diversity

MDPI

Article

Statistical Evaluation of Monophyly in the 'Broad-Nosed Weevils' through Molecular Phylogenetic Analysis Combining Mitochondrial Genome and Single-Locus Sequences (Curculionidae: Entiminae, Cyclominae, and Hyperinae)

Conrad P.D.T. Gillett [1,2,3,*], Christopher H. Lyal [2], Alfried P. Vogler [2,4] and Brent C. Emerson [1,5]

1 School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK
2 Department of Life Sciences, The Natural History Museum, London SW7 5BD, UK; c.lyal@nhm.ac.uk (C.H.L.); a.vogler@nhm.ac.uk (A.P.V.)
3 College of Tropical Agriculture and Human Resources, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA
4 Department of Life Sciences, Silwood Park Campus, Imperial College London, Ascot SL5 7QN, UK
5 Island Ecology and Evolution Research Group, Instituto de Productos Naturales y Agrobiología, 38206 La Laguna, Tenerife, Spain; bemerson@ipna.csic.es
* Correspondence: conradgillett@gmail.com

Received: 20 January 2018; Accepted: 2 April 2018; Published: 6 April 2018

Abstract: Establishing well-supported monophyletic groups is a key requirement for producing a natural classification that reflects evolutionary descent. In a phylogenetic framework this is best achieved through dense taxon sampling and the analysis of a robust character dataset, combined with statistical testing of topological hypotheses. This study assesses the monophyly of tribes and subfamilies within the diverse 'broad-nosed weevils' (Curculionidae: Entiminae, Cyclominae and Hyperinae) through analysis of single-locus sequence data for mitochondrial *cox1* and *rrnL* genes, in combination with a 'backbone' of complete and near-complete mitochondrial genome sequences. Maximum likelihood phylogenetic analyses incorporating topological constraints for various higher-taxa were statistically tested using the AU, SH, and KH tests, which indicated that three tribes within Entiminae, as presently classified, are not monophyletic. Moderate and high bootstrap support was also consistent with two entimine tribes (Peritelini and Cylydrorhinini) being each recovered as monophyletic in an unconstrained analysis. Furthermore, one genus of cyclomine weevils (*Aphela*) is recovered outside the clade of 'broad-nosed weevils', although its taxonomic placement remains uncertain. It is apparent that the present approach may be hampered by limited taxon sampling in the 'backbone' dataset, rendering it difficult for divergent taxa to robustly match to their closest lineages. However, with improved taxon sampling of the mitogenome tree, the general approach can be a useful taxonomic tool for weevils.

Keywords: constraint analysis; AU test; SH test; KH test; mitochondrial genomes; Curculionoidea

1. Introduction

The fundamental aim of phylogeny reconstruction is to summarise genealogically determined evolutionary relationships as phylogenetic trees, visually tracing the historical course of speciation, organised through the relative recency of common ancestry [1,2]. Together with other data, such as geographic distributions and ecological traits for species under consideration, phylogenies can be powerful tools for explaining observed patterns, and for testing hypothesised processes of speciation.

Of paramount importance when inferring biological and systematic meaning from trees is the formulation of a sound basis for identifying natural groups of taxa, from which broader conclusions and predictions can be made regarding the biology of the included species. Such predictions might include the identification of lineage-specific host-plant use, breeding behavior, or even geographic distribution. Deciphering which groups of organisms are natural (or monophyletic) is a prerequisite for constructing a hierarchical classification system that reflects their underlying evolutionary history. A well sampled dataset, containing taxa of as many potential lineages of the taxon of interest as possible, is crucial for the meaningful testing of monophyletic groups in order to increase confidence in the resulting topologies. However, because comprehensive taxon sampling in very diverse groups containing thousands of species, such as the weevils (Coleoptera: Curculionoidea), is very difficult in practice [3], alternative sources of data other than specifically collected specimens should be investigated to enhance taxon coverage. Such data can be obtained from public repositories of DNA sequence data held in freely accessible online databases such as the National Center for Biotechnology Information's GenBank [4]. Other databases also exist, for example The Barcode of Life Data System (BOLD) [5] but GenBank is by far the most comprehensive, at present holding more than 206 million sequences belonging to almost 260,000 described species, submitted by research laboratories across the world (NCBI GenBank Flat File release 223.0, 15 December 2017) [4].

Statistical tests available to undertake hypothesis testing between competing ML tree topologies generally utilise the likelihood values (for each tree this is the product of all per-site likelihoods in the input alignment) for calculation of test statistics. Such tests include the Shimodaira–Hasegawa (SH) test [6] and the Kishino–Hasegawa (KH) test [7] which both compare the log-likelihoods of two trees to produce a probability statistic for each of them. In the SH test, the trees tested are selected a posteriori, whereas in the KH test, the trees are selected a priori [8]. Both these tests have biases and limitations, including a correlation between the SH test results with the number of trees being tested (rendering the test conservative in rejecting trees) and the inability of the KH test to control for type 1 errors [9]. An alternative test that is able to correct for the tree selection bias is the approximately unbiased (AU) test [9]. The AU test is based upon bootstrap resampling of the per-site log-likelihoods of the input alignment, which allows for the alignment length to be altered and the newly bootstrapped probabilities being scaled to the original alignment length [8]. The AU test statistic is calculated from the change in BS probabilities for each bootstrapped set of replicates. This test is able to control for type 1 errors and is currently one of the most widely employed methods to assess topologies under the ML optimality criterion. To statistically test whether monophyly of any of the higher taxa constrained as described below could be rejected, the AU test was implemented to obtain the confidence set of trees. This was achieved through resampling the per-site log-likelihood of the input alignment by changing the alignment length and drawing new BS samples from these lengths. The number of times the hypothesis is supported by the BS replicates is used to calculate the BS support for different sequence lengths; the AU test then calculates a *p*-value from the change in bootstrap values along the changing sequence length [9].

We concur with recent opinion promoting the modern classification of weevils as a data-driven science [10], and the present study therefore aims to contribute to this by testing the monophyly of tribes and subfamilies within the diverse group of weevil subfamilies known as the 'broad-nosed weevils' (Curculionidae subfamilies Entiminae, Cyclominae, and Hyperinae) using sequences obtained from GenBank to enhance the taxon coverage of these groups in a phylogeny of Curculionoidea previously constructed from complete and near-complete mitochondrial genomes (mitogenomes) [11]. The approach used is analogous to that employed in a study which obtained short (<100 bp) phylogenetically informative amplicons (SPIAs) of the mitochondrial 16S ribosomal large subunit gene (*rrnL*) from DNA-degraded specimens of weevils and incorporated them into a 'backbone' phylogeny built from a concatenation of longer sequences from five loci (including *rrnL*) [12]. The process tested here differs in that, instead of SPIAs, longer 'complete' sequences of mitochondrial *cox1* and *rrnL* genes obtained from GenBank are added to the mitogenome 'backbone' phylogeny, containing sequences of

both those loci and 13 other genes, in order to identify the lineages to which the database sequences are most closely related under a maximum likelihood (ML) optimality criterion.

The concept of 'broad-nosed weevils' dates back to an 1863 work of Lacordaire [13], who divided his family 'Curculionides' into two groups: the Adelognatha and the Phanerognatha. The former of these represents the broad-nosed weevils, defined morphologically by having the prementum covering the maxillae and by the possession of deciduous processes on the adult mandibles [14,15], in addition to bearing the distinctive relatively short rostrum that eventually gave rise to their popular name. Interpretation of precisely which taxonomic groups are characterised as broad-nosed weevils has varied according to the opinion of different authors [14,16]. One widely used definition [17], which was tested in assessing the monophyly of broad-nosed weevils based upon larval and adult morphological characters [18], contained the following higher taxa sensu the family-level catalogue of Bouchard et al. [19]: Brachyceridae, Ithycerinae, and Microcerinae (subfamilies of Brentidae); Gonipterini (tribe of Curculioninae); Entiminae, Cyclominae, and Hyperinae (subfamilies of Curculionidae). That study concluded that broad-nosed weevils are not monophyletic, with the Ithycerinae, Microcerinae, and Brachyceridae recovered as forming three stepwise basal lineages (Ithycerinae most basal) and the Entiminae + Cyclominae forming an apical clade (Hyperinae was not analysed) [18]. This result, together with the results in the mitogenome phylogeny [11] and those based on other molecular data [20–24] represent strong independent evidence that Brachyceridae, Ithycerinae, and Microcerinae form separate basal lineages to those 'broad-nosed' weevils classified within Curculionidae sensu Bouchard et al. [19] except Platypodinae (i.e., Entiminae + Cyclominae + Hyperinae). For the purposes of this study, only the latter group is defined and henceforth referred to as the 'broad-nosed weevils', within which the monophyly of various taxa is tested. Throughout this article we employ an existing definition and terminology in naming the other large true weevil clade, containing the subfamilies Curculioninae + Conoderinae + Cossoninae + Molytinae + Scolytinae, as the 'CCCMS clade' [23].

Selection of the *cox1* 5′ region and *rrnL* as the short loci to be added to the mitogenome data was made based upon the fact that a large number of sequences for these genes are available on GenBank owing to their wide use in phylogenetics research, and in the case of *cox1*, its ubiquitous use as the 'barcode' region of choice for molecular-based species identifications [25]. The 'backbone' phylogeny of Curculionoidea, constructed with mitogenome data from 120 weevil taxa (in seven families, including 67 tribes of Curculionidae) [11] is highly congruent with other molecular hypotheses of weevil relationships [20–24] and clearly demonstrates the well supported division of the Curculionidae *s.str.* into two large clades, one of which represents the monophyletic 'broad-nosed weevils' as defined below, and recovered with 100% bootstrap (BS) support in that analysis. The 'broad-nosed weevils' are selected for further investigation of tribal relationships because of their unambiguous monophyly and the comparatively large number of taxa represented in the mitogenome phylogeny (33 species in 19 tribes), maximising the number of lineages available for study. Additionally, one of its component subfamilies, the Entiminae, is the most speciose subfamily-level taxon in Curculionidae, containing an estimated 12,000 described species globally [16]. Although Entiminae has generally been recovered as monophyletic [17] or paraphyletic [20–24] with respect to the other broad-nosed weevil subfamilies, Hyperinae and/or Cyclominae, in molecular analyses, its internal tribal structure is not well understood, with as many as 55, and as few as 5 tribes proposed [18,26]. Consequently, these relationships are in need of further investigation. General life-history in the Entiminae consists of adult feeding on leaves and shoots, and larval development in underground roots, with apparently low host-plant specificity [16]. Identification of lineage specific life-history traits is also a desirable goal of phylogenetic analyses combined with biological observations.

This study is therefore both an exploration of the phylogenetic utility of incorporating shorter sections of sequence data into a longer alignment, and a test of monophyly of the tribes and subfamilies for which sequences of more than one taxon are available, undertaken in a real-world scenario of combining newly generated sequences with publicly available ones.

2. Materials and Methods

2.1. 'Backbone' Phylogeny

The mitogenome sequences for 120 curculionid taxa analysed in a previous phylogeny [11], (Supplementary Figure S1 and File S1) were used in the phylogenetic reconstructions in this study, acting as a comprehensive phylogenetic framework insofar as they provided the 'backbone' in the resulting trees. Shorter single loci sequences for *cox1* 5′ and *rrnL* obtained from GenBank (as described below) were added to the data-matrix for a combined analysis.

2.2. Public Database Sequences

Automated extraction of sequence data from GenBank was achieved through the use of a series of Perl scripts originally developed as part of a custom-built bioinformatics pipeline for analysing public database sequence data [27,28]. These greatly facilitate the selection of both taxa and loci of interest from amongst all the sequences available, in addition to expediting the process of sequence retrieval. Similar scripts were successfully used to reconstruct a very large phylogeny of >8000 Coleoptera species from analysis of four nuclear and mitochondrial loci obtained from GenBank, indicating the importance of such databases as a source of freely available data [29]. The pipeline was used here only for the selection and retrieval of sequences; subsequent sequence alignment and phylogenetic analyses were undertaken separately. All scripts were run on the Natural History Museum 'ctag' Linux-based bioinformatics server.

The GenBank dataset was further reduced to a maximum of five species per genus following a preliminary ML analysis containing all downloaded GenBank broad-nosed weevil *cox1* and *rrnL* sequences (180 and 175 sequences respectively, representing 278 species-level taxa), combined with the mitogenome data from 120 taxa in Gillett et al. [11]. The alignment step and analysis was otherwise identical to that described below for the unconstrained analysis. The results of this allowed for objective selection of divergent species (sometimes recovered in clearly different lineages) within each genus to ensure that no bias for closely related species was made when choosing taxa to retain for further analysis. Wherever possible, taxa represented by both *cox1* and *rrnL* loci were preferentially selected to reduce missing data. Additionally, all taxonomic names were corrected for any mistakes and to ensure that genera had been assigned to tribes and subfamilies according to the generic catalogue of Alonso-Zarazaga and Lyal [26], except the genera *Aphela* and *Bronchus*, which are now classified in the cyclomine tribes Notiomimetini and Hipporhinini respectively, according to a recent review of the Cyclominae [30].

2.3. Multiple Sequence Alignment and Dataset Concatenation

Prior to alignment, the *cox1* 5′ and *rrnL* GenBank sequences were added to the corresponding whole mitogenome *cox1* and *rrnL* sequences to construct the combined GenBank + whole mitogenome dataset. Whole mitogenome sequences for the genes *nad5*, *nad4*, *nad4L*, and *nad1*, which are transcribed on the reverse strand of the mitogenome, were reverse complemented prior to alignment. Sequences for each of the 13 protein-coding and 2 ribosomal RNA genes were individually aligned using the MAFFT version 7 online server, incorporating the FFT-NS-i slow iterative refinement strategy [31], with the following parameter values: nucleotide scoring matrix 200PAM/k = 2, gap open penalty = 1.53, offset value = 0 [31]. Alignments were thereafter checked manually in Geneious 5.4 [32] for quality and to ensure that protein-coding genes were in the correct reading frame. The resulting individual gene alignments were concatenated together in mitogenome gene order to create the final dataset in Phylip format for phylogenetic analysis.

2.4. Monophyly Constraints

In order to test whether monophyly of any of the subfamilies Entiminae, Cyclominae, and Hyperinae, and any of the tribes within the subfamily Entiminae were consistent with

the combined dataset, a series of 20 constraint tree files in Newick format were constructed, each topologically constraining one subfamily or tribe within the broad-nosed weevils, as summarised and described in the results. Only groups with two or more species, and which were not recovered as monophyletic in the initial unconstrained ML analysis (an initial test of monophyly), were selected for constraint analysis.

2.5. Phylogenetic Analyses

Both an unconstrained and 20 constrained (as outlined above and in the results) ML analyses were undertaken using RAxML 7.6.6 [33] run on the CIPRES web-based server [34]. To assess nodal support, a rapid BS analysis with 1000 iterations was run simultaneously with tree-building. The dataset was analysed and partitioned by gene because previous analysis of the mitogenome dataset indicated that a partitioned analysis outperforms an unpartitioned one [11]. Therefore, separate estimated models of nucleotide substitution were specified for each gene region in the alignment. A GTRCAT model was implemented for the bootstrapping phase and a GTRGAMMA model was used for final tree inference (GTR + optimisation of substitution rates + optimisation of site-specific evolutionary rates). All trees were visualised in Dendroscope 3 [35] and were rooted with a divergent outgroup within Polyphaga (Chrysomelidae: *Crioceris duodecimpunctata*).

2.6. Statistical Hypothesis Testing

To undertake the AU test, the per-site log-likelihood was computed for each of the unconstrained and 20 constraint trees in RAxML using the —f g algorithm, and written to a Treepuzzle formatted file [8]. These values were then used in the program CONSEL [36] to perform the bootstrap resampling (100,000 replicates per tree) and to calculate the p-values for the AU, SH, and KH tests.

3. Results

3.1. Public Database Sequences

The GenBank-derived dataset obtained via the bioinformatics pipeline contained 107 species of Entiminae, Cyclominae, and Hyperinae. Within Entiminae, 22 tribes, 62 genera, and 92 species were represented. Within Cyclominae, 4 tribes, 10 genera, and 13 species were represented. The Hyperinae was represented by one genus and two species. A total of 68 *rrnL* and 63 *cox1* sequences were obtained and 24 species were represented by sequences from both loci, with 44 species only represented by *rrnL* and 39 species only by *cox1*. Sequence lengths varied between 113–558 bp for *rrnL* and 262–748 bp for *cox1*. Supplementary Table S1 summarises the GenBank-obtained sequence data matrix.

3.2. Phylogenetic Analyses

The GenBank-obtained sequences were combined with the existing mitogenome data (Supplementary Table S2) to yield an aligned matrix of 229 taxa, 15 genes and 13912 positions. The mitogenome sequence data is available in Supplementary File S1. The final dataset contained the following broad-nosed weevil taxa: 27 tribes, 74 genera, and 119 species (121 terminals) of Entiminae; 5 tribes, 14 genera and 18 species of Cyclominae; 1 genus and 3 species of Hyperinae. The following 18 tribes of Entiminae contained more than one species and therefore could be tested for monophyly, initially through the unconstrained ML analysis (as analysed by topology and BS support), and then through the individual constraint analyses: Brachyderini, Celeuthetini, Cylydrorhinini, Cyphicerini, Elytrurini, Eustylini, Geonemini, Laparocerini, Naupactini, Otiorhynchini, Peritelini, Polydrusini, Rhyncogonini, Sciaphilini, Sitonini, Tanymecini, Trachyphloeini, and Tropiphorini. Additionally, the subfamilies Entiminae, Cyclominae, and Hyperinae separately, and the three of them combined as the 'broad-nosed weevils', were each also tested for monophyly using constraint analyses.

The topology of weevil families and subfamilies recovered in the unconstrained ML tree (final ML optimisation likelihood: −789,416.469537) shown in Figure 1 (and Supplementary Figure S2) is highly

congruent with that in the tree generated using the mitogenome data alone [11], (Supplementary Figure S1). Only the position of *Ocladius* (Brachyceridae: Ocladiinae) differs in being placed within the Dryophthoridae + Platypodinae clade in the present analysis, and outside of it in the mitogenome analysis. One other intriguing disparity is the sister relationship recovered between *Aphela* (Cyclominae) and *Bagous* (Bagoinae) in a clade sister to all other Curculionidae *s.str. sensu* Bouchard et al. [19] except Platypodinae. Cyclominae is, in fact, recovered here in six separate lineages, whereas analysis of the mitogenomie data alone (containing much more restricted taxon coverage) resulted in a monophyletic Cyclominae [11], in contrast to most other molecular studies [20,21,23,24]. The division of the remainder of Curculionidae *s.str.* into two large clades is also recovered, although support for the dividing node is reduced to 31% BS from 100% BS in the mitogenome tree alone (Figure 1).

Relationships within the CCCMS clade are similarly highly congruent with the previous mitogenome analyses [11], (Supplementary Figure S1), consisting of a sister relationship between the Scolytini (Scolytinae) and the remaining taxa that are split into two clades, one containing the moderately well supported (70% BS) remaining Scolytinae (except *Coptonotus*) and the other containing the rest of the subfamilies with little support for the monophyly of any of them except Lixinae (100% BS).

Within the clade of focal interest, composed of the broad-nosed weevils, there is generally very low nodal support for the deeper nodes, although some of the more apical nodes are well supported, with 26 of them having support values of 80% BS or higher (Figure 1). Two tribes of Entiminae are recovered as a clade with moderate nodal support in this analysis: the Peritelini (88% BS) and the Cylydrorhinini (69% BS), each represented by two genera and two species.

Because of their monophyly as evaluated through bootstrap analysis, these last two tribes are therefore not considered for further constraint analyses. The remaining 16 tribes of Entiminae were recovered as paraphyletic or polyphyletic and were consequently each constrained as monophyletic (Table 1) in separate RAxML analyses (identical to the unconstrained analyses other than enforcing the topological constraint). The resulting per-site log likelihoods of these trees, estimated separately in RAxML, were used to calculate the AU test statistic as detailed below.

Table 1. Higher-taxa constrained as monophyletic for ML analysis and the AU test of monophyly.

Constrained Taxon	Generic Diversity (No. Genera) *	No. of Genera in Constraint	No. of Terminals in Constraint
Broad-nosed weevils	1585	89	142
Entiminae	1370	74	121
Cyclominae	180	14	18
Hyperinae	35	1	3
Brachyderini	24	2	6
Celeuthetini	75	8	8
Cyphicerini	120	1	2
Elytrurini	6	2	3
Eustylini	17	6	9
Geonemini	39	5	7
Laparocerini	9	3	9
Naupactini	65	9	19
Otiorhynchini	27	1	6
Polydrusini	14	3	6
Rhyncogonini	3	1	3
Sciaphilini	46	4	4
Sitonini	8	1	4
Tanymecini	42	5	6
Trachyphloeini	23	1	2
Tropiphorini	115	6	9
UNCONSTRAINED		147	229

* approximate count, data taken from [26].

3.3. Statistical Hypothesis Testing

Results of the statistical tests carried out in CONSEL indicate that at a significance level $\alpha = 0.05$, the confidence sets are the same across the AU, SH, and KH tests (Table 2), with only trees constraining Otiorhynchini, Brachyderini, and Tropiphorini as monophyletic rejecting the null hypothesis that there is no difference between the trees (i.e., that all unconstrained and constrained trees are equally good explanations of the data). Consequently, for these three tribes, the alternative hypothesis is accepted that their likelihoods are significantly different and therefore their monophyly is rejected.

Table 2. Results of the AU, KH, and SH tests of constrained monophyly of 20 higher taxa and the unconstrained analysis, ranked by likelihood. Log likelihood difference to the best tree is shown, except for the best tree, which shows the negative distance of the second best. The three *p*-values below a significance level $\alpha = 0.05$ are ranked 19–21, and represent the three tribes whose monophyly is rejected (Otiorhynchini, Brachycerini, and Tropiphorini).

Rank (By Likelihood)	Taxon Constrained in ML Tree	ΔLog Likelihood to Best Tree	AU Test *p*-Value	KH Test *p*-Value	SH Test *p*-Value
1	Sitonini	−4.1	0.621	0.526	0.971
2	UNCONSTRAINED	4.1	0.605	0.474	0.948
3	Hyperinae	8.7	0.527	0.396	0.968
4	Laparocerini	11.7	0.573	0.430	0.961
5	Rhyncogonini	18.0	0.513	0.409	0.921
6	Broad-nosed weevils	21.4	0.442	0.378	0.913
7	Polydrusini	23.7	0.431	0.357	0.942
8	Cyphicerini	24.1	0.425	0.362	0.865
9	Geonemini	26.7	0.411	0.355	0.873
10	Elytrurini	29.2	0.395	0.340	0.876
11	Celeuthetini	55.6	0.202	0.213	0.719
12	Naupactini	56.4	0.206	0.185	0.726
13	Cyclominae	70.6	0.132	0.174	0.627
14	Eustylini	72.6	0.176	0.125	0.619
15	Sciaphilini	78.0	0.119	0.153	0.573
16	Entiminae	88.0	0.080	0.100	0.505
17	Trachyphloeini	94.3	0.083	0.059	0.463
18	Tanymecini	99.3	0.054	0.059	0.426
19	Otiorhynchini	204.2	2×10^{-4}	0.006	0.048
20	Brachyderini	241.0	6×10^{-51}	3×10^{-5}	0.007
21	Tropiphorini	483.0	0.001	0	0

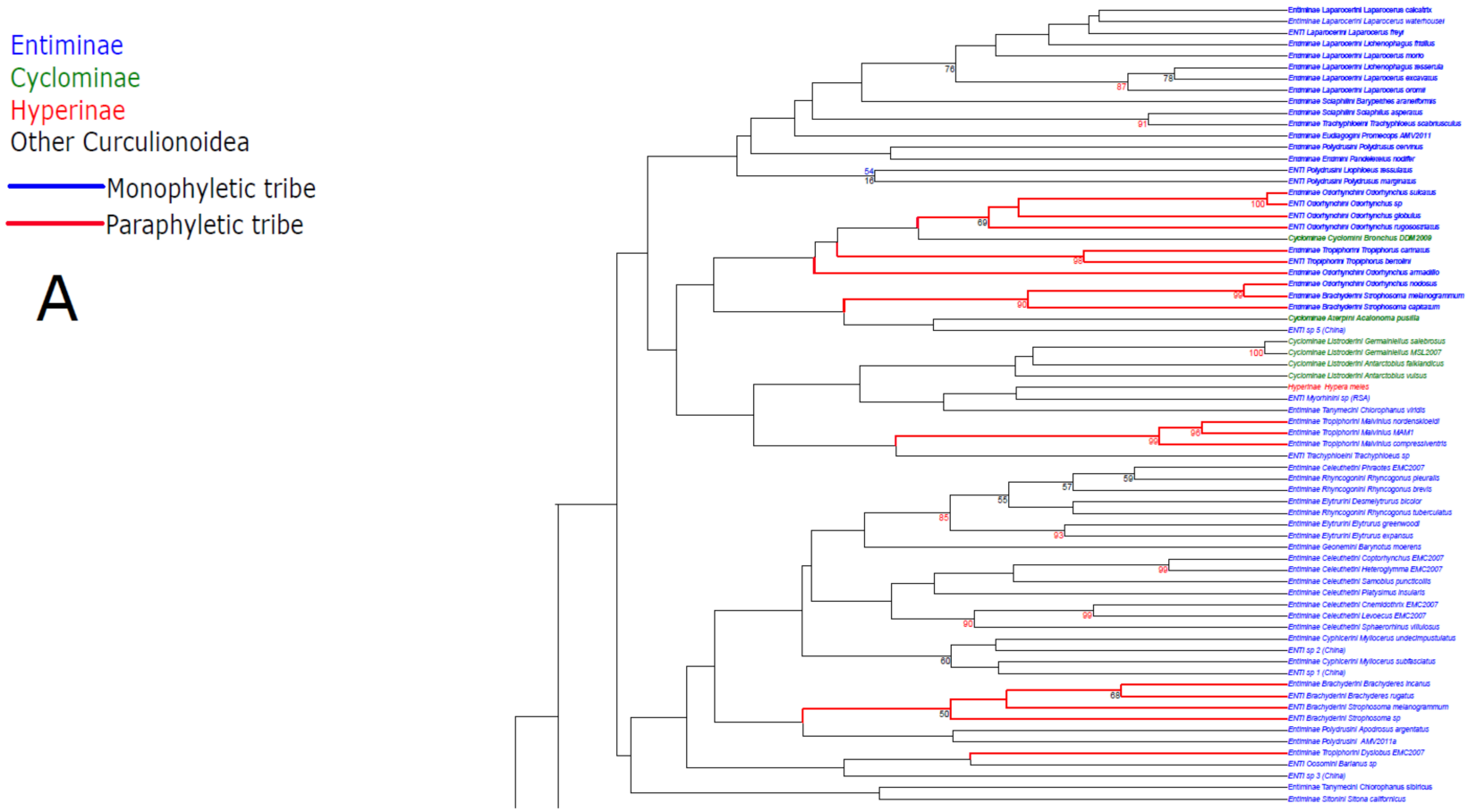

Figure 1. *Cont.*

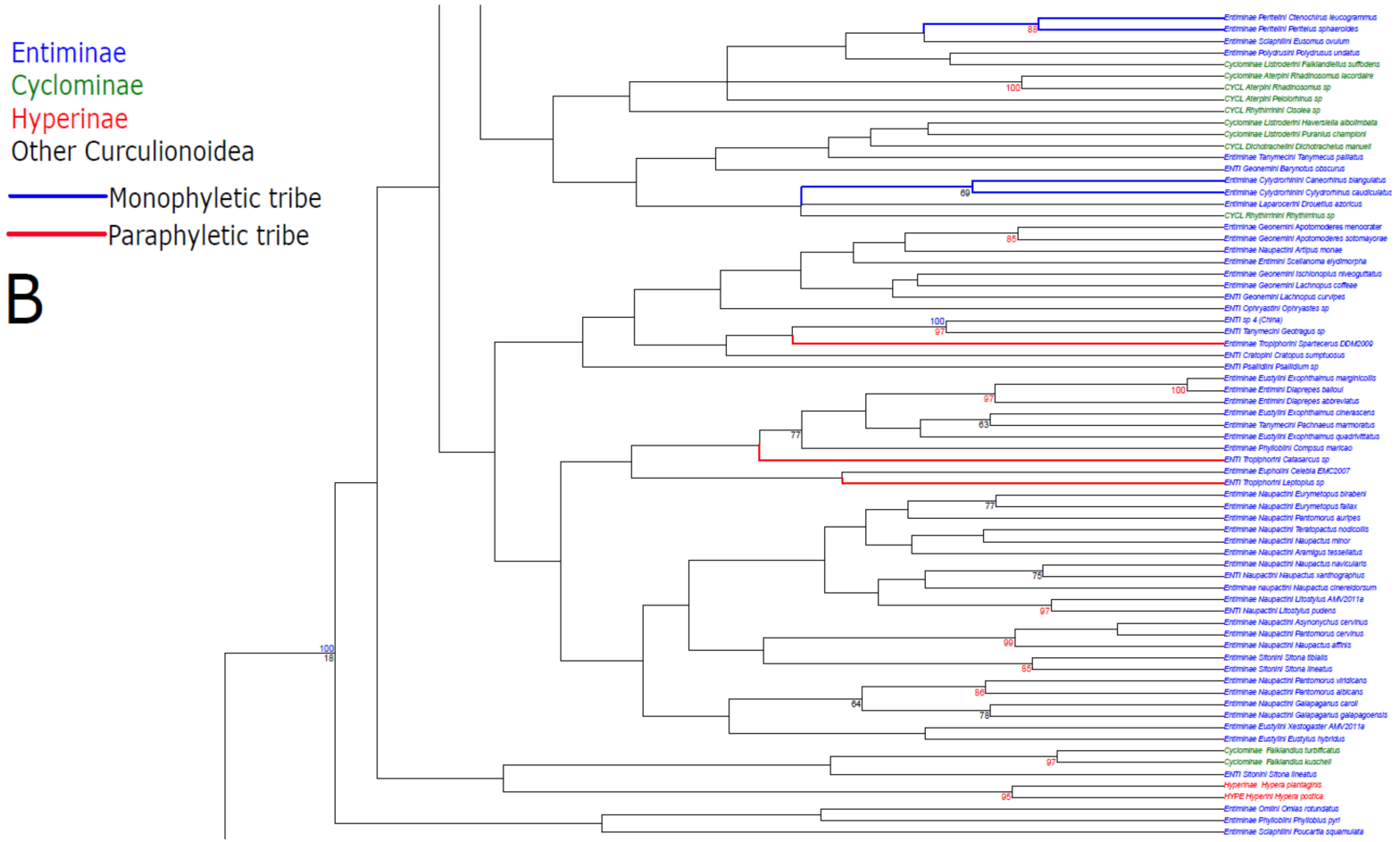

Figure 1. *Cont.*

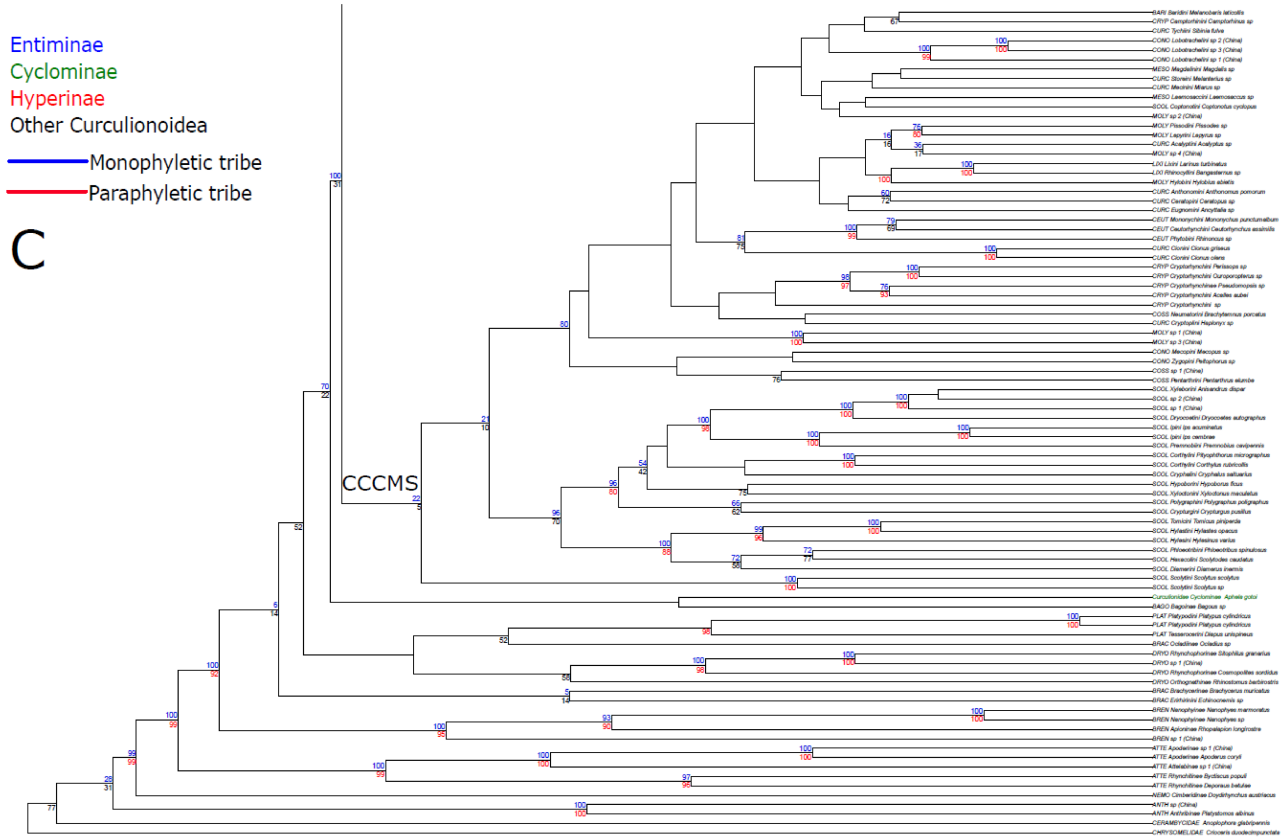

Figure 1. Unconstrained maximum likelihood tree of combined mitogenome and single-locus sequences, rooted at *Crioceris duodecimpunctata* (Chrysomelidae). Shown in three sequential sections (A–C). Bootstrap nodal supports are indicated below nodes, with those of 80% and higher indicated in red. Bootstrap values above nodes (in blue) are shown for consistent nodes in the 'backbone' mitogenome-only data ML tree [11]. Node labelled 'broad-nosed weevils' represents the clade consisting of subfamilies Entiminae + Hyperinae + Cyclominae (with exception of *Aphela*); node labelled CCCMS represents the clade consisting of subfamilies Curculioninae + Conoderinae + Cossoninae + Molytinae + Scolytinae [23]. In green is highlighted the aberrant position of *Aphela gotoi*, currently classified in Cyclominae, but recovered in the CCCMS clade. Taxa represented by mitogenome sequences have family and subfamily codes prefixes as follows: Anthribidae (ANTH), Attelabidae (ATTE), Brachyceridae (BRAC), Brentidae (BREN), Dryophthoridae (DRYO), Nemonychidae (NEMO), Bagoinae (BAGO), Baridinae (BARI), Ceutorhynchinae (CEUT), Conoderinae (CONO), Cossoninae (COSS), Cryptorhynchinae (CRYP), Curculioninae (CURC), Lixinae (LIXI), Mesoptiliinae, (MESO), Molytinae (MOLY), Platypodinae (PLAT), and Scolytinae (SCOL).

4. Discussion

4.1. Unconstrained Analysis

Augmenting the mitogenome dataset with the GenBank sequence data did not strongly affect the main topology with regards to family- and subfamily-level relationships compared to the mitogenome data alone. This was expected because the bulk of phylogenetic signal is present in the full mitogenome alignment and no additional taxa in the deeper portion of the tree were incorporated into this analysis. The single aberrant placement of *Aphela gotoi*, currently classified in the Cyclominae [30], outside the broad-nosed weevil clade, and together with Bagoinae, was the only inconsistency. Although the *Aphela* + *Bagous* relationship has only very weak nodal support (47% BS), it is nevertheless striking that *Aphela*, an apparent broad-nosed weevil, is recovered outside the large Entiminae + Cyclominae + Hyperinae clade, which is otherwise monophyletic in the mitogenome phylogeny of Gillett et al. [11].

When *Aphela* was separately constrained within the Cyclominae and within the broad-nosed weevils, neither of the resulting ML trees was rejected by the AU, SH or KH test, prohibiting a definitive systematic placement. *Aphela* was previously classified within the cyclomine tribe Minyopini, now considered synonymous with the molytine subtribe Plinthina on morphological grounds [26], although *Aphela* itself is presently classified in the cyclomine tribe Notiomimetini [30] and clearly this fact, together with the present molecular findings, indicate that this taxon warrants further investigation with additional sequence data (ideally a full mitogenome sequence). The Cyclominae have been considered a "'subfamily of convenience' for now, sharing no obvious synapomorphic characters" [16], and has consistently been shown to be a paraphyletic taxon in recent molecular studies [20,21,23,24]. The uncertain placement of *Aphela* in our analyses supports these previous results.

The unconstrained analysis indicated that the tribes Peritelini and Cylydrorhinini are each monophyletic in our dataset, although due to the limited taxon sampling of each, interpretation of monophyly beyond the included genera remains putative. Nevertheless, inclusion of the type genera of both these tribes (*Peritelus* and *Cylyndrorhinus* respectively) in the dataset increases objectivity and confidence in at least establishing that each of the other genera included per tribe is correctly classified at present (*Ctenochirus* in Peritelini and *Caneorhinus* in Cylydrorhini), which would not have been the case had the type genera not been analysed.

The tribe Peritelini is large, containing 76 genera with a wide distribution in the Holarctic, Afrotropical, and Australian regions, with new species being continuously discovered even in the relatively well studied European fauna [37,38]. However, morphologically it has not been well defined, and in particular, lacks apomorphies enabling a clear separation from Otiorhynchini [39]. Additionally, at least one genus, *Caenopsis*, has been recently transferred to the tribe Trachyphloeini [39], further highlighting the uncertain monophyly of the group.

In contrast, the tribe Cylydrorhinini is much smaller, containing only six genera, and is of restricted distribution, occurring only in the Australian and southern Neotropical regions. It had previously been classified as a subfamily (Cylyndrorhininae) consisting of two tribes: the Cylyndrorhinini and Listroderini [40]. However, study of larval characters led to the conclusion that the Cylyndrorhinini (in particular the genera *Caneorhinus* and *Cylydrorhinus*, also evaluated here with molecular data) belong in the Entiminae, and the Listroderini belong in the 'Rhytirrhininae', i.e., within the current subfamily Cyclominae [40]. The molecular data indicate that Listroderini is paraphyletic, consisting of three lineages, only one of which, *Germainiellus* + *Antarctobius*, has low support (56% BS), with the two included *Germainiellus* species being well supported as monophletic (100% BS). Whilst the limited taxon sampling in the present study suggests that Cylydrorhinini is monophyletic, no firm conclusions can be drawn with regards to its relationship with Listroderini because of low nodal support in the intervening parts of the tree. This specific relationship was not investigated further with constraint analyses although constraining the Cyclominae as a whole did not lead to the resulting tree being rejected by the AU test statistic, suggesting that the molecular data is consistent with larval morphology and that Listroderini is distinct from Cylydrorhinini.

Although the unconstrained analysis failed to recover any of the remaining 16 tribes of Entiminae as monophyletic, some of these were recovered in two or more well supported clades. Therefore, within the Tropiphorini, *Tropiphorus carinatus*, and *T. bertolini* form one clade (98% BS), *Malvinius* (three species) forms another (99% BS), with the remaining four genera (and species) of Tropiphorini distributed across the tree with low support. In the Celeuthetini, *Cnemidothrix*, *Levoecus*, and *Sphaerorhinus* form a clade (90% BS), as do *Coptorhynchus* and *Heteroglymma* (99% BS). With the addition of *Samobius* and *Platysimus*, all seven aforementioned genera form a clade, but with low support (14% BS); the remaining genus of Celeuthetini, *Phraotes*, is recovered away from this last clade with one moderately supported (85% BS) intervening node that groups it with members of the tribes Rhyncogonini and Elytrurini. Whilst such clades with moderate and high nodal support appear to offer evidence for the paraphyly of several tribes, the generally low nodal supports in the intermediate nodes between such clades preclude conclusions to be drawn based on bootstrap values alone.

4.2. Constraint Analyses and Statistical Tests of Monophyly

In supplement to the bootstrap support results, the AU tests rejecting the three ML trees respectively containing the constrained monophyly of the tribes—Otiorhynchini, Brachyderini, and Tropiphorini—provide further evidence for the paraphyly of these higher taxa.

Otiorhynchini is a particularly species-rich tribe containing 10 genera, of which the *Otiorhynchus* 'complex' contains about 1500 species exclusive to the Palaearctic region (except for a few introduced species in the Nearctic) which have been divided into 105 subgenera [41]. No detailed phylogenetic analysis has been undertaken within this group, although a karyotype analysis of three genera was in accordance with the current classification [41]. The taxa analysed in this study belong to five subgenera: *O.* (*Otiorhynchus*) *armadillo*, *O.* (*Postaremus*) *nodosus*, *O.* (*Dorymerus*) *sulcatus*, *O.* (*Nihus*) *globulus*, and *O.* (*Zustalestus*) *rugosostriatus* [39]. Four of these species (*O. globulus*, *O. sulcatus*, *O. rugosostriatus*, and *Otiorhynchus* sp.) were retrieved in a clade in the unconstrained ML analysis (69% BS), with a high support for the sister relationship between *O. sulcatus* and *Otiorhynchus* sp. (100% BS). Of the two remaining species, *O. nodosus* was retrieved with high support as sister to *Strophosoma melanogrammum*, belonging to the tribe Brachyderini (98% BS), and *O. armadillo* was weakly supported as a lineage sister to a clade containing the first group of four *Otiorhynchus* + two members of Tropiphorini (two *Tropiphorus* spp.) and one Hipporhinini (*Bronchus* sp.). It is difficult to be confident about the relationships amongst these Otiorhynchini, and the retrieval of *O. nodosus* sister to *S. melanogrammum* is particularly surprising. Sequences for *cox1* for these last two species were obtained from GenBank, and both originated from the same study investigating clonality and polyploidy in *Otiorhynchus* [40]. A BLAST search against the GenBank database revealed that the *S. melanogrammum cox1* sequence very closely matches sequences from four *Otiorhynchus* species in the same study (98–99% identity over 100% of the 552 bp sequence; E = 0.0) indicating a close relationship between these two genera. It is unlikely that the sample was mislabelled on GenBank, although this cannot be ruled out with certainty. *Strophosoma melanogrammum* is also represented in the present data matrix by a partial mitogenome sequence, lacking both *cox1* and *rrnL* [11], and not recovered together with the GenBank sequence represented *S. melanogrammum*, but in another clade containing three other Brachyderini taxa (*Brachyderes* spp.), most likely explaining the relationship with *Otiorhynchus* described above being driven by the closest-matching *cox1* sequence. A previous molecular analysis, based on a smaller dataset than the present, resulted in the recovery of monophyletic Otiorhynchini and Brachyderini [20].

5. Conclusions

The approach used here has confirmed the utility of combining shorter sequences into a longer alignment insofar as several interesting relationships were identified, both supporting and rejecting monophyly of currently classified higher taxa. The extent to which meaningful conclusions can be made regarding how accurately shorter sequences are able to match to their correct lineages is undoubtedly a function of the depth of taxon coverage in the backbone mitogenome alignment, from which most of

the phylogenetic signal is derived. The mitogenome dataset contained members of less than a third (19 out of 63) of the tribes within the broad-nosed weevils, so it is hardly surprising that nodal BS support for many internal nodes within this group were poorly supported with the addition of taxa represented by single mitochondrial genes from GenBank. This is a direct result of the small amount of shared comparative data for calculating BS support between taxa with long mitogenome sequences and the taxa solely represented by short sequences.

The inability to reject several of the apparently paraphyletic clades through constraint analyses highlights the presence of conflicting or insufficient data, and demonstrates the complex systematics of the Curculionoidea, wherein particular genera cannot confidently be ascribed to even a particular subfamily. Other limitations in this study included the use of taxa incompletely identified only to the level of subfamily, therefore not allowing for possible further scrutiny of tribal- or generic-level relationships. Additionally, several sequences from the mitogenome dataset lacked the *cox1* and *rrnL* genes [12], confounding their utility here to act as 'backbone' sequences due to the missing data for the critical loci. Alternative or additional mitochondrial loci, such as *cytB* and *cox2* that have been used in the phylogeny of Coleoptera, could have been also incorporated in the alignment which may have increased the number of taxa available for study. Another potential limitation with this approach is that taxonomic coverage within the public databases is currently rather patchy, being dependent upon a multitude of sources such that in many cases certain higher taxa are represented by a small number of potentially highly aberrant or localised species e.g., most of the Cyclominae obtained from GenBank stemmed from a single study based on the fauna of the Falkland Islands [42]. Additionally, a potential general criticism of mitogenome data, despite its consistency with bifurcating phylogenetic trees [43] owing to its maternal inheritance and its unambiguous orthology [44], is that phylogenetic analyses may be confounded by inconsistencies of the coalescent history.

Whilst some results obtained here are cautionary in highlighting the necessity for the careful use of publicly available sequences, it has been demonstrated that it is possible to both single out interesting relationships that warrant further investigation and to test for monophyly, whilst attempting to maximise taxon sampling. One avenue of possible investigation for reconstructing supra-specific phylogenies may involve the use or concatenation of several congeneric GenBank-obtained sequences to represent genus-level or higher taxa, rather than relying only on conspecific sequences, as used here. This may be particularly useful where inter-generic limits may already be well established a priori for such taxa.

Supplementary Materials: The following are available online at http://www.mdpi.com/1424-2818/10/2/21/s1. Figure S1: Mitogenome-derived 'backbone' ML tree. Figure S2: Unconstrained ML tree of combined mitogenome and single-locus sequences. Table S1: List of single-locus sequences analysed; Table S2: List of mitogenome sequences analysed. File S1: Mitogenome DNA sequence data analysed [11].

Acknowledgments: The authors are indebted to Karen Siu-Ting (NUI Maynooth) for her assistance with some analyses. This work was supported by a Natural Environment Research Council CASE PhD studentship (NE/G01194X/1) to C.P.D.T.G., the University of East Anglia, and The Natural History Museum.

Author Contributions: C.P.D.T.G., A.P.V., C.H.L., and B.C.E. conceived and designed the project; C.P.D.T.G. obtained the data, and performed the experiments and analyses; B.C.E. A.P.V., and C.H.L. contributed reagents/materials/analysis tools; C.P.D.T.G., A.P.V., C.H.L., and B.C.E. wrote the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Harrison, C.J.; Langdale, J.A. A step by step guide to phylogeny reconstruction. *Plant J.* **2006**, *45*, 561–572. [CrossRef] [PubMed]
2. Wiley, E.O.; Lieberman, B.S. *Phylogenetics: Theory and Practice of Phylogenetic Systematics*, 2nd ed.; Wiley-Blackwell: Hoboken, NY, USA, 2011; pp. 1–432, ISBN 978-0-470-90596-8.
3. Franz, N.M.; Engel, M.S. Can higher-level phylogenies of weevils explain their evolutionary success? A critical review. *Syst. Entoml.* **2010**, *35*, 597–606. [CrossRef]

4. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2013**, *41*, D36–D42. [CrossRef] [PubMed]
5. Ratnasingham, S.; Hebert, P.D.N. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol. Ecol. Notes* **2007**, *7*, 355–364. [CrossRef] [PubMed]
6. Shimodaira, H.; Hasegawa, M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **1999**, *16*, 1114–1116. [CrossRef]
7. Kishino, H.; Hasegawa, M. Evaluation of the maximum-likelihood estimate of the evolutionary tree topologies from DNA-sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **1989**, *29*, 170–179. [CrossRef] [PubMed]
8. Schmidt, H.A. Testing tree topologies. In *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd ed.; Lemey, P., Salemi, M., Vandamme, A.M., Eds.; Cambridge University Press: Cambridge, UK, 2009; pp. 381–404. ISBN 9780521730716.
9. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **2002**, *51*, 492–508. [CrossRef] [PubMed]
10. Jordal, B.H.; Smith, S.M.; Cognato, A.I. Classification of weevils as a data-driven science: Leaving opinion behind. *Zookeys* **2014**, *439*, 1–18. [CrossRef] [PubMed]
11. Gillett, C.P.D.T.; Crampton-Platt, A.; Timmermans, M.J.; Jordal, B.H.; Emerson, B.C.; Vogler, A.P. Bulk *de novo* mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Mol. Biol. Evol.* **2014**, *31*, 2223–2237. [CrossRef] [PubMed]
12. Hernández-Vera, G.; Caldara, R.; Tosevski, I.; Emerson, B.C. Molecular phylogenetic analysis of archival tissue reveals the origin of a disjunct southern African-Palaearctic weevil radiation. *J. Biogeogr.* **2013**, *40*, 1348–1359. [CrossRef]
13. Lacordaire, T. *Histoire Naturelle des Insectes. Genera des Coléoptères, ou Exposé Méthodique et Critique de Tous Les Genres Proposés Jusqu'ici Dans cet Ordre D'insectes*; pp Tome sixième contenant la famille des curculionides; Librairie Encyclopédique de Roret: Paris, France, 1863; pp. 1–637.
14. Thompson, R.T. Observations on the morphology and classification of weevils (Coleoptera, Curculionoidea) with a key to major groups. *J. Nat. Hist.* **1992**, *26*, 835–891. [CrossRef]
15. Velazquez De Castro, A.J.; Alonso-Zarazaga, M.A.; Outerelo, R. Systematics of Sitonini (Coleoptera: Curculionidae: Entiminae), with a hypothesis on the evolution of feeding habits. *Syst. Entomol.* **2007**, *32*, 312–331. [CrossRef]
16. Oberprieler, R.G.; Marvaldi, A.E.; Anderson, R.S. Weevils, weevils, weevils everywhere. *Zootaxa* **2007**, *1668*, 491–520. [CrossRef]
17. Kuschel, G. A phylogenetic classification of Curculionoidea to families and subfamilies. *Mem. Entomol. Soc. Wash.* **1995**, *14*, 5–33.
18. Marvaldi, A.E. Higher level phylogeny of Curculionidae (Coleoptera: Curculionoidea) based mainly on larval characters, with special reference to broad-nosed weevils. *Cladistics* **1997**, *13*, 285–312. [CrossRef]
19. Bouchard, P.; Bousquet, Y.; Davies, A.E.; Alonso-Zarazaga, M.A.; Lawrence, J.F.; Lyal, C.H.C.; Newton, A.F.; Reid, C.A.M.; Schmitt, M.; Ślipiński, S.A.; et al. Family-group names in Coleoptera (Insecta). *Zookeys* **2011**, *88*, 1–972. [CrossRef] [PubMed]
20. Hundsdoerfer, A.K.; Rheinheimer, J.; Wink, M. Towards the phylogeny of the Curculionoidea (Coleoptera): Reconstructions from mitochondrial and nuclear ribosomal DNA sequences. *Zool. Anz.* **2009**, *248*, 9–31. [CrossRef]
21. McKenna, D.D.; Sequeira, A.S.; Marvaldi, A.E.; Farrell, B.D. Temporal lags and overlap in the diversification of weevils and flowering plants. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7083–7088. [CrossRef] [PubMed]
22. Haran, J.; Timmermans, M.J.T.N.; Vogler, A.P. Mitogenome sequences stabilize the phylogenetics of weevils (Curculionoidea) and establish the monophyly of larval ectophagy. *Mol. Phylogenet. Evol.* **2013**, *67*, 156–166. [CrossRef] [PubMed]
23. Gunter, N.L.; Oberprieler, R.G.; Cameron, S.L. Molecular phylogenetics of Australian weevils (Coleoptera: Curculionoidea): Exploring relationships in a hyperdiverse lineage through comparison of independent analyses. *Austral Entomol.* **2015**, *55*, 217–233. [CrossRef]
24. Shin, S.; Clarke, D.J.; Lemmon, A.R.; Aitken, A.L.; Haddad, S.; Farrell, B.D.; Marvaldi, A.E.; Oberprieler, R.G.; McKenna, D.D. Phylogenomic data yield new and robust insights into the phylogeny and evolution of weevils. *Mol. Biol. Evol.* **2017**. [CrossRef] [PubMed]

25. Hebert, P.D.N.; Cywinska, A.; Ball, S.L.; DeWaard, J.R. Biological identifications through DNA barcodes. *Proc. Biol. Sci.* **2003**, *270*, 313–321. [CrossRef] [PubMed]
26. Alonso-Zarazaga, M.A.; Lyal, C.H.C. *A World Catalogue of Families and Genera of Curculionoidea (Insecta: Coleoptera) (Excepting Scolytidae and Platypodidae)*; Entomopraxis: Barcelona, Spain, 1999; pp. 1–315, ISBN 84-605-9994-9.
27. Hunt, T.; Bergsten, J.; Levkanicova, Z.; Papadopoulou, A.; John, O.S.; Wild, R.; Hammond, P.M.; Ahrens, D.; Balke, M.; Caterino, M.S.; et al. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **2007**, *318*, 1913–1916. [CrossRef] [PubMed]
28. Hunt, T.; Vogler, A.P. A protocol for large-scale rRNA sequence analysis: Towards a detailed phylogeny of Coleoptera. *Mol. Phylogenet. Evol.* **2008**, *47*, 289–301. [CrossRef] [PubMed]
29. Bocak, L.; Barton, C.; Crampton-Platt, A.; Chesters, D.; Ahrens, D.; Vogler, A.P.V. Building the Coleoptera tree-of-life for >8000 species: Composition of public DNA data and fit with Linnaean classification. *Syst. Entomol.* **2013**, *39*, 97–110. [CrossRef]
30. Oberprieler, R.G. A reclassification of the weevil subfamily Cyclominae (Coleoptera: Curculionidae). *Zootaxa* **2010**, *2515*, 1–35. [CrossRef]
31. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef] [PubMed]
32. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef] [PubMed]
33. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [CrossRef] [PubMed]
34. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010; pp. 1–8.
35. Huson, D.H.; Scornavacca, C. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Syst. Biol.* **2012**, *61*, 1061–1067. [CrossRef] [PubMed]
36. Shimodaira, H.; Hasegawa, M. CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* **2001**, *17*, 1246–1247. [CrossRef] [PubMed]
37. Pierotti, H.; Fink, T. New and interesting Peritelini of the Western Mediterranean fauna. XX. A novel *Meira* (Jacquelin du Val, 1852) species from the Ligurian Alps. *Zootaxa* **2013**, *3716*, 595–598. [CrossRef] [PubMed]
38. Pierotti, H.; Germann, C.; Braunert, C. New or interesting Peritelini of the West-Mediterranean fauna. XXIV. Two new Simmeiropsis Pierotti & Bello, 2013 from Portugal (Coleoptera, Curculionidae, Entiminae). *Zootaxa* **2013**, *3734*, 273–280. [CrossRef] [PubMed]
39. Pierotti, H.; Bello, C.; Alonso-Zarazaga, M.A. Contribution to the systematic rearrangement of the Palaearctic Peritelini. VI. A synthesis of the Spanish Peritelini (Coleoptera: Curculionidae: Entiminae). *Zootaxa* **2010**, *2376*, 1–96.
40. Marvaldi, A.E. Larvae of South American Entimini (Coleoptera: Curculionidae), and phylogenetic implications of certain characters. *Rev. Chil. Entomol.* **1998**, *25*, 21–44.
41. Lachowska, D.; Rozek, M.; Holecova, M. Cytotaxonomy and karyology of the tribe Otiorhynchini (Coleoptera: Curculionidae). *Eur. J. Entomol.* **2008**, *105*, 175–184. [CrossRef]
42. Papadopoulou, A.; Jones, A.G.; Hammond, P.M.; Vogler, A.P. DNA taxonomy and phylogeography of beetles of the Falkland Islands (Islas Malvinas). *Mol. Phylogenet. Evol.* **2009**, *53*, 935–947. [CrossRef] [PubMed]
43. Curole, J.P.; Kocher, T.D. Mitogenomics: Digging deeper with complete mitochondrial genomes. *Trends Ecol. Evol.* **1999**, *14*, 394–398. [CrossRef]
44. Botero-Castro, F.; Tilak, M.K.; Justy, F.; Catzeflis, F.; Delsuc, F.; Douzery, E.J.P. Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). *Mol. Phylogenet. Evol.* **2013**, *69*, 728–739. [CrossRef] [PubMed]