversification of South American murid rodents: evidence from mitochondrial DNA sequence data for the akodontine tribe. *Biol J Linn Soc* **50**: 149–177.