Blackwell Publishing

Cladistics 24 (2008) 34–50

Cladistics

10.1111/j.1096-0031.2007.00167.x

Multilocus ribosomal RNA phylogeny of the leaf beetles (Chrysomelidae)

Jesús Gómez-Zurita*†, Toby Hunt and Alfried P. Vogler

Department of Entomology, Natural History Museum, London, UK; Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire, UK

Accepted 26 March 2007

Abstract

Basal relationships in the Chrysomelidae (leaf beetles) were investigated using two nuclear (small and partial large subunits) and mitochondrial (partial large subunit) rRNA (≈ 3000 bp total) for 167 taxa covering most major lineages and relevant outgroups. Separate and combined data analyses were performed under parsimony and model-based tree building algorithms from dynamic (direct optimization) and static (Clustal and BLAST) sequence alignments. The performance of methods differed widely and recovery of well established nodes was erratic, in particular when using single gene partitions, but showed a slight advantage for Bayesian inferences and one of the fast likelihood algorithms (PHYML) over others. Direct optimization greatly gained from simultaneous analysis and provided a valuable hypothesis of chrysomelid relationships. The BLAST-based alignment, which removes poorly aligned sequence segments, in combination with likelihood and Bayesian analyses, resulted in highly defensible trees obtained in much shorter time than direct optimization, and hence is a viable alternative when data sets grow. The main taxonomic findings include the recognition of three major lineages of Chrysomelidae, including a basal "sagrine" clade (Criocerinae, Donaciinae, Bruchinae), which was sister to the "eumolpine" (Spilopyrinae, Eumolpinae, Cryptocephalinae, Cassidinae) plus "chrysomeline" (Chrysomelinae, Galerucinae) clades. The analyses support a broad definition of subfamilies (i.e., merging previously separated subfamilies) in the case of Cassidinae (cassidines + hispines) and Cryptocephalinae (chlamisines + cryptocephalines + clytrines), whereas two subfamilies, Chrysomelinae and Eumolpinae, were paraphyletic. The surprising separation of monocot feeding Cassidinae (associated with the eumolpine clade) from the other major monocot feeding groups in the sagrine clade was well supported. The study highlights the need for thorough taxon sampling, and reveals that morphological data affected by convergence had a great impact when combined with molecular data in previous phylogenetic analyses of Chrysomelidae.

The Chrysomelidae (leaf beetles) includes over 35 000 described species feeding on green parts of plants, while some groups secondarily feed on pollen, flowers, roots, seeds and ant nests debris (Jolivet and Verma, 2002). Despite their great economic importance (Jolivet and Verma, 2002) and use as a model for plant–herbivore coevolution (Ehrlich and Raven, 1964; Mitter and Farrell, 1991; Becerra, 1997), basal relationships in Chrysomelidae are not well understood. The family has been subdivided in up to 16 subfamilies (Seeno and Wilcox, 1982), although the most conservative classification considers 11 subfamilies, lumping several well recognized higher taxa (Reid, 1995). In addition, the chrysomelids are now well established to include the Bruchidae (seed beetles), and the subfamily Spilopyrinae was recently added for a divergent set of species removed from the Eumolpinae (Reid, 1995, 2000; Gómez-Zurita et al., 2005).

Whereas most subfamilies are clearly monophyletic, and some higher groups are easily identifiable, relationships between major lineages have been difficult to resolve, possibly because they originated in short succession (Gómez-Zurita et al., 2007). The most detailed attempts to resolve basal relationships in Chrysomelidae have used a range of morphological characters either

*Corresponding author:
E-mail address: jgzurita@um.es
†Present address: Zoologische Staatssammlung München, Münchhausenstrasse 21, 81247 Munich, Germany.

separately (Reid, 1995, 2000), or combined with 18S rRNA sequences (Farrell, 1998; Duckett et al., 2004; Farrell and Sequeira, 2004). Other studies have used specific character systems (e.g., larval morphology, Lee, 1993; male genitalia, Verma, 1996; reproductive system and hind wing venation, Suzuki, 1996), but their impact has been less significant. To date, no sister-group relationship between two subfamilies is free from criticism (Reid, 2000). Even so, current conclusions on relationships in Chrysomelidae are strongly influenced by morphological characters, and where combined with molecular data, the preferred trees from these analyses were obtained under weighting schemes that increased the contribution of morphology.

We recently obtained sequences for three ribosomal markers, the nuclear 18S rRNA (SSU) and 28S rRNA (LSU), and mitochondrial 16S rRNA (*rrnL*) for a representative sample of Chrysomelidae, with the aim of investigating the evolutionary timeframe of coradiation with their angiosperm host plants (Gómez-Zurita et al., 2007). These data are now available for a detailed phylogenetic analysis to reassess the systematics of this group and the impact of morphological characters on combined analyses. Length variation is of particular concern for phylogenetic analysis using ribosomal RNA genes, and a variety of strategies for analyzing alignment variable markers based on statistical analysis of similarity or explicit inferences of homology can be applied. As data sets grow in size and complexity, analytical approaches have to be tailored to provide fast and reliable tree construction. While not all procedures are equally defensible on theoretical grounds, fast alignment and tree building methods may result in sufficiently high accuracy of trees. Here, we applied a range of procedures to the analysis of basal relationships and classification of the Chrysomelidae, addressing questions about the disputed monophyly of the family and the constitution and relationships among subfamilies.

Materials and methods

Taxon coverage and data used

Taxon sampling included 147 representatives of Chrysomelidae from 134 genera and all subfamilies except Sagrinae and Lamprosomatinae, plus two representatives of Orsodacnidae, one Megalopodidae, 16 Cerambycidae and one Vesperidae; the latter was used as an outgroup in all analyses. All data used here are from Gómez-Zurita et al. (2007) and include sequences for partial mtDNA *rrnL* and nuclear LSU, and complete SSU sequences. The matrices were complete for the SSU and LSU data sets, but *rrnL* sequences were missing for 10 taxa (four chrysomelids, six cerambycids). Sequence data were deposited in the EBI DNA sequence database under accession numbers AJ841299–AJ841670. For combined analyses, a morphological data set comprising 56 morphological characters (Reid, 2000) was obtained from the original paper. Character states in the morphological matrix were provided for each subfamily as a single terminal, but separating two tribes of Chrysomelinae (Chrysomelini and Timarchini) and three tribes of Eumolpinae (Synetini, Eumolpini and Megascelidini).

Phylogenetic analyses

Homology assignment of nucleotides is a key step in phylogenetic analyses of length-variable RNA markers. Three principally different approaches were applied, including fixed and dynamic homology searches. The latter was implemented using direct optimization (Wheeler, 1996; Ogden et al., 2005) to find the most parsimonious tree by optimizing nucleotide changes and indels in a one-step approach, with the cost of substitutions (transitions and transversions) and insertion–deletions (indels) specified by a step matrix. Optimal trees are obtained by rearrangements to the tree topology and correspondences of nucleotide positions, to minimize substitutions and length variation simultaneously. The output of the analysis are (1) the shortest tree topology (defined by the cost matrix), and (2) the so-called implied alignment, which is a visual display of the assigned homologies and is derived from the tree secondarily by tracing back the original character optimizations through the cladogram (Wheeler, 2003). Direct optimization was performed in POY 3.0.11 (Wheeler, 1996; Wheeler et al., 2002) on a parallel processing system using a 16 dual-processor (2.8GHz P4, 2GB RAM) cluster at Imperial College London for a maximum of 48 h for each run. Tree searches included three consecutive stages, each computationally more intensive than the previous (Giannini and Simmons, 2003; see Appendix 1). The first step consisted of 40 random sequence addition replicates keeping the optimal trees from each independent replicate (*-repintermediate*), followed by up to 10 000 tree fusings (Goloboff, 1999). The second step consisted of several TBR ratchet rounds (Nixon, 1999) performed on the shortest tree from the previous tree fusing and on the shortest and longest tree, respectively, obtained from each random addition replicate. Finally, the shortest tree from these analyses was submitted to a TBR search under iterative pass optimization (Wheeler, 2003). The latter was time consuming but resulted in a significant reduction of the tree length even when searches were not run to completion. All tree searches were done under a scheme of equal costs for nucleotide changes and indels. An identical search strategy was followed using a combined matrix of molecular and Reid's (2000) morphological data set, whereby the same character states were applied

to each taxon classified as members of a given subfamily. Analyses of character evolution were performed with MacClade 4.07 (Maddison and Maddison, 2005).

Node robustness was assessed using a heuristic Bremer support search in POY (command *-bremer*) constraining the topology from the iterative pass search to estimate the decay values. Searches consisted of 30 random sequence addition replicates (three tree builds with 10 rounds of random sequence addition) with several rounds of SPR and TBR branch rearrangements, and tree fusing and drifting (Appendix 1). Because this search was less efficient than the original parsimony search, greatly inflating the decay values, we discarded it and estimated support by bootstrapping the associated implied alignment in PAUP*, although here support values were also inflated because uncertainty from homology assignment is not considered. Finally, we established bootstrap proportions on the matrix excluding all characters with alignment gaps, representing characters less affected by alignment ambiguities (heuristic search with three random sequence addition replicates and 2.5×10^8 TBR rearrangements on 100 bootstrap pseudoreplicates).

In addition to direct optimization, static alignment procedures using two-step analyses (separate alignment and tree searches) were applied. First, the implied alignment generated by POY can be used as a primary alignment on which to perform further tree searches (Wheeler, 2003). We also used a "progressive" alignment procedure based on penalties in pair-wise alignment as implemented in ClustalW (Thompson et al., 1994). Finally, we employed an alignment based on the *blastn* algorithm, which identifies short non-gapped segments of high similarity between pairs of sequences as implemented in BlastAlign (Belshaw and Katzourakis, 2005). These High-scoring Segment Pairs (HSP) act as seeds for initiating searches to find longer segments in both directions and can be displayed as "flat query-anchored alignments" that contain mainly the alignment-conservative regions of the sequences, improving homology assignments. The resulting alignments were used as the primary alignments for tree searches. Default settings of ClustalW (version 1.86) and BlastAlign were used in each case.

Each alignment was the basis for tree searches using four different algorithms and optimality criteria: (1) parsimony in PAUP* 4.0b10 for Unix (Swofford, 2003); (2) maximum likelihood using PHYML 2.4 (Guindon and Gascuel, 2003); (3) maximum likelihood searches based on the genetic algorithm implemented in MetaPIGA 1.0.2b (Lemmon and Milinkovitch, 2002); and (4) Bayesian reconstruction of phylogenetic relationships as implemented in MrBayes 3.0b4 (Huelsenbeck and Ronquist, 2001).

PAUP* searches consisted of 50 replicates of random sequence addition with TBR branch-swapping and saving multiple trees per replicate. PHYML likelihood searches were run under the evolutionary models and estimated parameters for each matrix as obtained from ModelTest 3.06 (Posada and Crandall, 1998) and starting from a tree obtained using the modified neighbor-joining algorithm BIONJ (Gascuel, 1997). Node robustness was assessed using non-parametric bootstrapping by 100 bootstrap pseudoreplicates. "Genetic" algorithms for ML searches were implemented in MetaGA (Lemmon and Milinkovitch, 2002), performing searches with probability consensus pruning among four populations and applying the HKY + Γ + I model (Hasegawa et al., 1985; the most complex model available to MetaPIGA), with the Ti/Tv ratio estimated from the data. Searches started from neighbor joining trees, and a single best tree per population was kept. Twenty-five replicates were run (200 for the three ribosomal markers combined) and the resulting trees were used to compute a majority-rule consensus tree. Bayesian tree reconstructions were done under the appropriate model of nucleotide substitution and running four chain searches for 10^6 generations. Chains were sampled every 100th trees and initial trees obtained before reaching the stationary phase were removed (burn-in). The estimated tree topology, branch lengths and the Bayesian posterior probabilities for each node were saved. In all non-parsimony analyses, gaps were treated as missing data.

Trees were assessed for the recovery of monophyletic groups considered to constitute well established clades in the traditional systematics of the Chrysomeloidea, similar to the "key nodes" in Ribera et al. (2002) to evaluate stability under different analytical conditions. This taxonomic congruence approach exclusively relied on well established groups whose validity is little controversial. Critical nodes for this analysis were (1) the monophyly of each subfamily of Chrysomelidae and Cerambycidae and the well established chrysomeline tribe Timarchini (sometimes assigned subfamily status); (2) the monophyly of groups of subfamilies that have been established with high confidence on morphological grounds, including "Camptosoma" (Cryptocephalinae + Clytrinae + Chlamisinae = Cryptocephalinae sensu Reid, 1995), "Cryptostoma" (Cassidinae + Hispinae = Hispinae sensu Reid, 1995) and "Trichostoma" (Galerucinae + Alticinae = Galerucinae sensu Reid, 1995); and (3) the monophyly of Chrysomelidae, Orsodacnidae and Cerambycidae.

Specific hypotheses of monophyly were further tested using a maximum likelihood framework and the Shimodaira–Hasegawa test (SH test; Shimodaira and Hasegawa, 1999) in PAUP*. Parsimony trees obtained with and without topological constraint were evaluated using the SH test implementing the most appropriate evolutionary model (according to the hierarchical likelihood ratio test in ModelTest) and resampling of

estimated log-likelihoods using 1000 bootstrap replicates.

Results

Aligned matrices and parsimony analysis

Among the three types of alignment performed, the implied alignment required the largest number of aligned positions, more than twice the length of the longest sequence in the case of *rrnL* (Table 1). This reflects the fact that the procedure generates additional alignment columns when nucleotides that correspond to each other on the tree are displayed as linearly arrayed alignments. Implied alignments for the two other markers showed only an 11.7% increase of matrix size over the longest sequence for SSU, and 57.5% for the faster evolving LSU (Table 1). Clustal and BlastAlign alignments were more compressed (Table 1). The number of variable sites was highest for *rrnL*, followed by the less variable LSU and particularly the SSU rRNA genes, although the latter provided a large number of potentially informative indel characters (Table 1).

Standard parsimony analyses on these aligned data matrices under equal weighting showed a low CI of 0.12 for *rrnL* (RI = 0.47) on the Clustal alignment (554 positions), increasing to 0.14 when gaps were treated as a fifth character state (RI = 0.54), and much higher CIs and RIs for the nuclear genes (Table 1). Homoplasy levels in BlastAlign alignments were similar, e.g., for the *rrnL* alignment comprising 617 positions we found a CI = 0.16 (0.17 including gaps) and RI = 0.45 (0.59). The implied alignment from direct optimization was much increased in size (e.g., 1089 positions for *rrnL*) and also resulted in broadly similar homoplasy levels, but homoplasy was lower compared with the other methods when gaps were treated as a fifth character state (Table 1). Apparently, the compact alignment achieved with Clustal and BlastAlign was at the expense of higher homoplasy of gap characters, while the stretched alignments improved the congruence of aligned positions. However, when calculated as the proportion of the maximum possible homoplasy (RI), the values from all alignments were very similar, indicating a similar level of synapomorphic character information.

The aligned matrices were then subjected to various types of tree searches, using each of the three markers separately and in simultaneous analysis. We first tested the recovery of subfamilies and well established groups of subfamilies, for a total of 22 groups (Fig. 1, Table 2). By this criterion, mitochondrial *rrnL* recovered the fewest groups and contradicted at least half of them in any type of analysis, while LSU performed better and SSU rRNA was best (Table 2). Both the alignment strategy and the method of tree reconstruction had an impact on the trees obtained. Parsimony analysis in particular performed poorly on the single partition data sets, and also on the BLAST-based alignment, indicating that the reduced amount of data affected this type of analysis more so than the model-based analyses. Among the two likelihood approaches used, the PHYML procedures performed generally better than MetaPIGA, while the best performance by this criterion was by Bayesian analysis.

Any of the three markers separately performed worse than the simultaneous analysis. Again, the Bayesian analysis performed best on the combined data, and retrieved essentially the same 16 key nodes regardless of the alignment strategy used (PHYML on the BLAST alignment also retrieved almost the same 16 nodes; Fig. 1, Table 2). Direct optimization gained most from

Table 1
Size and composition of aligned data matrices, pair-wise divergences (both range and average ± SD are given for Chrysomelidae and Cerambycidae) and basic tree statistics for the three ribosomal markers obtained with various alignment procedures

	Aligned length	Var. sites/ informative*	$[d_{Chr}]$†	d_{Chr}†	$[d_{Cer}]$†	d_{Cer}†	CI*	RI*
Clustal								
rrnL	554	403 (422)/362 (377)	0.010–0.402	0.240 ± 0.012	0.103–0.251	0.176 ± 0.012	0.12 (0.14)	0.47 (0.54)
SSU rRNA	1942	489 (473)/283 (341)	0–0.052	0.021 ± 0.001	0.002–0.019	0.011 ± 0.001	0.37 (0.38)	0.74 (0.79)
LSU rRNA	756	241 (273)/171 (203)	0–0.091	0.037 ± 0.004	0.006–0.094	0.033 ± 0.004	0.31 (0.31)	0.69 (0.76)
BLAST								
rrnL	617	337 (598)/295 (483)	0.012–0.266	0.153 ± 0.010	0.083–0.175	0.130 ± 0.011	0.16 (0.17)	0.45 (0.59)
SSU rRNA	1946	368 (547)/243 (300)	0–0.041	0.018 ± 0.001	0.002–0.015	0.010 ± 0.001	0.36 (0.40)	0.72 (0.72)
LSU rRNA	741	185 (240)/122 (148)	0–0.075	0.030 ± 0.003	0.003–0.039	0.019 ± 0.003	0.32 (0.36)	0.71 (0.71)
IA								
rrnL	1089	437 (943)/365 (503)	0–0.286	0.137 ± 0.017	0.078–0.239	0.151 ± 0.011	0.15 (0.23)‡	0.48 (0.55)‡
SSU rRNA	2127	382 (639)/250 (356)	0–0.034	0.014 ± 0.001	0.002–0.013	0.007 ± 0.001	0.35 (0.45)‡	0.71 (0.76)‡
LSU rRNA	1120	187 (628)/125 (221)	0–0.070	0.027 ± 0.004	0.003–0.042	0.022 ± 0.004	0.27 (0.50)‡	0.70 (0.74)‡

*The value considering gaps as a fifth character state in parentheses.
†Divergences calculated applying the K2P correction.
‡Measured from the tree obtained by POY.

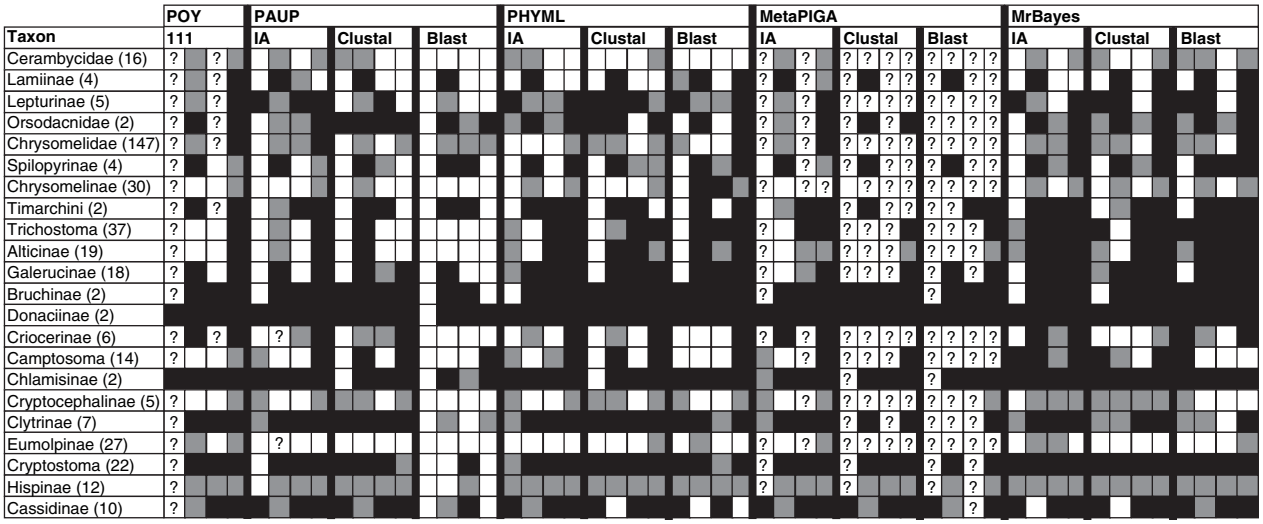

Fig. 1. Graphical summary of the results of different types of phylogenetic analyses. The five types of phylogenetic analyses implemented were direct optimization (POY), parsimony (PAUP), maximum likelihood (PHYML, MetaPIGA) and Bayesian (MrBayes), each applied to three different alignments, including an implied alignment from the direct optimization analysis (IA), a ClustalW alignment (Clustal) and a BlastAlign alignment (BLAST). Each analysis was carried out for *rrnL*, SSU, LSU and their combination, corresponding to the four columns in each section of the figure. A filled black cell in the matrix indicates recovery of the relevant group as monophyletic, a gray cell as paraphyletic, an empty cell as polyphyletic, and a question mark as an inconclusive result due to unresolved polytomies. A numerical summary of these results is also presented in Table 2.

Table 2
Recovery of "key nodes" in different types of phylogenetic analyses. Each column shows the number of monophyletic and paraphyletic, respectively, groups recovered from a set of 22 well-established groups in the Chrysomeloidea

Alignment	Tree search	rrnL	SSU	LSU	all
IA	PAUP	4/3	8/9	8/5	15/5
	PHYML	4/10	10/6	10/4	15/4
	MetaPIGA	2/4	8/7	8/3	14/7
	MrBayes	6/4	14/7	10/8	16/4
Clustal	PAUP	6/3	13/8	9/4	9/4
	PHYML	7/3	13/5	11/2	11/9
	MetaPIGA	3/0	9/2	5/1	10/2
	MrBayes	7/8	9/7	10/4	16/5
BLAST	PAUP	0/0	8/4	6/4	4/2
	PHYML	7/4	13/3	7/6	16/3
	MetaPIGA	2/0	7/2	4/0	9/3
	MrBayes	9/7	12/7	10/2	16/4
DO	POY111	2/0	10/7	6/1	15/7

combining the data, presumably because synergistic effects combine to reveal the shared historical signal (Gatesy et al., 1999) and alignment is optimized accordingly. The equally weighted POY tree as well as the implied alignment used for tree building with any of the available methods recovered from 19 to 22 (the maximum possible) higher-level taxa as monophyletic (14–16) or narrowly paraphyletic (Fig. 1, Table 2). All other alignment–tree building method combinations (except direct optimization, see below) showed lower key node recovery, in particular with parsimony and MetaPIGA searches (Fig. 1, Table 2).

Not a single higher-level taxonomic group was stable across all analyses, although some groups such as Bruchinae, Donaciinae, Chlamisinae (each represented by only two exemplars), Clytrinae (seven terminals) and Cryptostoma (22 terminals) appeared monophyletic in most analyses. To a lesser degree, Orsodacnidae, Galerucinae, Cassidinae and Spilopyrinae were recovered in most analyses, and Hispinae, Cryptocephalinae and the family Chrysomelidae were frequently paraphyletic (Fig. 1).

Focusing on the tree topology of groups beyond the key nodes, the tree from direct optimization was selected to represent parsimony analysis (Fig. 2). This tree subdivided the monophyletic Chrysomelidae s. str. in three major clades, corresponding to: (1) a "chrysomeline" clade, including Chrysomelinae (paraphyletic with Timarchini at the base) plus the reciprocally monophyletic Alticinae and Galerucinae (= Trichostoma) nested within it; (2) a "eumolpine" clade, including Hispinae (paraphyletic) plus Cassidinae (= Cryptostoma), Clytrinae plus Chlamisinae within a paraphyletic Cryptocephalinae (= Camptosoma), nested within a paraphyletic Eumolpinae (incl. Synetinae), with Spilopyrinae at the base; and (3) a "sagrine" clade, including Criocerinae as sister to Donaciinae plus Bruchinae, the latter presumably including the Sagrinae not sampled here (see below). The sister taxon to Chrysomelidae were Orsodacnidae, considered by some authors as primitive chrysomelids, but raised to family status recently (Kuschel and May, 1990). The megalopodid *Zeugophora varians* was placed outside of the

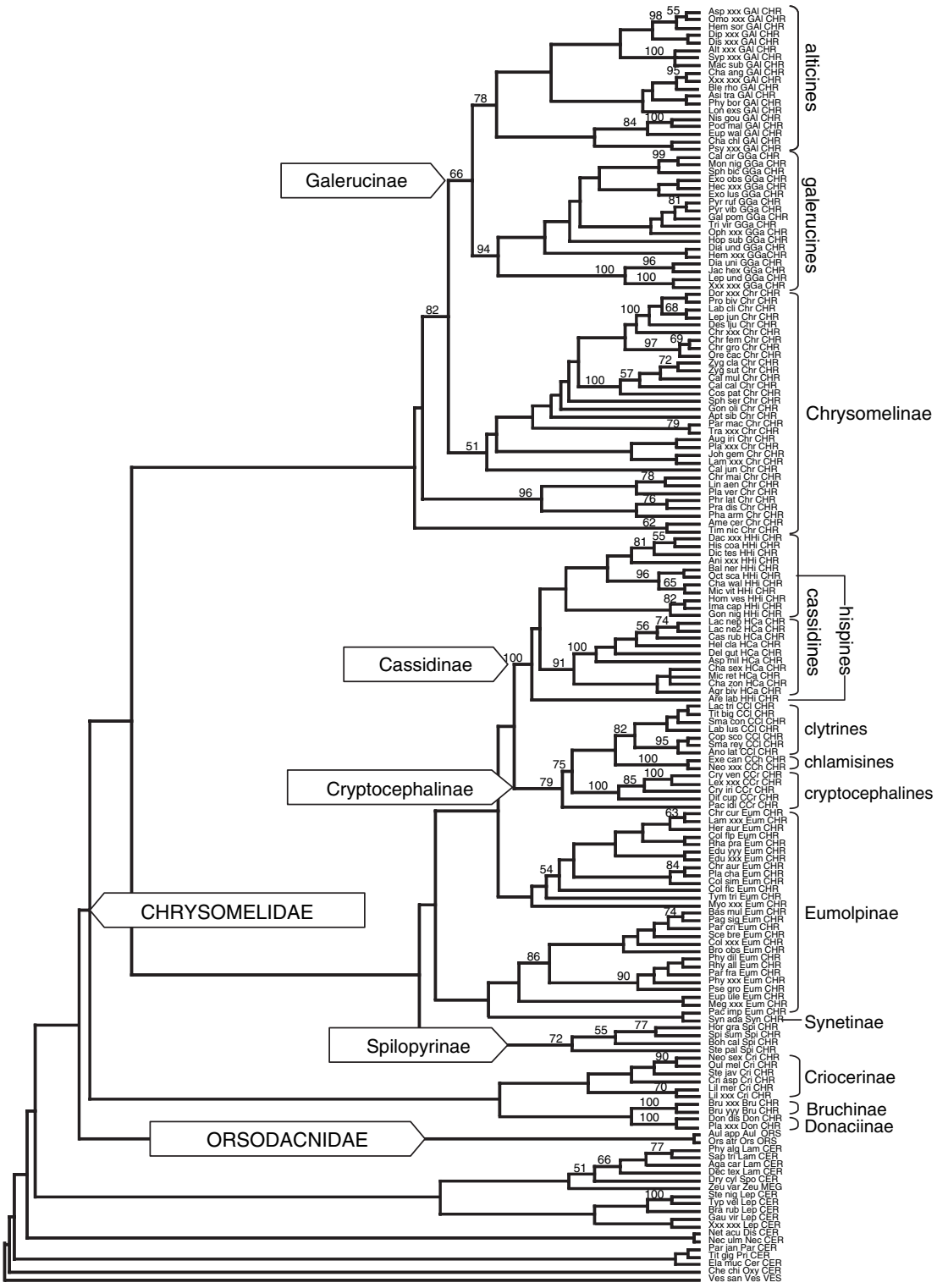

Fig. 2. Most parsimonious phylogenetic hypothesis for the Chrysomelidae based on *rrnL*, SSU and LSU genes from direct optimization analysis under equal weighting (10 105 steps). Numbers above branches represent bootstrap support values above 50% using a matrix excluding all gapped positions. Clades and relevant taxonomic groups are indicated with brackets or arrows pointing to the corresponding nodes.

Chrysomelidae, as expected (Crowson, 1981; Kuschel and May, 1990). Basal relationships were supported only weakly under the most conservative estimates based on bootstrapping of ungapped positions, although their support was high for all nodes if gapped sites in the implied alignment are included (not shown). Support increased for nodes near the tips and the monophyly of most subfamilies received high support, except for Criocerinae and the paraphyletic Eumolpinae and Chrysomelinae (Fig. 2).

Maximum likelihood and Bayesian trees, and comparisons of methods

The likelihood scores of trees from various model-based methods and separate and simultaneous analyses of the three markers were compared (Table 3). As in the analysis of taxonomic congruence above, both the effects of alignment and methods of tree construction affected these scores. Generally, tree searches performed on the implied alignments provided better trees compared with those on Clustal alignments. (The BLAST-based alignments were not strictly comparable for this purpose because the corresponding matrices contained a reduced number of characters.) When comparing the performance of the three model-based methods on a given alignment, MetaPIGA consistently produced worse likelihood scores, although it should be noted that a less complex model was used (HKY versus GTR in the other programs). The Bayesian and PHYML analyses using the same model produced broadly similar scores, although with somewhat unpredictable results as to which performed better (Table 3).

For the combined data set on the BLAST-based alignment, PHYML found a tree with logL = −35314.65 showing high bootstrap proportions for shallow nodes, but lower support near the base of the tree (Fig. 3). A Bayesian analysis under the same model generated a best tree with logL = −34698.07 (arithmetic mean) or −34727.24 (harmonic mean) (Fig. 4). The posterior probabilities were generally high, even at basal nodes and those nodes defining relationships among subfamilies of Chrysomelidae. Differences compared with the topology of the direct optimization were mainly regarding the eumolpine clade where (1) Cassidinae appeared polyphyletic with respect to two subclades of Hispinae (only in ML, not Bayesian analysis); (2) Chlamisinae were sister to the other Camptosoma, not just to Clytrinae; (3) *Syneta* did not appear nested within Eumolpinae, but moved to the base of the sagrine clade; (4) the paraphyly of Eumolpinae was broadened as the primitive eumolpine *Eupales ulema* was placed basal to the Spilopyrinae; (5) Donaciinae was sister to Criocerinae and not to Bruchinae (in Bayesian analysis only); and (6) Orsodacnidae were not recovered as the sister to Chrysomelidae s. str. but nested within the Cerambycidae/Vesperidae.

Inclusion of morphological data

When the morphological characters of Reid (1995, 2000) were added here under direct optimization, the resulting tree topology was identical to that obtained with molecular data only, except for a monophyletic Eumolpinae (with Synetinae subordinated), and the Cassidinae/Hispinae + Cryptocephalinae s.l. as its sister group. The simultaneous analysis added 295 steps to the tree from the molecular data set, increased homoplasy slightly (CI = 0.197, versus CI = 0.203 in separate analysis), and produced moderate incongruence (ILD = 124 steps; WILD = 0.012). This indicated only mild conflict between the two data sets, but nonetheless resulted in some critical differences in the way key

Table 3
Comparison of parsimony tree lengths and ML scores between tree search and alignment methods

Tree search	Alignment	rrnL	SSU	LSU	all
PAUP*	IA	6969 (8)	1385 (17000)	1592 (701760)	8209 (1890)
	Clustal	7387 (231)	2110 (294730)	1711 (10692)	12948 (13)
	BLAST	5582 (187)	1994 (5000)	1296 (20000)	12462 (166)
PHYML (GTR + Γ + I)	IA	−24801.94	−11512.67	−6751.94	−46071.15
	Clustal	−26937.26	−12317.57	−7199.74	−47035.15
	BLAST	−16211.72	−11040.73	−5428.05	−35301.17
MetaPIGA (HKY + Γ + I)	IA	−29063.22	−13273.57	−8099.98	−60127.84
	Clustal	−32521.02	−14860.98	−9091.98	−60748.08
	BLAST	−19626.62	−12966.94	−6637.31	−46127.43
MrBayes† (GTR + Γ + I)	IA	−24773.36	−11723.42	−6904.92	−45562.56
	Clustal	−26355.86	−12557.45	−7395.54	−49210.05
	BLAST	−16043.01	−11285.22	−5606.75	−34698.33

*The number in brackets is the number of trees obtained in a search.

†The likelihood score provided is the arithmetic mean discarding the initial 250 000 trees of the chain (burn-in), consistently lower than the corresponding harmonic mean.

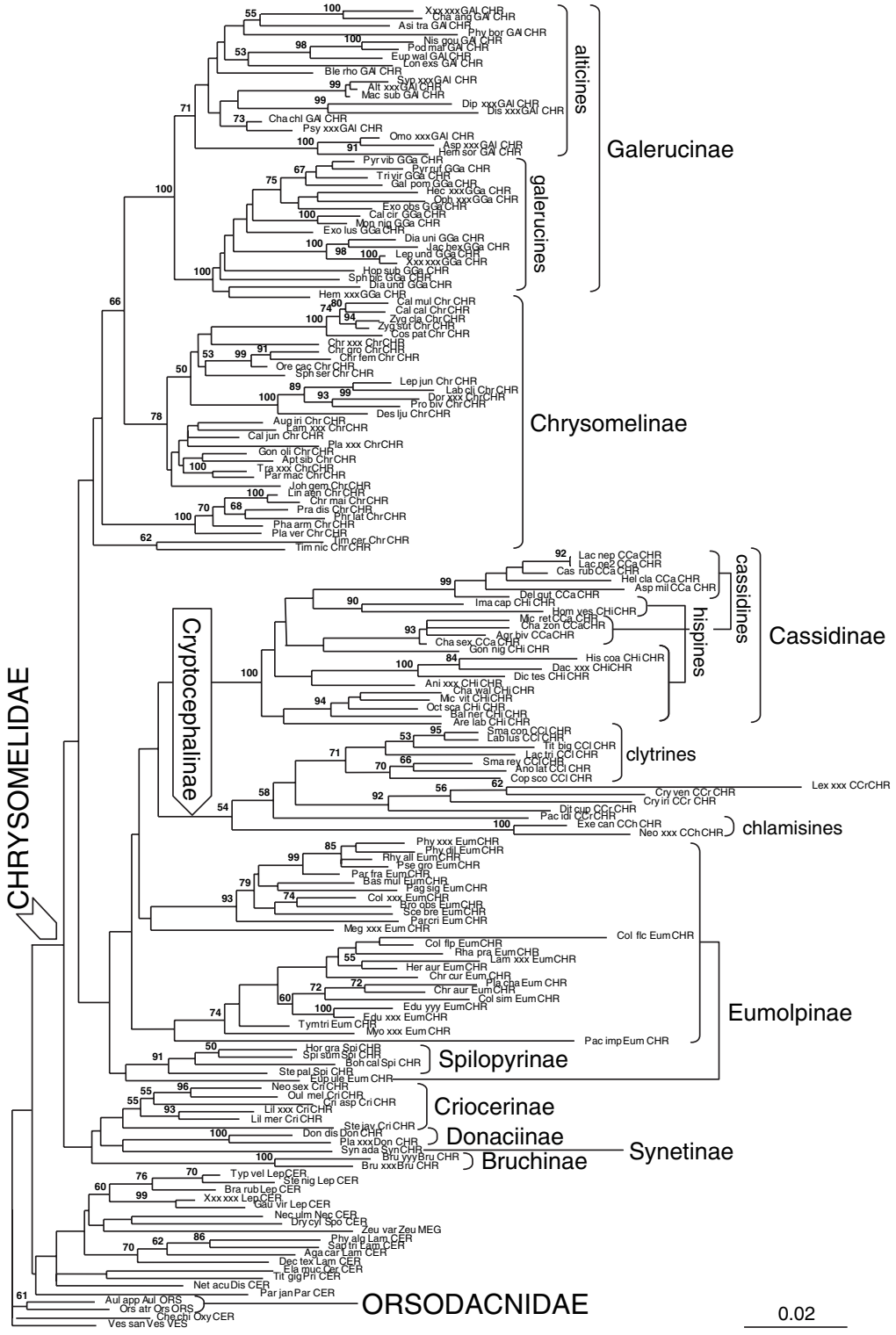

Fig. 3. Maximum likelihood phylogenetic hypothesis based on *rrnL*, SSU and LSU gene sequences. The tree (logL = −35314.65) was obtained under a GTR + Γ + I evolutionary model and sequence alignment was based on the BLASTN algorithm. Numbers above branches are bootstrap support values above 50%.

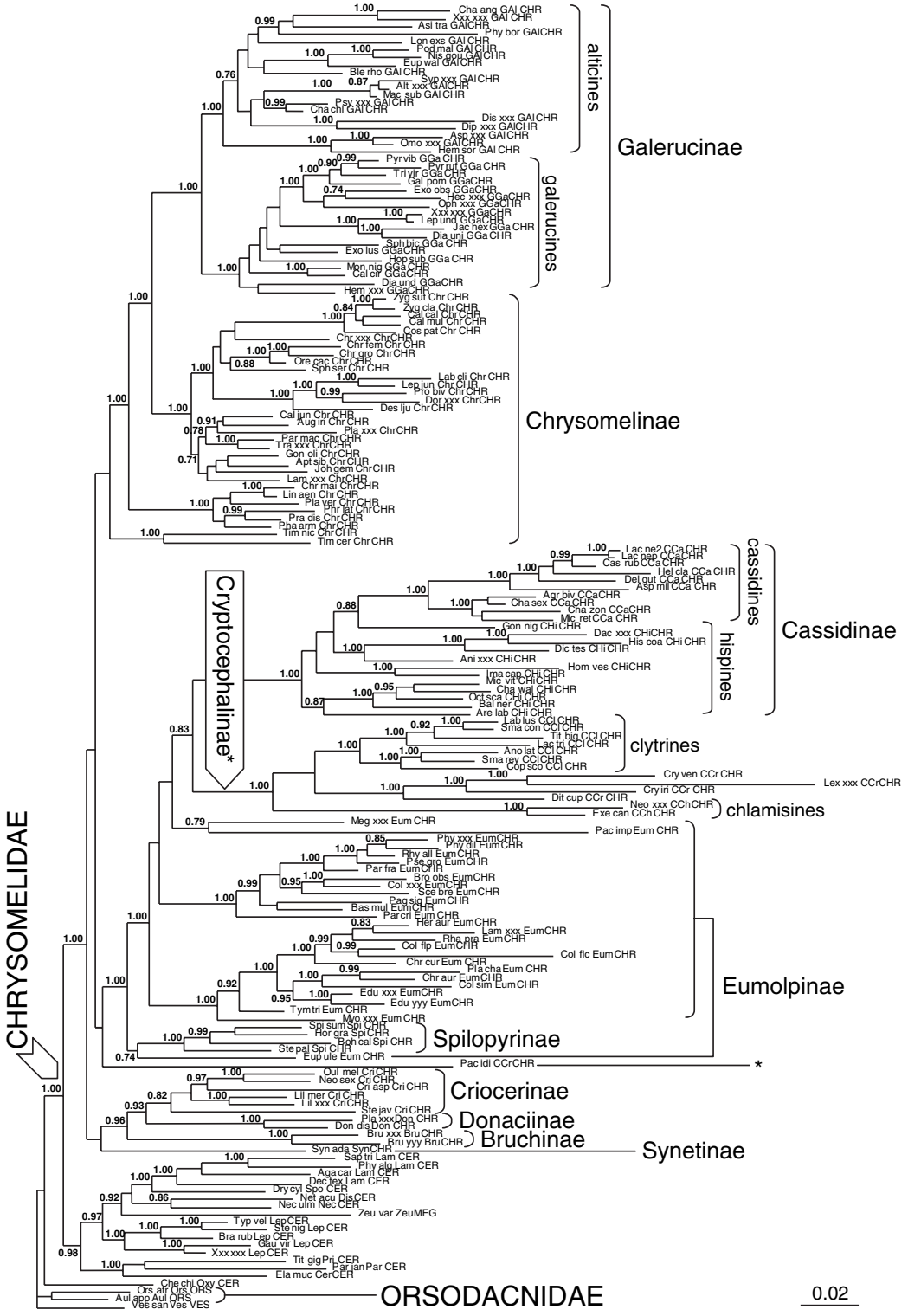

Fig. 4. Bayesian phylogenetic hypothesis based on *rrnL*, SSU and LSU obtained under a GTR + Γ + I evolutionary model and sequence alignment based on the BLASTN algorithm. Numbers above branches are posterior probabilities above 0.70.

morphological character were optimized on the tree. For example, the occurrence of bifid tarsal setae in Cassidinae/Hispinae on the one hand and Donaciinae on the other, is a convergent trait (CI = 0.50; two changes in the tree) in the two monocot feeding lineages. The only other character linking these two lineages, the shared

absence of larval thoracic eggbursters, would also have to be explained by an independent loss of the trait in both lineages (CI = 0.333; three changes in the tree).

Finally, a combined analysis with the inclusion of Sagrinae and Lamprosomatinae, two subfamilies for which molecular markers were unavailable, produced three most parsimonious trees of 10 414 steps in POY analysis. As expected, this placed Sagrinae as sister to Bruchinae in a clade also including Donaciinae. The Lamprosomatinae was confirmed as part of the Camptosoma to which it had been associated before, and was found to be sister to Cryptocephalinae s.l. There were small rearrangements in the tree topology compared with that in the combined analysis without these taxa described above, namely a paraphyletic "sagrine" lineage, with Criocerinae sister to the "eumolpine" and "chrysomeline" clades, and Synetinae sister to the (Donaciinae,(Bruchinae,Sagrinae)) clade.

Discussion

Comparison of phylogenetic methods

Tree alignment procedures and dynamic homology remain highly defensible for phylogenetic analysis of length variable sequences as those used here, because indels and nucleotides can be optimized on the tree simultaneously (Sankoff, 1975; Hein, 1990; Wheeler, 1995, 1996; Fleissner et al., 2005). However, these methods need to set relative weights arbitrarily for indels and nucleotide substitutions, and testing for the most suitable parameters based on taxonomic or character congruence (e.g., Wheeler, 1995). This may be circular because the approach assumes prior knowledge of the trees or character distributions (Simmons and Ochoterena, 2000; Pons and Vogler, 2006). Here we minimized the assumptions about character evolution by assigning equal weight to all changes including indels (Grant and Kluge, 2003). As had been shown for a subset of the data for the Eumolpinae using the same markers and broader exploration of parameters (Gómez-Zurita et al., 2005), direct optimization performed better or at least as well as other methods by the criterion of taxonomic congruence, provided that enough characters were present in the analysis for the emergence of "hidden support" (*sensu* Gatesy et al., 1999) (Table 2).

Separate alignment and tree building two-step approaches here included a primary alignment from short ungapped fragments of similarity obtained with *blastn*, followed by model-based tree construction. The remarkable success of this procedure can be assigned to the unique method of approaching length variable sequences with large indels or regions lacking homology (Belshaw and Katzourakis, 2005). As the algorithm performs pair-wise searches for fragments of sequence identity between any two sequences in the alignment, a "most representative" sequence is selected that has the overall greatest number of identical residues to the set of sequences in the analysis (*Trachymela* sp., in our case). The latter is used to build the "flat query-anchored multiple alignment" of the *blastn* output, keeping only regions with sequence similarity to that sequence. Hence, the procedure provides an objective method for removal of divergent or unalignable regions, which can be performed under varying degrees of stringency (set by the parameters in BLAST, e.g., for the length of the original High-scoring Segment Pairs, or HSPs, and the permitted number of mismatches when extending the HSP). In contrast to established procedures for character exclusion of alignment ambiguous base positions (e.g., "culling"; Gatesy et al., 1993), this procedure removes nucleotides from particular sequences (terminals) rather than eliminating entire columns, hence targeting specifically those regions with no apparent similarity elsewhere and reducing the effects of particular divergent taxa or sequence fragments on the final alignment.

The degree of character exclusion in our data set was correlated with the level of sequence variation: only 1.6% and 3.9% of nucleotide positions were removed for LSU and SSU, respectively, while the more variable *rrnL* data set was reduced by 18.6% mainly from positions around the hypervariable regions. The retention of nucleotide positions was also dependent on the overall composition of the data set, where highly divergent regions may be kept in cases that still provide meaningful homologies with reference to the "most representative" sequence. For example, the "most representative" *Trachymela* sp. grouped in the subfamily Chrysomelinae where data exclusion in *rrnL* affected only 7.7% of nucleotides, compared with 13.6% in the Cerambycidae or 22.3% in Cryptocephalinae. It remains to be tested how data exclusion depends on the clustering of variation in the overall data set, how it increases at the periphery of the space of sequence variation and what are the consequences for phylogenetic inferences of distantly related taxa. Establishing local homology with this procedure therefore is not unlike the hierarchical optimization in POY, but here nucleotides that lack homology with the bulk of sequences (i.e., the set of sequences most similar to the most representative sequence) are removed, rather than used to establish their homology with its closest relatives and included for phylogenetic analyses of subclades. The strength of this approach, not shared by direct optimization, is that portions of questionable homology without matches in other sequences are removed prior to tree construction. As this affects the most heterogeneous portions of the data, problems of long-branch attraction are reduced and the data are more

appropriate for molecular clock estimates (Gómez-Zurita et al., 2007).

Given the great choice of methods for tree construction and alignment, what criteria should be applied when choosing among them? Leaving aside the broader justification for data exploration in phylogenetics (Grant and Kluge, 2003), comparing the performance of methods is inappropriate where different optimization methods are applied. For example, comparing the cost of trees under an optimality criterion of parsimony versus likelihood is immaterial because the searches are not attempting the same goal, and only under very specific conditions of data structure can these searches be expected to favor the same tree (i.e., the parsimony reconstruction approximates the likelihood estimate). This also applies to the direct optimization approach, which has recently been implemented in a likelihood context (although the method is not viable at the scale necessary here). The direct comparison of parsimony and likelihood scores (e.g., Whiting et al., 2006) is only of limited value because different optimality criteria would necessarily result in a different tree, except perhaps under very specific properties of the data.

This leaves the comparison of different implementations of either parsimony or model-based approaches. As observed in other studies, POY-derived alignments were thousands of steps shorter than parsimony trees (e.g., Whiting et al., 2006) based on Clustal alignments, indicating superior optimization of nucleotide homologies. However, this reduction in number of steps did not reduce the level of homoplasy, nor did it increase the proportion of synapomorphic sites as established by the RI (Table 1). This observation suggests that the improvement is largely obtained from the avoidance of nucleotide changes at the expense of greater number of indels. Yet, the overall improved parsimony score (when applying the same weight for indels, as done here) appears defensible as a criterion to choose between different parsimony alignments. Finally, when comparing a range of model-based methods on the various alignments, the likelihood scores from the two best-performing methods, PHYML and MrBayes, varied substantially dependent on the genes these methods were applied to, nuclear LSU and SSU (better PHYML performance) or mitochondrial *rrnL* (MrBayes favored) (Table 3). This leaves questions about how the character variation in different genes might affect the model parameters in either method, and how these differences could be explained by the mechanics of the searches. More importantly, the implied alignments from direct optimization consistently provided better scores than the Clustal equivalent, mirroring the results from parsimony tree lengths. This indicates that likelihood searches on the implied alignment could provide an approximation to likelihood-based optimization of homologies where data sets are too large for the ML implementation of direct optimization.

Beyond epistemological considerations of choosing a tree building method, computing time becomes an important factor as data sets grow. For example, the multistage protocol for direct optimization led to increasingly shorter trees, from 10 234 to 10 283 steps after 40 random sequence addition replicates with TBR branch swapping, to a single tree of 10 215 steps after tree fusing, to trees of 10 213–10 220 after ratcheting, and the final single shortest tree of 10 105 steps under the TBR-iterative pass procedure. This search took more than 112 h (38 h 52 min for random addition + TBR; 25 h 23 min for fusing and ratcheting; 48 h for iterative pass before being timed out) on a 16-dual processor cluster, equivalent to over 3584 h (actual CPU time may have been less due to the lower utilization rate of individual processors). Likelihood searches using MetaPIGA took on average over 35 h on a desktop PC (2.4 GHz, 512 MB RAM), while the Bayesian and parsimony searches took 21 and 9 h, respectively, and PHYML searches completed in only 31 min. Sequence alignments prior to the tree searches added 3 h 4 min to the Clustal alignment and just under 7 min for the BlastAlign alignment on the same desktop PC (the sum of the time needed to align each marker separately). As the solutions from different phylogenetic reconstruction strategies largely converged in the case of this data set, the available very fast procedures based on a BlastAlign alignment become sufficiently reliable, while direct optimization searches will have to be very superficial with increasing numbers of terminals.

Molecular systematics of Chrysomelidae

Monophyly of the Chrysomelidae. The Chrysomelidae has been claimed to be polyphyletic (Monrós, 1960; Chen, 1985; Suzuki, 1988, 1994; Schmitt, 1994) or parallelophyletic (i.e., close polyphyly of groups evolving rapidly from a common ancestral stock showing similarity in traits, which, however, are derived repeatedly due to "hidden propensities" in the ancestor; Mayr and Ashlock, 1991; Jolivet and Verma, 2002). This was based on the realization that depending on the characters analyzed, different systematic arrangements would emerge, or that some groups traditionally considered chrysomelids shared characteristics (internal reproductive systems, hindwing venation or the mesoscutopronotal stridulatory organ) with the Cerambycidae, resulting in their ambiguous placement between both families (e.g., Schmitt, 1994; Suzuki, 1994; Jolivet and Verma, 2002). However, this problem has partly been due to a poor understanding of hierarchical structure at the base of the Chrysomeloidea, which was improved

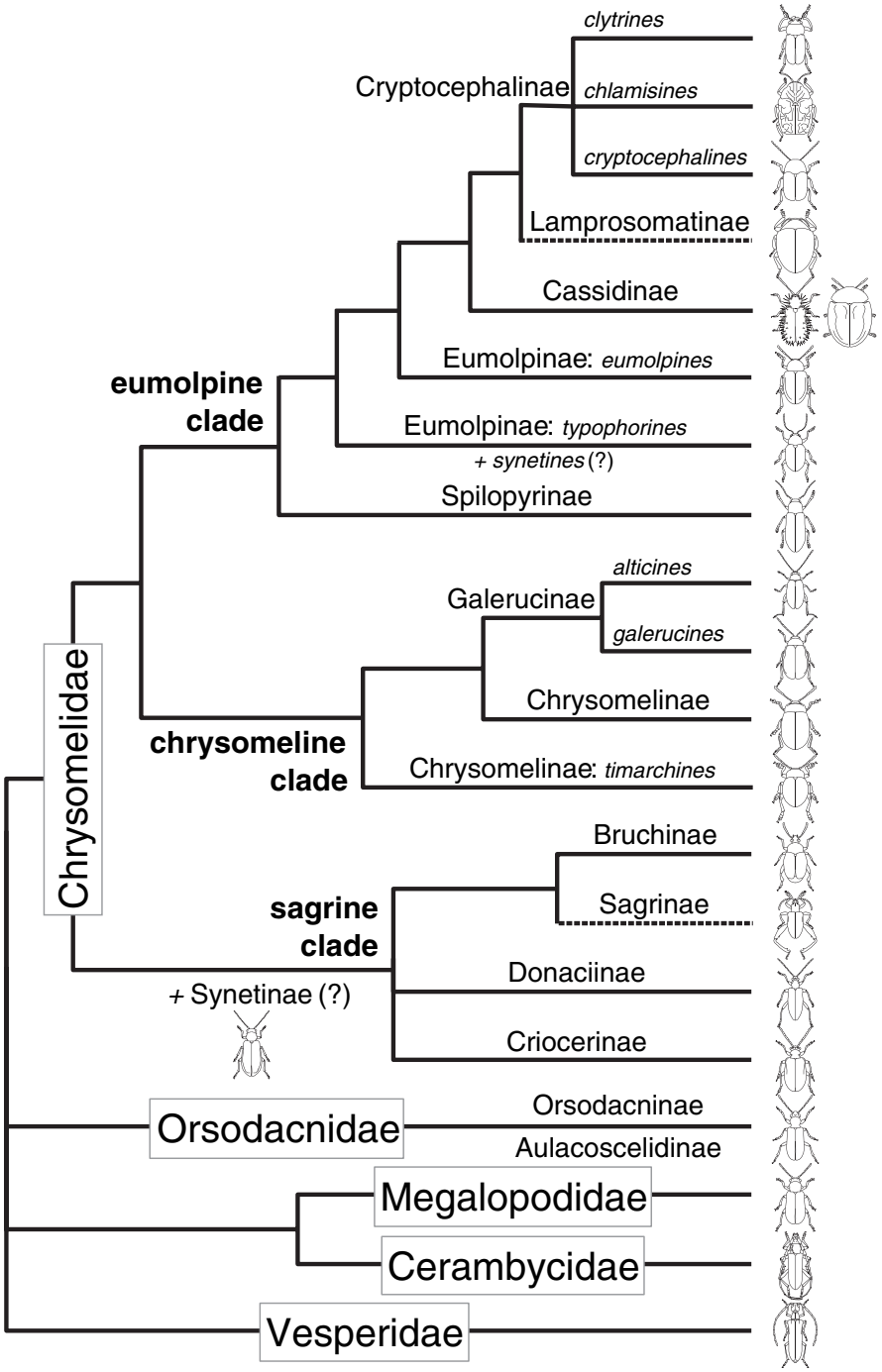

Fig. 5. Summary tree depicting the phylogenetic relationships among leaf beetle subfamilies and outgroups. Three main chrysomelid lineages (the sagrine, eumolpine and chrysomeline clades) are distinguished based on findings of the current study. Some of the traditionally recognized subfamilies were subsumed within well established larger groupings (e.g., alticines plus galerucines = Galerucinae s.l.), while Eumolpinae and Chrysomelinae are shown as paraphyletic and separated into further well supported subgroups that deserve subfamily rank, such as Timarchini. The placement of Synetinae remains ambiguous, as indicated. Lamprosomatinae and Sagrinae were not sampled but were placed confidently according to conclusions from the literature and the combined parsimony analysis with this taxa represented by morphological characters only.

with the recognition of Orsodacnidae and Megalopodidae (Chen, 1985; Suzuki, 1988; Kuschel and May, 1990; Reid, 1995, 2000; Cox and Windsor, 1999; Suzuki, 2003). The Megalopodidae was erected as a family to include the former chrysomelid Megalopodinae, Zeugophorinae and Orsodacninae, to reflect their primitive

traits and the similarity of their larvae to Cerambycidae (Chen, 1985), but later split into orsodacnids and megalopodids + zeugophorids (Suzuki, 1988), remaining as two separate families in modern leaf beetle systematics (e.g., Kuschel and May, 1990; Reid, 1995, 2000). The Megalopodidae also include the recently created Australian Palophaginae (Kuschel and May, 1990), and the Orsodacnidae include the Orsodacninae and more recently the Aulacoscelidinae (Reid, 1995; Cox and Windsor, 1999; Suzuki, 2003). Our results generally support these views of the Chrysomelidae. The megalopodid *Zeugophora varians* appeared basal or within the Cerambycidae, as suggested by characters in the reproductive system and hindwing venation (Suzuki, 1988, 1994). The orsodacnids *Orsodacne atra* (Orsodacninae) and *Aulacoscelis appendiculata* (Aulacoscelidinae) were monophyletic and outside of the Chrysomelidae (confirming Duckett et al., 2004; Farrell and Sequeira, 2004). This supersedes other views that suggested an association of the Orsodacninae to the Galerucinae based on larval and aedeagal characters (Cox, 1981; Crowson and Crowson, 1996) or of the Aulacoscelidinae to either Chrysomelinae or Sagrinae (reviewed in Cox and Windsor, 1999).

Bruchid seed beetles are another group whose relationships with Chrysomelidae have been contentious. Traditionally they were considered a separate family, but their clear affinities to some groups of Chrysomelidae (especially the Sagrinae) led Crowson (1955) and Mann and Crowson (1981) to rank them as a subfamily within the Chrysomelidae. The latter was confirmed here and in previous phylogenetic studies (Reid, 1995; Farrell, 1998; Farrell and Sequeira, 2004), and the debate should now be resolved (Reid, 2000).

The constitution of chrysomelid subfamilies. Three groups in particular have been debated in the past to be paraphyletic, while for two subfamilies paraphyly is reported here for the first time:

1 Clytrinae was established by Crowson (1955) by merging tribes Clytrini, Cryptocephalini, Chlamisini and Lamprosomatini (formerly regarded as subfamilies), a treatment followed by Suzuki (1988) and more recently by Reid (1995), although excluding the Lamprosomatinae and renaming the subfamily as Cryptocephalinae on the basis of priority (the "Camptosomes"; Chapuis, 1874; Jacoby, 1908). This group shares several morphological characters (Erber, 1988; Reid, 1995), and is further characterized by the production of an egg mantle protection at oviposition, which is re-utilized by the larvae and enlarged into a protective case carried until pupation (Erber, 1988). In our study, Cryptocephalinae s.l. was a well supported monophyletic clade, but the relationships among these lineages of leaf beetles were not entirely resolved, while Cryptocephalinae s.str. appeared as paraphyletic or polyphyletic.

2 Hispines (leaf-miners) and cassidines (tortoise beetles)—"Cryptostomes" of Chapuis (1874)—have been treated as a single (e.g., Crowson, 1955; Reid, 1995, 2000) or separate subfamilies (e.g., Seeno and Wilcox, 1982), characterized by mining larvae with lateral thoracic processes and abdominal apical furci in the former, and exophagous larvae without these structures in the latter. Ambiguous placement of some species (Borowiec, 1995) and paraphyly of Hispinae (Farrell, 1998; Hsiao and Windsor, 1999; Duckett et al., 2004; Farrell and Sequeira, 2004) was confirmed here with a more extensive sample, clearly favoring their treatment as a single subfamily, the Cassidinae (Borowiec, 1995; Staines, 2002; = Hispinae *sensu* Reid, 1995, 2000).

3 Galerucinae and Alticinae (flea beetles)—"Galerucides" of Chapuis (1874), "Trichostomes" of Jacoby (1908)—have been treated either as two related subfamilies or a single subfamily with alticines subordinated to galerucines (Böving, 1929; Lingafelter and Konstantinov, 2000, and references therein). The former is supported by two apomorphies in the flea beetles, the inflated hind femora with a specialized internal sclerite (Konstantinov, 1994) and the division of the lateral rubbing patch on the underside of the elytra (Samuelson, 1994). The cladistic analysis of morphological characters (Lingafelter and Konstantinov, 2000) and molecular data alone or combined with Lingafelter and Konstantinov's (2000) data resulted in an alternative hypothesis with the Alticinae paraphyletic including a monophyletic Galerucinae (Kim et al., 2003; Duckett et al., 2004), or reciprocally monophyletic Galerucinae and Alticinae (Farrell, 1998; Farrell and Sequeira, 2004). The latter was supported here. However, key genera (see Furth and Suzuki, 1994) remain to be included before this question can be fully resolved.

4 Chrysomelinae has been a very stable group taxonomically, characterized by many apomorphies of larvae and adults (Chen, 1934). Previous molecular studies did not appreciate the potential paraphyly due to sampling problems. Only two Chrysomelinae from a single clade were sampled initially (Farrell, 1998), while after the inclusion of further exemplars (Duckett et al., 2004; Farrell and Sequeira, 2004) only the addition of morphological data retrieved this lineage as monophyletic. Our sampling of 30 exemplars is a better representation of the full diversity, indicating paraphyly, although trees constrained for the monophyly of Chrysomelinae were not significantly worse in the SH test.

5 Eumolpinae has been considered monophyletic with the problematic inclusion of Synetinae usually subordinated as an early separated tribe (Reid, 2000; Verma and Jolivet, 2002; Gómez-Zurita et al., 2005). Previous studies initially included only a few exemplars of the well-defined tribes Eumolpini and Megascelidini

(Farrell, 1998; Duckett et al., 2004; Farrell and Sequeira, 2004), which were recovered as monophyletic. Represented here by 27 taxa, the monophyly of Eumolpinae was clearly rejected in the SH test. However, when combined with Reid's (2000) morphological characters the Eumolpinae appears monophyletic, including the Synetinae occupying a long branch defined by 15 homoplasious changes and one apomorphy (or with exclusion of Synetinae when the matrix is expanded to include Sagrinae and Lamprosomatinae). This could be an erroneous result of the coding scheme, which, following the practice of previous combined analyses, scored all members of a taxonomic group as having the same character states, favoring the monophyly of subfamilies. The critical finding is that the monocot feeding Cassidinae + Cryptocephalinae s.l. maintain their position closely allied to Eumolpinae and separation from the Donaciinae, supporting the separate origin of the two monocot feeding lineages.

Higher-level systematics of the Chrysomelidae. The results of all molecular analyses converged on the same systematic classification for the Chrysomelidae, differing only in the position of the Synetinae (Fig. 5). Chrysomelidae were separated into three well-differentiated lineages, including:

1 A "sagrine" clade with Bruchinae, Donaciinae, Criocerinae, and putatively Sagrinae (unambiguously associated to the Bruchinae in the parsimony analysis with this taxon represented by morphological characters and previous studies; Monrós, 1960; Crowson, 1981; Mann and Crowson, 1981; Reid, 1995, 2000; Verma, 1996; Farrell and Sequeira, 2004). This clade and its basal placement in the Chrysomelidae have been recognized widely (Farrell, 1998; Duckett et al., 2004; Farrell and Sequeira, 2004), but with some uncertainty about the affinities of Cassidinae and Criocerinae (see Reid, 1995). Our results strongly contradict the suggested relationship of the Cassidinae with the "sagrine" clade (the SH test rejects their monophyly). The former had been grouped among sagrines mostly because they share the unusual bifid tarsal setae as opposed to simple setae in the other leaf beetles (Stork, 1980; Mann and Crowson, 1981; Reid, 1995). This result was reinforced in the early combined morphological plus SSU analysis (Farrell, 1998), and treated as *fait accompli* in the study of the hispine radiation on monocot plants (Wilf et al., 2000; Duckett et al., 2004). However, the DNA data strongly suggest that the occurrence of bifid tarsal setae is convergent, perhaps driven by their parallel colonization of monocots (Gómez-Zurita et al., 2007).

2 The "eumolpine" clade (Eumolpinae, Cryptocephalinae, Cassidinae, Spilopyrinae). The major change in the placement of Cassidinae proposed here to a "eumolpine" clade, specifically as sister to Cryptocephalinae s.l. within a paraphyletic Eumolpinae, is supported by male abdominal characters and the shared ability to produce oothecas (Reid, 1995), internal reproductive organs and hindwing venation Suzuki (1988, 1994), and the male soft reproductive organs (Jolivet and Verma, 2002). It was also supported by the analysis of SSU plus morphology in Duckett et al. (2004), but dismissed by the authors in favor of a reweighted analysis placing Cassidinae basal to the Chrysomelinae + Galerucinae clade, a secondary solution suggested by Reid (1995). The remaining "eumolpines" consist of two clades of Eumolpinae roughly corresponding to the traditional Eumolpini and the Typophorini + Colasposomini (Seeno and Wilcox, 1982; Gómez-Zurita et al., 2005). Among the taxa associated with the latter are Megascelidinae (Cox and Windsor, 1999; references therein) here represented by *Megascelis* sp., which have been consistently recovered within Eumolpinae (Jolivet, 1959; Iablokoff-Khnzorian, 1966; Suzuki, 1988; Reid, 1995; but see Cox and Windsor, 1999) including recent molecular studies (Reid, 1995, 2000; Farrell, 1998; Duckett et al., 2004; Farrell and Sequeira, 2004; Gómez-Zurita et al., 2005). In contrast, Spilopyrinae, stand out as separate at the base of the "eumolpine" lineage, consistent with their plesiomorphic nature (Reid, 2000; Verma and Jolivet, 2004), and clearly deserving subfamilial status. Similarly, *Syneta* should be retained as a subfamily although its precise affinities remain questionable, as they were recovered nested within the Eumolpinae in the parsimony tree, but within the "sagrine" lineage in the ML and Bayesian trees, while morphological analyses have also failed to place them with confidence (Verma and Jolivet, 2000; Gómez-Zurita et al., 2005).

3 The "chrysomeline" clade (Chrysomelinae, Galerucinae and Alticinae). The Galerucinae s.l. were closely linked to Chrysomelinae as reported before (Reid, 1995, 2000; Farrell, 1998; Duckett et al., 2004), but were embedded within a paraphyletic Chrysomelinae in all our analyses. From a morphological perspective, the Chrysomelinae is well defined by several apomorphies, but internal subdivisions are unclear (e.g., Daccordi, 1994). We found three major groups of Chrysomelinae, with the deepest split separating the tribe Timarchini, consistent with morphology (e.g., Daccordi, 1994), which should be considered a separate subfamily Timarchinae as suggested before (Jolivet and Verma, 2002; Gómez-Zurita, 2004). These three clades, Timarchinae–Chrysomelina/Phyllodectina–other Chrysomelini, are highly congruent with larval types (Kimoto, 1962; Takizawa, 1976) and chemical defensive compounds (Pasteels et al., 1994, 2003): larvae without defensive glands and undergoing three molts (Timarchinae), larvae with several defensive glands and three molts and production of nitropropanoic acid and isoxazolinone glucosides (Chrysomelina and Phyllodectina), and larvae with one pair of defensive glands and four molts and production of cardenolides and the so-called

dipeptide 43 (Chrysolinina, Doryphorina, Gonioctenina, Paropsina, and others).

Conclusions

With the ever-increasing size of molecular phylogenetic data sets, fast procedures for sequence alignment and tree searches are required. Here we compared direct optimization, arguably an analytically and philosophically well justified method (Wheeler et al., 2006), with other methods of two-step alignment procedures, including an alignment obtained with the very fast BLAST algorithm and approximate ML tree searches. Although not a strict comparison of methodology, our analysis shows that these methods are defensible in terms of the quality of homology assignment (overall level of synapomorphy) and criteria of taxonomic congruence.

Equally, only the combining of three markers provided sufficient phylogenetic signal to resolve basal relationships in the Chrysomelidae, although support levels at many nodes remained low, possibly exacerbated by the rapid early diversification of the leaf beetles (Gómez-Zurita et al., 2007). The study also highlights the problems of previous molecular sytematics studies of Chrysomelidae. While they also used the SSU gene, which provided the strongest signal of the three markers employed here, due to a perceived lack of power the tree searches were conducted in simultaneous analyses with an existing morphological data set. However, the latter, upweighted relative to molecular characters in these analyses, have greatly influenced the combined analysis topology, while little effort was made to test conflicting signal with the molecular data. Hence, apparent convergences in morphological character systems, as those uniting the monocot feeding groups, remained undetected. Only a greatly increased taxon sampling and addition of new molecular markers provided a satisfactory tree. This can now be used towards a consensus in the systematic arrangement of this fascinating beetle group and the reassessment of conflicting morphological characters.

Acknowledgments

We thank all the colleagues providing samples for this research. Thanks also to the Computer Center at Imperial College London and their assistance to use the Viking cluster. Dr J. M. Carpenter and two anonymous referees provided useful criticism to improve this paper. This study has benefited from funding from Marie Curie Actions HPMF-CT-2000-00744 and MERG-CT-2004-50004, and by a Humboldt Research Fellowship (SPA-1121941) to JGZ, a NERC CASE studentship to TH, and Leverhulme Trust grant F/969/H to APV and P. Hammond.

References

Becerra, J.X., 1997. Insects on plants: macroevolutionary chemical trends in host use. Science 276, 253–256.

Belshaw, R., Katzourakis, A., 2005. BlastAlign: a program that uses blast to align problematic nucleotide sequences. Bioinformatics 21, 122–123.

Borowiec, L., 1995. Tribal classification of the cassidoid Hispinae (Coleoptera. Chrysomelidae). In: Pakaluk, J., Slipinski, S.A. (Eds.), Biology, Phylogeny and Classification of Coleoptera: papers celebrating the 80th birthday of Roy A. Crowson. Muzeum i Instytut Zoologii PAN, Warszawa, pp. 541–558.

Böving, A.G., 1929. Beetle larvae of the subfamily Galerucinae. Proc. US Natl Mus. 75, 1–48.

Chapuis, F., 1874. Famille Des Phytophages. In: Lacordaire, T. (Eds.), Histoire Naturelle des insectes, genera des coléoptères, Vol. 10. Encyclopédique de Roret, Lib., Paris, pp. 1–455.

Chen, S., 1934. Recherches sur les Chrysomelinae de la Chine et du Tonkin. PhD Thesis, Société Entomologique de France, Paris.

Chen, S., 1985. Phylogeny and classification of the Chrysomeloidea. Entomography 3, 465–475.

Cox, M.L., 1981. Notes on the biology of *Orsodacne*, with a subfamily key to the larvae of British Chrysomelidae. Entomol. Gaz. 32, 123–136.

Cox, M.L., Windsor, D.A., 1999. The first instar larva of *Aulacoscelis sp. & Megascelis puella* Lacordaire (Coleoptera: Chrysomelidae: Aulacoscelinae, Megascelinae) and their value in the placement of the Aulacoscelinae and Megascelinae. In: Cox, M.L. (Ed.), Advances in Chrysomelidae Biology. Backhuys Publishers, Leiden, pp. 51–70.

Crowson, R.A., 1955. The Natural Classification of the Families of Coleoptera. E.W. Classey, Hampton, UK.

Crowson, R.A., 1981. The Biology of Coleoptera. Academic Press, London.

Crowson, R.A., Crowson, E.A., 1996. The phylogenetic relations of Galerucinae-Alticinae. In: Jolivet, P., Cox, M.L. (Eds.), Chrysomelidae Biology, Vol. 1: The Classification, Phylogeny and Genetics. SPB Academic Publishing, Amsterdam, pp. 97–118.

Daccordi, M., 1994. Notes for phylogenetic study of Chrysomelinae, with descriptions of new taxa and a list of all the known genera (Coleoptera: Chrysomelidae, Chrysomelinae). In: Furth, D.G. (Ed.), Proceedings of the Third International Symposium on the Chrysomelidae, Beijing 1992. Backhuys Publishers, Leiden, pp. 60–84.

Duckett, C.N., Gillespie, J.J., Kjer, K.M., 2004. Relationships among the subfamilies of Chrysomelidae inferred from small subunit ribosomal DNA and morphology, with special emphasis on the relationship among the flea beetles and the Galerucinae. In: Jolivet, P., Santiago-Blay, J.A., Schmitt, M. (Eds.), New Developments in the Biology of Chrysomelidae. SPB Academic Publishing, The Hague, pp. 3–18.

Ehrlich, P.R., Raven, P.H., 1964. Butterflies and plants: a study in coevolution. Evolution, 18, 586–608.

Erber, D., 1988. Biology of Camptosomata–Clytrinae–Cryptocephalinae–Chlamisinae–Lamprosomatinae. In: Jolivet, P., Petitpierre, E., Hsiao, T.H. (Eds.), Biology of Chrysomelidae. Kluwer Academic Publishers, Dordrecht, pp. 513–552.

Farrell, B.D., 1998. 'Inordinate Fondness' explained: Why are there so many beetles? Science 281, 555–559.

Farrell, B.D., Sequeira, A.S., 2004. Evolutionary rates in the adaptive radiation of beetles on plants. Evolution 58, 1984–2001.

Fleissner, R., Metzler, D., Haeseler, A.V., 2005. Simultaneous statistical multiple alignment and phylogeny reconstruction. Syst. Biol. 54, 548–561.

Furth, D.G., Suzuki, K., 1994. Character correlation studies of problematic genera of Alticinae in relation to Galerucinae (Coleoptera: Chrysomelidae). In: Furth, D.G. (Ed.), Proceedings of the Third International Symposium on the Chrysomelidae, Beijing 1992. Backhuys Publishers, Leiden, pp. 116–135.

Gascuel, O., 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. 14, 685–695.

Gatesy, J., DeSalle, R., Wheeler, W., 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. Mol. Phylogenet. Evol. 2, 152–157.

Gatesy, J., O'Grady, P., Baker, R., 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. Cladistics 15, 271–313.

Giannini, N.P., Simmons, N.B., 2003. A phylogeny of megachiropteran bats (Mammalia: Chiroptera: Pteropodidae) based on direct optimization analysis of one nuclear and four mitochondrial genes. Cladistics 19, 496–511.

Goloboff, P.A., 1999. Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics 15, 415–428.

Gómez-Zurita, J. 2004. Molecular systematics and time-scale for the evolution of Timarcha, a leaf-beetle genus with a disjunct Holarctic distribution. Mol. Phylogenet. Evol. 32, 647–665.

Gómez-Zurita, J., Jolivet, P., Vogler, A.P., 2005. Molecular systematics of Eumolpinae and the relationships with Spilopyrinae (Coleoptera, Chrysomelidae). Mol. Phylogenet. Evol. 34, 584–600.

Gómez-Zurita, J., Hunt, T., Kopliku, F., Vogler, A.P., 2007. Recalibrated tree of Leaf Beetles (Chrysomelidae) indicates independent diversification of angiosperms and their insect herbivores. PLoS ONE, 2(4), e360. doi: 10.1071/journal.pone.0000360.

Grant, T., Kluge, A.G., 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. Cladistics 19, 379–418.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies. Syst. Biol. 52, 696–704.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174.

Hein, J., 1990. Unified approach to alignment and phylogenies. Methods Enzymol. 183, 626–645.

Hsiao, T.H., Windsor, D.M., 1999. Historical and biological relationships among Hispinae inferred from 12S mtDNA sequence data. In: Cox, M.L. (Ed.), Advances in Chrysomelidae Biology. Backhuys Publishers, Leiden, pp. 39–50.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Iablokoff-Khnzorian, S.M., 1966. Considérations sur l'édéage des Chrysomelidae et son importance phylogénique. L'Entomologiste 22, 115–136.

Jacoby, M., 1908. Fauna of British India: Coleoptera. Chrysomelidae, Vol. I. Taylor & Francis, London.

Jolivet, P., 1959. Recherche sur l'aile des Chrysomeloidea (Col.). Mém. Inst. R. Sci. Nat. Belg. 2, 1–180.

Jolivet, P., Verma, K.K., 2002. Biology of Leaf Beetles. Intercept Publishers, Andover.

Kim, S.J., Kjer, K.M., Duckett, C.N., 2003. Comparison between molecular and morphological-based phylogenies of galerucine/alticine leaf beetles (Coleoptera: Chrysomelidae). Insect Syst. Evol. 34, 53–64.

Kimoto, S., 1962. A phylogenic consideration of Chrysomelinae based on immature stages of Japanese species (Coleoptera). J. Fac. Agric., Kyushu Univ. 12, 67–87.

Konstantinov, A.S., 1994. Comparative morphology and some evolutionary trends in flea beetles. In: Jolivet, P.H., Cox, M.L., Petitpierre, E. (Eds.), Novel Aspects of the Biology of Chrysomelidae. Kluwer Academic Publishers, Dordrecht, pp. 383–391.

Kuschel, G., May, B.M., 1990. Palophaginae, a new subfamily for leaf beetles, feeding as adult and larva on araucarian pollen in Australia (Coleoptera: Megalopodidae). Invert. Taxon 3, 697–719.

Lee, J.E., 1993. Phylogenetic studies on the larvae of the Chrysomelidae (Coleoptera) from Japan. Jap. J. Entomol. 61, 409–424.

Lemmon, A.R., Milinkovitch, M.C., 2002. The metapopulation genetic algorithm: an efficient solution for the problem of large phylogeny estimation. Proc. Natl Acad. Sci. USA 99, 10516–10521.

Lingafelter, S.W., Konstantinov, A.S., 2000. The monophyly and relative rank of alticine and galerucine leaf beetles: a cladistic analysis using adult morphological characters (Coleoptera: Chrysomelidae). Entomol. Scand. 30, 397–416.

Maddison, D.R., Maddison, W.P., 2005. MacClade 4: Analysis of Phylogeny and Character Evolution. Sinauer Assoc, Sunderland, MA.

Mann, J.S., Crowson, R.A., 1981. The systematic position of *Orsodacne* and *Syneta* in relation to characters of larvae, internal anatomy and tarsal vestiture. J. Nat. Hist. 15, 727–748.

Mayr, E., Ashlock, P.D., 1991. Principles of Systematic Zoology, 2nd edn. McGraw-Hill, New York.

Mitter, C., Farrell, B.D., 1991. Macroevolutionary aspects of insect/plant relationships. In: Bernays, E.A. (Ed.), Insect/Plant Interactions, Vol. 3. CRC Press, Boca Raton, FL, pp. 35–78.

Monrós, F., 1960. Los géneros de Chrysomelidae (Coleoptera). Opera Lilloana 3, 1–336.

Nixon, K.C., 1999. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15, 407–414.

Ogden, T.H., Whiting, M.F., Wheeler, W.C., 2005. Poor taxon sampling, poor character sampling, and non-repeatable analyses of a contrived dataset do not provide a more credible estimate of insect phylogeny: a reply to Kjer. Cladistics 21, 295–302.

Pasteels, J.M., Rowell-Rahier, M., Braekman, J.-C., Daloze, D., 1994. Chemical defence of adult leaf beetles updated. In: Jolivet, P.H., Cox, M.L., Petitpierre, E. (Eds.), Novel Aspects of the Biology of Chrysomelidae. Kluwer Academic Publishers, Dordrecht, pp. 289–301.

Pasteels, J.M., Termonia, A., Daloze, D., Windsor, D.M., 2003. Distribution of toxins in chrysomeline leaf beetles: Possible taxonomic inferences. In: Furth, D.G. (Ed.), Special Topics in Leaf Beetle Biology. Proceedings of the Fifth Symposium on the Chrysomelidae. Pensoft Publishers, Sofia, pp. 261–275.

Pons, J., Vogler, A.P., 2006. Size, frequency, and phylogenetic signal of multiple-residue indels in sequence alignment of introns. Cladistics 22, 144–156.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Reid, C.A.M., 1995. A cladistic analysis of subfamilial relationships in the Chrysomelidae sensu lato (Chrysomeloidea). In: Pakaluk, J., Slipinski, S.A., (Eds.), Biology, Phylogeny and Classification of Coleoptera: papers celebrating the 80th birthday of Roy A. Crowson. Muzeum i Instytut Zoologii PAN, Warszawa, pp. 559–631.

Reid, C.A.M., 2000. Spilopyrinae Chapuis: a new subfamily in the Chrysomelidae and its systematic placement (Coleoptera). Invert. Taxon. 14, 837–862.

Ribera, I., Hogan, J.E., Vogler, A.P., 2002. Phylogeny of Hydradephagan water beetles inferred from 18S rRNA sequences. Mol. Phylogenet. Evol. 23, 43–62.

Samuelson, G.A., 1994. An elytron to body meshing mechanism of possible significance in the higher classification of Chrysomelidae. In: Furth, D.G. (Ed.), Proceedings of the Third International Symposium on the Chrysomelidae, Beijing 1992. Backhuys Publishers, Leiden, pp. 136–147.

Sankoff, D., 1975. Minimal mutation trees of sequences. Siam J. Appl. Math., 28, 35–42.

Schmitt, M., 1994. The position of Megalopodinae and Zeugophorinae in a phylogenetic system of the Chrysomeloidea (Coleoptera). In: Furth, D.G. (Ed.), Proceedings of the Third International

Symposium on the Chrysomelidae, Beijing 1992. Backhuys Publishing, Leiden, pp. 38–44.
Seeno, T.N., Wilcox, J.A., 1982. Leaf beetle genera (Coleoptera: Chrysomelidae). Entomography 1, 1–221.
Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.
Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49, 369–381.
Staines, C.L., 2002. The new world tribes and genera of hispines (Coleoptera: Chrysomelidae: Cassidinae). Proc. Entomol. Soc. Wash. 104, 721–784.
Stork, N.E., 1980. A scanning electron microscope study of tarsal adhesive setae in the Coleoptera. Zool. J. Linn. Soc. 68, 173–306.
Suzuki, K., 1988. Comparative morphology of the internal reproductive system of the Chrysomelidae (Coleoptera). In: Jolivet, P., Petitpierre, E., Hsiao, T.H. (Eds.), Biology of Chrysomelidae. Kluwer Academic Publishing, Dordrecht, pp. 317–355.
Suzuki, K., 1994. Comparative morphology of the hindwing venation of the Chrysomelidae (Coleoptera). In: Jolivet, P.H., Cox, M.L., Petitpierre, E. (Eds.), Novel Aspects of the Biology of Chrysomelidae. Kluwer Academic Publishing, Dordrecht, pp. 337–354.
Suzuki, K., 1996. Higher classification of the family Chrysomelidae (Coleoptera). In: Jolivet, P.H.A., Cox, M.L. (Eds.), Chrysomelidae Biology, Vol. 1: the Classification, Phylogeny and Genetics. SPB Academic Publishing, Amsterdam, pp. 3–54.
Suzuki, K., 2003. Systematic position of the subfamilies Megalopodinae and Megascelinae (Chrysomelidae) based on the comparative morphology of internal reproductive system. In: Furth, D.G. (Ed.), Special Topics in Leaf Beetle Biology, Proceedings of the 5th International Symposium on the Chrysomelidae, Iguassu Falls, 2000. Pensoft Publishers, Sofia, pp. 105–116.
Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, MA.
Takizawa, H., 1976. Larvae of the genus *Gonioctena* Chevrolat (Coleoptera, Chrysomelidae): Descriptions of Japanese species and the implicaitons of larval characters for the phylogeny. Kontyû, 44, 444–468.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Verma, K.K., 1996. Inter-subfamily relations among Chrysomelidae (Coleoptera) as suggested by organization of the male genital system. In: Jolivet, P., Cox, M.L. (Eds.), Chrysomelidae Biology, Vol. 1: The Classification, Phylogeny and Genetics. SPB Academic Publishing, Amsterdam, pp. 317–351.
Verma, K.K., Jolivet, P., 2000. Phylogeny of Synetinae – reconsidered. Nouv. Rev. Entomol. 17, 35–49.
Verma, K.K., Jolivet, P., 2002. Comments on Spilopyrinae (Col. Chrys.). Nouv. Rev. Entomol. 19, 99–110.
Verma, K.K., Jolivet, P., 2004. The primitive Eumolpinae and the Gondwana hypothesis. In: Jolivet, P., Santiago-Blay, J.A., Schmitt, M. (Eds.), New Developments in the Biology of Chrysomelidae. SPB Academic Publishing, The Hague, The Netherlands, pp. 395–406.
Wheeler, W.C., 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. Syst. Biol. 44, 321–331.
Wheeler, W.C., 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics? Cladistics 12, 1–9.
Wheeler, W.C., 2003. Iterative pass optimization of sequence data. Cladistics 19, 254–260.
Wheeler, W.C., Aagesen, L., Arango, C.P., Faivovich, J., Grant, T., D'Haese, C., Janies, D., Smith, W.L., Varón, A., Giribet, G. 2006. Dynamic Homology and Phylogenetic Systematics: A Unified Approach Using POY. American Museum of Natural History, New York.
Wheeler, W.C., Gladstein, D.S., De Laet, J., 2002. POY, Version 3.0. ftp.amnh.org/pub/molecular/poy/
Whiting, A.S., Sites, J.W. Jr, Pellegrino, K.C.M., Rodrigues, M.T., 2006. Comparing alignment methods for inferring the history of the new world lizard genus *Mabuya* (Squamata: Scincidae). Mol. Phylogenet. Evol. 38, 719–730.
Wilf, P., Labandeira, C.C., Kress, W.J., Staines, C.L., Windsor, D.M., Allen, A.L., Johnson, K.R., 2000. Timing the radiation of leaf beetles: Hispines on gingers from latest Cretaceous to recent. Science 289, 291–294.

Appendix 1

Commands used for direct optimisation tree search in POY 3.0.11 following the strategy of Giannini and Simmons (2003).

RAS + TBR search

poy -parallel -jobspernode 2 [data files] -outgroup Ves_san_Ves_VES -gap 1 -replicates 40 -tbr -nospr -maxtrees 2 treefuse -fuselimit 10000 fusemingroup 1 -fusemaxtrees 1000 –time -printtree -norandomizeoutgroup -impliedalignment -repintermediate > [output file]

Ratchet

poy -parallel -jobspernode 2 [data files] -outgroup Ves_san_Ves_VES -gap 1 -replicates 0 -nospr -notbr -topofile shorttree.txt -ratchettbr 20 -ratchetpercent 15 -ratchettrees 2 -ratchetseverity 4 -time -printtree -poytreefile [RAS + TBR tree file] -norandomizeoutgroup -impliedalignment > [output file]

Iterative pass search

The search was incomplete; not enough memory; it ran for 48 h):

poy -parallel -jobspernode 2 [data files] -gap 1 -replicates 0 -tbr -nospr -slop 1 -checkslop 5 -iterativepass -iterativelowmem -topofile [ratchet tree file] -time -printtree -norandomizeoutgroup -impliedalignment -poystrictconsensuscharfile [output tree file] > [output file]

Bremer support

poy -parallel -jobspernode 2 [data files] -noleading -outgroup Ves_san_Ves_VES -norandomizeoutgroup -molecularmatrix [equal weights matrix] -replicates 10 -sprmaxtrees 2 -tbrmaxtrees 3 -maxtrees 5 -holdmaxtrees 20 -fitchtrees -seed −1 -buildsperreplicate 3 -buildspr -buildtbr -approxbuild -buildmaxtrees 2 -minstop 3 -treefuse -fuselimit 7 -treefusespr -treefusetbr -numdriftchanges 2 -driftspr -numdriftspr 2 -drifttbr -numdrifttbr 2 -bremer -constrain [iterative pass tree file] > [output file]