Systematic Entomology

Systematic Entomology (2017), **42**, 60–81 DOI: 10.1111/syen.12210

A molecular phylogeny and revised higher-level classification for the leaf-mining moth family Gracillariidae and its implications for larval host-use evolution

AKITO Y. KAWAHARA[1], DAVID PLOTKIN[1,2], ISSEI OHSHIMA[3], CARLOS LOPEZ-VAAMONDE[4,5], PETER R. HOULIHAN[1,6], JESSE W. BREINHOLT[1], ATSUSHI KAWAKITA[7], LEI XIAO[1], JEROME C. REGIER[8,9], DONALD R. DAVIS[10], TOSIO KUMATA[11], JAE-CHEON SOHN[9,10,12], JURATE DE PRINS[13] and CHARLES MITTER[9]

[1]Florida Museum of Natural History, University of Florida, Gainesville, FL, U.S.A., [2]Department of Entomology and Nematology, University of Florida, Gainesville, FL, U.S.A., [3]Kyoto Prefectural University, Kyoto, Japan, [4]INRA, UR0633 Zoologie Forestière, Orléans, France, [5]IRBI, UMR 7261, CNRS/Université François-Rabelais de Tours, Tours, France, [6]Department of Biology, University of Florida, Gainesville, FL, U.S.A., [7]Center for Ecological Research, Kyoto University, Kyoto, Japan, [8]Institute for Bioscience and Biotechnology Research, University of Maryland, College Park, MD, U.S.A., [9]Department of Entomology, University of Maryland, College Park, MD, U.S.A., [10]Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC, U.S.A., [11]Hokkaido University Museum, Sapporo, Japan, [12]Department of Environmental Education, Mokpo National University, Muan, South Korea and [13]Department of Entomology, Royal Belgian Institute of Natural Sciences, Brussels, Belgium

Abstract. Gracillariidae are one of the most diverse families of internally feeding insects, and many species are economically important. Study of this family has been hampered by lack of a robust and comprehensive phylogeny. In the present paper, we sequenced up to 22 genes in 96 gracillariid species, representing all previously recognized subfamilies and genus groups, plus 20 outgroups representing other families and superfamilies. Following objective identification and removal of two rogue taxa, two datasets were constructed: dataset 1, which included 12 loci totalling 9927 bp for 94 taxa, and dataset 2, which supplemented dataset 1 with 10 additional loci for 10 taxa, for a total of 22 loci and 16 167 bp. Maximum likelihood analyses strongly supported the monophyly of Gracillariidae and most previously recognized subfamilies and genus groups. On this basis, we propose a new classification consisting of eight subfamilies, four of which are newly recognized or resurrected: Acrocercopinae Kawahara & Ohshima **subfam. n.**; Gracillariinae Stainton; Lithocolletinae Stainton; Marmarinae Kawahara & Ohshima **subfam. n.**; Oecophyllembiinae Réal & Balachowsky; Parornichinae Kawahara & Ohshima **subfam. n.**; Ornixolinae Kuznetzov & Baryshnikova **stat. rev.**; and Phyllocnistinae Zeller. The subfamily Gracillariinae is restricted to the monophyletic group comprising *Gracillaria* Haworth and closely related genera. We also formally transfer *Acrocercops scriptulata* Meyrick to Ornixolinae and use the name *Diphtheroptila* Vári, creating *Diphtheroptila scriptulata* **comb. n.** An exploratory mapping of larval host-use traits on the phylogeny shows strong conservation of modes of leaf mining but much higher lability of associations with host plant orders and families, suggesting that host shifts could play a significant role in gracillariid diversification.

This published work has been registered in ZooBank, http://zoobank.org/urn:lsid:zoobank.org:pub:942814A2-DE66-41D4-8AB6-FF0B18C87EDB.

Correspondence: Akito Y. Kawahara, Florida Museum of Natural History, McGuire Center for Lepidoptera and Biodiversity, University of Florida, 3215 Hull Road, Gainesville, FL 32611, U.S.A. E-mail: kawahara@flmnh.ufl.edu

Introduction

Leaf-mining moths in the family Gracillariidae constitute one of the most diverse groups of internal-feeding Lepidoptera, including nearly 2000 described species in more than 100 genera (van Nieukerken *et al.*, 2011; De Prins & De Prins, 2016) with many more yet to be described (Lees *et al.*, 2014). Larvae of most Gracillariidae create serpentine and/or blotch mines in leaves, but some species instead mine/bore fruit, mine stems, roll leaves, or make galls (Davis, 1987; De Prins & De Prins, 2016). Some are obligate pollination mutualists (Kawakita *et al.*, 2004, 2010), while others can delay senescence of abscised leaves (Giron *et al.*, 2007; Kaiser *et al.*, 2010; Gutzwiller *et al.*, 2015). Individual species are mostly monophagous or oligophagous, but the family feeds collectively on an extraordinary diversity of host plants on continents (De Prins & De Prins, 2016) and remote island archipelagos (Zimmerman, 1978). Some gracillariids are invasive and cause substantial economic damage as introduced pests (Heppner, 1993; Gilbert *et al.*, 2005; Shapiro *et al.*, 2008; Lopez-Vaamonde *et al.*, 2010). Other species have been used as models in the study of insect–plant interactions (Lopez-Vaamonde *et al.*, 2006; Ohshima, 2012). Gracillariid larvae undergo spectacular ontogenetic and behavioural changes in morphology and feeding behaviour – the number of instars can vary from four to 11 depending on the species (Davis, 1987). Larvae are unique within Lepidoptera in that some transition from a sap-feeding form (with a flattened head and body, modified mandibles, absence of a functional spinneret, and absence of legs), to a strikingly different tissue-feeding form (with a cylindrical body, round head, chewing mouthparts, legs, and a functional spinneret) that resembles a typical lepidopteran larva (Kumata, 1978; Davis, 1987; Wagner *et al.*, 2000; Body *et al.*, 2015). Some larvae also have a transitional quiescent instar that does not feed (Kumata, 1978; Davis, 1987; Wagner *et al.*, 2000). The marked morphological and behavioural transitions displayed by gracillariids are often regarded to be hypermetamorphic.

To understand the origins of these unique adaptations, and to provide a sound classification supporting the basic and applied study of Gracillariidae, a robust and comprehensive phylogeny is needed. At present, however, our knowledge of gracillariid phylogeny is very incomplete. The present study builds on previous morphology-based (Fig. 1A) and molecular-based (Fig. 1B, C) classifications and reconstructions of gracillariid phylogeny.

Synthetic higher-level morphological work on Gracillariidae began in the early 1900s. Spuler (1910) was the first to divide the family into two subfamilies: Gracillariinae and Lithocolletinae. Vári (1961) examined the wing venation of many gracillariid genera and used venation characters to classify them, providing an initial framework for subsequent gracillariid classifications. Kuznetzov & Stekol'nikov (1987) examined the musculature of the male genitalia of Russian gracillariids, and proposed that the Gracillariinae and Lithocolletinae form a monophyletic group, which is in turn a sister group to Parornichinae (consisting of the genera *Callisto* Stephens and *Parornix* Spuler). Kuznetzov & Stekol'nikov (1987) treated Phyllocnistinae as a separate family, 'Phyllocnistidae', and hypothesized that it was the sister group to taxa currently placed in the Gracillariidae (Fig. 1A), albeit without explicit cladistic analysis. Kumata *et al.* (1988a,1988b) recognized three subfamilies (Gracillariinae, Lithocolletinae, Phyllocnistinae) and also applied informal names to four groups of genera in the Gracillariinae, namely the *Acrocercops*, *Gracillaria*, *Parectopa* and *Parornix* groups.

Kawahara *et al.* (2011) generated a molecular phylogeny based on the largest gracillariid dataset to that point (39 species and up to 21 protein-coding nuclear genes), and tested Kumata's hypotheses. Their analysis showed that the Gracillariidae are monophyletic, that most species can be placed in one of the subfamilies, and that most Gracillariinae can be placed in one of Kumata's genus groups, with high confidence (Fig. 1C). The sample of gracillariids (22 taxa) in the broad lepidopteran study of Regier *et al.* (2013) yielded similar conclusions (Fig. 1B). In both of these studies, however, taxon sampling was limited, and enigmatic genera of uncertain placement (e.g. *Callicercops* Vári, *Ornixola* Kuznetzov, etc.) were not included. The complexity and diversity of Gracillariidae have hampered clarification of its internal relationships, and the placement of many genera remains unknown.

We present here the most extensive sequence-based phylogenetic analysis of gracillariid genera thus far, with more than double the taxon sample size of the largest previous study (Kawahara *et al.*, 2011). We targeted taxa that span the morphological diversity within the family, type genera and species, and taxa that have been historically challenging to place. We use this new dataset to test the monophyly of Gracillariidae and its subfamilies and propose a revised higher classification. Using our phylogeny, we also provide an initial overview of evolutionary patterns in larval host plant use.

Methods

Taxon and gene sampling

We sampled as many Gracillariidae genera as possible in order to construct a robust phylogeny for the family and test prior hypotheses on their relationships. In total, 96 species in 59 genera were included as ingroup taxa, and 20 species were included as outgroups. Two taxa were removed after rogue taxon analysis (see later), leaving a total of 94 species in 58 genera. We initially included 11 nuclear genes and one mitochondrial gene that were known to amplify well and provide strong phylogenetic signal, based on previous phylogenetic studies of leaf miners and relatives (e.g. Kawahara *et al.*, 2011; Sohn *et al.*, 2013). We constructed two datasets: (i) 12 genes totalling 9927 bp sampled for 94 gracillariid species (dataset 1); and (ii) the 12 genes in dataset 1 plus 10 genes for 10 distantly related gracillariid taxa – thus 22 genes for 10 taxa plus 12 genes for 86 taxa, totalling 16 167 bp (dataset 2). Data for the latter 10 genes added to dataset 2 were originally used in the phylogenetic analyses of Kawahara *et al.* (2011). Dataset 2 included more genes but also had a larger amount of missing data than dataset 1. Although adding data for a diverse subset of taxa can help

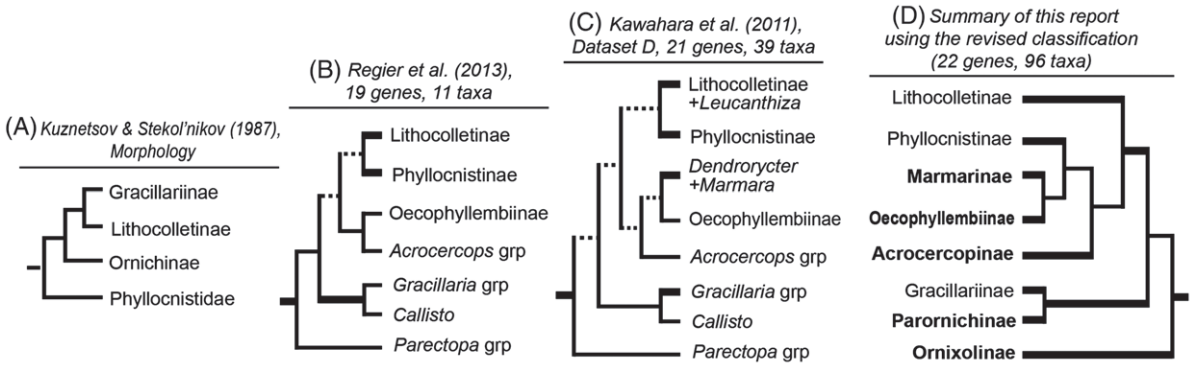

Fig. 1. Summaries of hypotheses of higher-level phylogenetic relationships within Gracillariidae. (A) Higher classification based on morphology (Kuznetzov & Stekol'nikov, 1987). (B) Higher classification based on a large-scale molecular phylogeny of the entire order Lepidoptera (Regier *et al.*, 2013). (C) Partial redrawing of Fig. 2 from Kawahara *et al.* (2011), dataset D. (D) Tree based on our maximum likelihood topology generated in the current study using dataset 2, nt123, with rogue taxon set 1 removed. For (B)–(D), line dashing and line thickness indicate bootstrap support that was present in both degen1 and nt123 analyses. Thick lines indicate bootstrap support ≥80% in both degen1 and nt123 analyses, thin lines indicate support of 50–79%, and dashed lines indicate groups that are favoured but that received <50% support. Nomenclature in C corresponds to that of Kumata (1998) and Davis & Robinson (1998), as used in Kawahara *et al.* (2011). Nomenclature in D is based on the current study (name changes in bold).

deep phylogeny estimation, adding sequence data can also have misleading results if those data are only available for a subset of taxa (e.g. de la Torre-Bárcena *et al.*, 2009; Philippe *et al.*, 2011; Roure *et al.*, 2013). Thus, we analysed the two datasets in order to more accurately assess the impacts of increased gene sampling and missing data on our results.

Total sequence lengths and names of the 12 genes in dataset 1 are: a trimmed 46F–1028R region of *pyrimidine biosynthesis* (*CAD*; 2886 bp), the 1.7sF–4sR region of *dopa decarboxylase* (*DDC*; 708 bp), the 28LF–406R region of *enolase* (1135 bp), the 2F–4R region of *acetyl-coA carboxylase* (*ACC*; 501 bp), the 1F-2R region of *gelsolin* (*109fin*; 561 bp), the 2F–3R region of *histidyl-tRNA synthetase* (*265fin*; 447 bp), the 1F–2R region of *AMP deaminase* (*268fin*; 768 bp), the 1F–2R region of *glucose phosphate dehydrogenase* (*3007fin*; 620 bp), the Wg1aF–Wg2aR region of *wingless* (402 bp), a trimmed region of the 30F–41.21R region of *elongation factor-1α* (*ef1α*; 990 bp), *histone 3* (*H3*; 273 bp), and a trimmed 'barcode' region of *cytochrome oxidase 1* (*CO1*; 657 bp). Dataset 2 includes the following additional genes: a trimmed 177sF–532sR region of *period* (1014 bp), a trimmed region of *phosphogluconate dehydrogenase* (*40fin*; 750 bp), a trimmed *putative GTP-binding protein* (*42fin*; 840 bp), a trimmed *glutamyl- & prolyl-tRNA synthetase* (*192fin*; 402 bp), *triosephosphate isomerase* (*197fin*; 444 bp), trimmed *proteasome subunit* (*262fin*; 501 bp), the 1F-2R region of *tetrahydrofolate synthase* (*3017fin*; 594 bp), the 4F-5R region of *alanyl-tRNA synthetase* (*3070fin*; 705 bp), the *nucleolar cysteine-rich protein* (*8028fin*; 324 bp), and the 1F–2R region of *glucose phosphate isomerase* (*8091fin*; 666 bp). These gene names follow Regier (2008) and GenBank numbers for each sequence are listed in Table S1.

Sequencing, alignment, and contamination

Polymerase chain reaction (PCR) primers, amplification strategies, and laboratory protocols largely followed Kawahara *et al.* (2011). For nearly all genes, sequences were initially generated from mRNAs amplified by reverse transcription PCR. Amplicons were then gel-isolated and purified. Nested amplifications were conducted whenever necessary using internal primers outlined in Regier (2008). Contig assemblies were visually checked for errors before creating a consensus for each species with the software GENEIOUS 8.0 (Biomatters Ltd, Auckland, New Zealand). Each consensus was edited and checked for base call error. All single-gene datasets began with the first codon position (nt1) and were separately aligned using MAFFT 7.703 (Katoh & Standley, 2013), implementing the E-INS-i option (mafft–genafpair maxiterate 1000). Alignments were visually inspected and checked for frame-shifts and the presence of termination codons. Sequences were also assessed for contamination and sample-switching error by generating maximum likelihood (ML) bootstrap trees in RAXML (Stamatakis, 2014) for each gene (see below for ML tree-building methods). The final dataset was concatenated in GENEIOUS from aligned single gene datasets. We have deposited the concatenated alignment in the Dryad Data Repository (http://www.datadryad.org, study accession number doi:10.5061/dryad.j316c). Specimen voucher data, sequences, and images can be found in the BOLD dataset, DS-PHYGRAC (dx.doi.org/10.5883/DS-PHYGRAC).

Datasets and character partitioning

Sequence evolution varies among codon positions, and synonymous and nonsynonymous changes can produce potentially conflicting and misleading results if not analysed properly (e.g. Regier *et al.*, 2008, 2009; Cho *et al.*, 2011). To see how synonymous and nonsynonymous signals affect our results at different phylogenetic levels, we first analysed datasets 1 and 2 using all nucleotide substitutions (nt123) and subsequently analysed them using nonsynonymous change only (degen1, Regier *et al.*, 2010; Zwick, 2010). Degen1 datasets were created by running the

Table 1. Results of approximately unbiased (AU) significance tests for nonmonophyly of predicted clades on dataset 2.

	Group	AU score	
		degen1	nt123
1	Oecophyllembiinae + Phyllocnistinae (OP)[a]	0.062	0.063
2	Lithocolletinae + Phyllocnistinae (LP)[b]	**<0.001**	**<0.001**
3	Marmarinae + Phyllocnistinae (MP)	0.618	0.737
4	Marmarinae + Phyllocnistinae + Parornichinae (MP + R)	0.515	0.340
5	Acrocercopinae + Gracillariinae + Marmarinae + Ornixolinae + Lithocolletinae (AGMX + L)	**0.001**	**<0.001**
6	Acrocercopinae + Gracillariinae + Marmarinae + Ornixolinae + Parornichinae (AGMX + R)[c]	**0.005**	**<0.001**
7	Acrocercopinae + Gracillariinae + Marmarinae + Ornixolinae + Parornichinae + Lithocolletinae (AGMX + R + L)[d]	**0.008**	**<0.001**
8	Acrocercopinae + Gracillariinae + Marmarinae + Ornixolinae + Oecophyllembiinae + Parornichinae (AGMX + O + R)[e]	>0.999	>0.999
9	Lithocolletinae minus *Leucanthiza*	**<0.001**	**<0.001**

Statistically significant results ($P < 0.05$) are shown in bold.
[a]Oecophyllembiinae + Phyllocnistinae *sensu* Kumata (1998).
[b]Lithocolletinae *sensu* Clemens (1859).
[c]Gracillariinae *sensu* Stainton (1854).
[d]Gracillariinae *sensu* Kuznetzov & Stekol'nikov (1987).
[e]Gracillariinae *sensu* Davis & Robinson (1998).
A, Acrocercopinae; G, Gracillariinae; L, Lithocolletinae; M, Marmarinae; O, Oecophyllembiinae; P, Phyllocnistinae; R, Parornichinae; X, Ornixolinae.

degen1.pl Perl script of Zwick (2010). On average, synonymous change occurs more rapidly, leading to multiple substitutions per site and nonhomogeneous base composition. This can degrade phylogenetic signal, especially for deep divergences, although it can be informative at the tips of the tree. For example, the major taxonomic finding in Cho *et al.* (2011), the identification of Gracillarioidea + Yponomeutoidea as sister group to all other non-tineoid Ditrysia, only received strong support from nonsynonymous change.

We used PARTITIONFINDER v1.1.1 (Lanfear *et al.*, 2012) to determine the best character partitions and nucleotide substitution models for the aligned datasets. Because all datasets were subsequently analysed using RAXML, model selection was limited to the substitution models available in RAXML (e.g. GTR + G, GTR + I + G). The best substitution models for each dataset, utilized in RAXML, are listed in Table S2.

Phylogenetic analyses

Phylogenetic analyses were conducted with ML methods as implemented in RAXML v.8.0 (Stamatakis, 2014) for the nt123 and degen1 versions of datasets 1 and 2. For each dataset, the RAXML run consisted of a 1000-pseudoreplicate bootstrap analysis followed by a search for the best-scoring tree, incorporating the best partitioning scheme obtained from PARTITIONFINDER. In order to control for the use of different programs in assessing the effects of differences in taxon and gene sampling between this study and a prior study with fewer taxa and genes (Kawahara *et al.*, 2011), we also conducted 2000 ML tree searches in GARLI (Zwickl, 2006) with the same settings as those of Kawahara *et al.* (2011). Trees were visualized using FIGTREE v.1.4.2 (Rambaut, 2014). All phylogenetic analyses were conducted on the University of Florida High Performance Computing Cluster (http://www.hpc.ufl.edu/).

Rogue taxon analyses

Bootstrap values for some deeper nodes remained low regardless of the dataset analysed. One possible cause of low support is the sensitivity of bootstrap values to taxa of unstable placement (Sanderson & Shaffer, 2002). We investigated the potential contribution of rogue taxa to low bootstrap values in our dataset using the ROGUENAROK (RNR) approach of Aberer *et al.* (2013). To identify rogue taxa, we downloaded the code from GitHub (https://github.com/aberer/RogueNaRok) and ran the program locally. We tested for rogue taxa with the nt123, 22-gene dataset, as initial analyses it gave the highest bootstrap support overall. The rogue taxa identified in this test were then removed from each gene file in all datasets, and the trimmed datasets were subsequently concatenated in GENEIOUS and analyzed with RAXML.

Tests of alternative phylogenetic hypotheses

Morphology and larval mining patterns predict the monophyly of Gracillariinae *sensu* Kuznetzov & Stekol'nikov (1987), Gracillariinae *sensu* Davis & Robinson (1998), Oecophyllembiinae + Phyllocnistinae (Kumata, 1998), and a number of other groups. However, nine of these predicted groups, listed in Table 1, were not recovered in our molecular analyses. To test whether these differences between the morphological/behavioural-based tree topologies and molecular-based topologies could be attributable to sampling error in the molecular data, we used the approximately unbiased (AU) test of Shimodaira (2002) on dataset 2. The AU test determines whether trees estimated under a topological constraint describe the data significantly worse than the best-fitting tree. We conducted a separate analysis for each morphological/behaviour grouping contradicted by the molecular data. Topologies corresponding to the constraint of monophyly for

each predicted group were constructed in MESQUITE v.3.01 (Maddison & Maddison, 2014). Each topology was then used as a constraint for a subsequent 100-tree search in RAXML using the default algorithm. Per-site log-likelihood values from the best tree under each alternative hypothesis and the best unconstrained tree were then input and read into CONSEL v.0.20 (Shimodaira & Hasegawa, 2001) to conduct the AU test.

Inference of phylogenetic trends in larval feeding habits

One goal of this study is to increase understanding of the evolution of gracillariid larval host plant use and its potential role in diversification. To gain an initial phylogenetic overview of larval feeding ecology and the relative degree of conservatism of different aspects of host use, we compiled a synopsis of larval host plant use traits in the species sampled and superimposed it on the phylogeny. Host taxon use was tabulated as both plant family and plant order. Host growth form was categorized as trees, shrubs, vines or herbs. Host range was scored as oligophagous (using plants of a single order) versus polyphagous (using two or more plant orders). Finally, we divided variation in the mode of leaf mining into the following categories: blotch mining; tentiform blotch mining; retention of a fully serpentine mine throughout larval development; exiting the mine and rolling the edge of the leaf during the final instar; flower/fruit feeding; and galling. The evolutionary history of each of these traits on the molecular phylogeny was estimated using the unordered parsimony criterion. Data on host plant use traits were largely compiled from De Prins & De Prins (2016).

Results

The best-scoring ML tree is shown in Fig. 2, with bootstrap values for the different combinations of character coding (nt123 and degen1) and gene sampling (datasets 1 and 2) superimposed. In the treatment below, we will repeatedly refer to clades on this tree that are labelled with acronyms (Table 1, Fig. 2). The number of taxa in the current study was more than double that of Kawahara *et al.* (2011) and branch support for relationships among subfamilies, in nearly all cases, was also higher (Figs 1, 2). The most robust result was from the nt123 analysis of the 22-gene dataset (dataset 2), with 73.1% (68/93) of the ingroup nodes having strong bootstrap support (≥80%). Figure 3 shows the same topology in cladogram format, with thickened branches denoting strong bootstrap support (≥80%) from the 22-gene nt123 analysis. Trees from all analyses were mostly congruent; none of the nodes that conflicted with the 22-gene nt123 tree were strongly supported (i.e. bootstrap support was <80%).

Dataset 1 (12 genes) yielded strong bootstrap support for Gracillariidae in both the nt123 (100%) and degen1 (81%) analyses (Table 2). Dataset 2 generally provided stronger support than dataset 1. The nt123 and degen1 analyses resulted in the monophyly of the following subfamilies with strong (≥80% bootstrap) support: Acrocercopinae, Gracillariinae, Lithocolletinae, Marmarinae, Oecophyllembiinae, Parornichinae, and Phyllocnistinae (nodes 4, 7, 10–14; Fig. 2). There were few topological differences between the trees generated from nt123 and degen1 analyses. Ornixolinae was recovered with 97% bootstrap support in the nt123 analyses, but was not monophyletic in the degen1 tree (*Chileoptilia* Vargas & Landry moved in the degen1 tree). However, bootstrap support for this alternative was <50% (Figure S1). The only notable conflict was the node including *Parornix anglicella* (Stainton) and *Parornix fagivora* (Frey), for which bootstrap percentage (BP) = 86% under degen1 but <50% in the nt123 tree. Datasets 1 and 2 yielded lower bootstrap support for degen1 than nt123 for all cross-subfamily non-100% bootstrap nodes for which dataset comparisons could be made (Table 2). The average bootstrap difference for these nine nodes for which a comparison could be made was 11.75% for dataset 1 and 10.57% for dataset 2 (Table 2).

To assess the effect of increased taxon sampling beyond that of Kawahara *et al.* (2011), we conducted a GARLI analysis on a 10-gene, 94-taxon dataset. This dataset included the same 10 loci and 39 taxa from Kawahara *et al.*'s dataset C with an additional 55 taxa from the present study (two rogue taxa being excluded). GARLI was run with the same parameters used in Kawahara *et al.* (2011). The 94-taxon gracillariid analysis resulted in a tree with improved bootstrap support for some key nodes (e.g. Marmarinae, BP = 64%), but lower bootstrap support for other key nodes (e.g. Gracillariidae, BP = 70%, Figure S7).

Our rogue taxon analysis using ROGUENAROK (Aberer *et al.*, 2013) identified nine rogue taxa from the 22-gene nt123 dataset (Table 3). Three of the rogue taxa detected were among the more distant outgroups. Selection of outgroups as rogue might stem from increased uncertainty in position due to lower sampling density among these taxa. Removal of two rogue gracillariid taxa (*Callicercops iridocrossa* (Meyrick), *Gibbovalva kobusi* Kumata & Kuroko; rogue taxon set 1 in Table 3) increased the 'relative bipartition information criterion' (RBIC; Aberer & Stamatakis, 2011; Aberer *et al.*, 2013) from 75.2 to 78.1% (2.9% difference), whereas additional removal of the remaining seven rogue taxa (rogue taxon set 2 in Table 3) only further increased the RBIC by ∼0.3%. Removal of those two rogue taxa in taxon set 1 (Table 3) resulted in increased bootstrap values for 11 nodes and decreased values for eight nodes in the tree (Figure S2). However, most of these increases and decreases were just 1 or 2% changes in bootstrap value. Larger changes in bootstrap value (up to 29%) were observed for six nodes (five increases, one decrease). When rogue taxon sets 1 and 2 were excluded, there were increased bootstrap values for 14 nodes and decreased values for 17, with a total of 21 changes that were greater than 2%. Among the clades undergoing the strongest improvements in support are Gracillariidae (Fig. 2, node 10; BP = 99/82, after/before rogue removal); the Gracillariinae + Parornichinae (Fig. 2, node 6; BP = 83/60); and the *Melanocercops* clade (Fig. 2, node 16; BP = 78/49). Nearly all of the increases in bootstrap values can be explained by deletion of *Callicercops iridocrossa* and *Gibbovalva kobusi* alone (Fig. 2, Figures S3–S5). For this reason we focus on the 94-taxon dataset, from which only these two taxa are deleted.

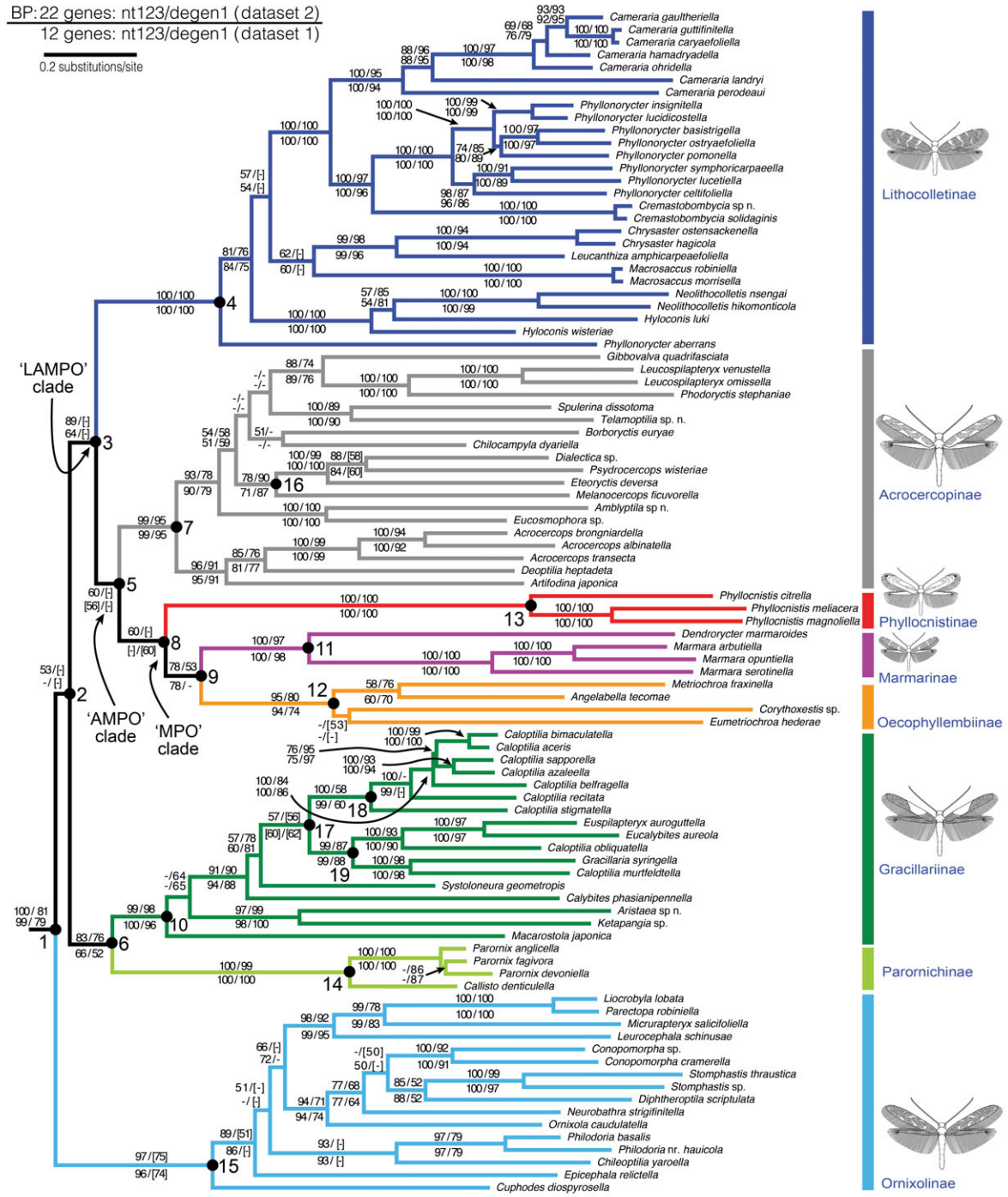

Fig. 2. Maximum likelihood topology found with the 22-gene, 114-taxon dataset (dataset 2), analysed with nt123 showing bootstrap percentages (BPs). The top pair of BPs are separately calculated for nt123 (left of slash) and degen1 (right of slash) datasets. The bottom pair of BPs are separately calculated for a 12-gene, 114-taxon dataset (dataset 1) analysed with nt123 (left of slash) and degen1 (right of slash). Dashes denote bootstrap support <50% and brackets around a BP denote failure to recover that node in the best tree in the corresponding analysis. Higher-level taxonomic names, adopted in this study, are shown on the right. Branch lengths of the phylogram are proportional to total nucleotide substitutions per character as calculated under the 22-gene nt123 model. Relationships of outgroups can be found in Figures S2, S3. [Colour figure can be viewed at wileyonlinelibrary.com].

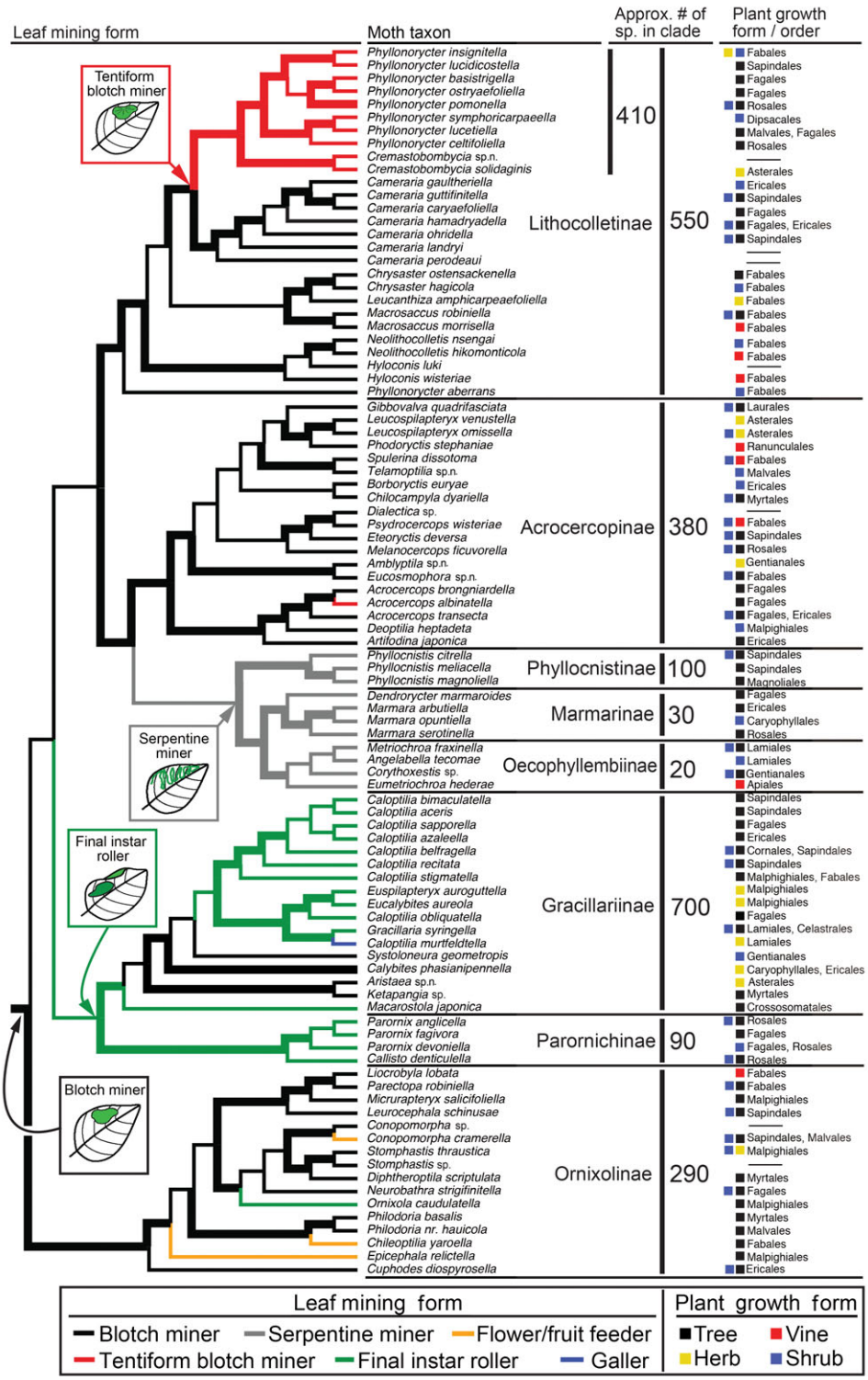

Fig. 3. Synopsis of diversity, host-plant use, and leaf-mining form for the eight gracillariid subfamilies sampled and supported in this study. Leaf-mining form is mapped onto the tree condensed from Fig. 2; branches representing species with unknown leaf-mining forms have been coloured in based on their probable form, as predicted by a parsimony-based ancestral state reconstruction. Thick branches indicate bootstraps ≥80%. For polyphagous species, only the two plant orders with the most records are shown. [Colour figure can be viewed at wileyonlinelibrary.com].

Table 2. Comparison of bootstrap support values for all-nucleotide (nt123) and nonsynonymous only (degen1).

		Dataset 1			Dataset 2		
Node	Taxonomic group	nt123	degen1	Δ	nt123	degen1	Δ
1	Gracillariidae	99	79	+20	100	81	+19
2	Gracillariidae minus Ornixolinae	[<50]	[<50]	n/a	53	[<50]	n/a
3	'LAMPO' clade	64	[< 50]	n/a	89	[<50]	n/a
4	Lithocolletinae	100	100	0	100	100	0
5	'AMPO' clade	[56]	[<50]	n/a	60	[<50]	–
6	'GR' clade	66	52	+14	83	76	+7
7	Acrocercopinae	99	95	+4	99	95	+4
8	'MPO' clade	[<50]	[60]	n/a	60	[<50]	–
9	Marmarinae + Oecophyllembiinae	78	49	+29	78	53	+25
10	Gracillariinae	100	96	+4	99	98	+1
11	Marmarinae	100	98	+2	100	97	+3
12	Oecophyllembiinae	94	74	+20	95	80	+15
13	Phyllocnistinae	100	100	0	100	100	0
14	Parornichinae	100	99	+1	100	100	0
15	Ornixolinae	96	[74]	n/a	97	[75]	n/a
Mean difference		–	–	+11.75	–	–	+10.57
Nodes with 'high' (BP ≥ 80%) support		67	57	–	68	59	–
Percentage of nodes with 'high' support		72.04	61.29	–	73.12	63.44	–

BP, bootstrap percentage; 'high support' refers to BP ≥ 80%. 'Node' refers to corresponding node numbers in Fig. 2. Bootstrap values that were not recorded in the ML tree for that analysis are shown in square brackets. Δ indicates the difference between nt123 and degen1 bootstrap support values.

Table 3. Rogue taxa identified by the ROGUENAROK (RNR) analyses.

Rogue taxon set	SC (%)	Taxon	Raw improvement	RBIC
None	–	–	–	0.751 956
1	18.3	*Callicercops iridocrossa*	1.777	0.767 681
	14.6	*Gibbovalva kobusi*	1.507	0.781 018
2	35.5	*Neurobathra strigifinitella*	0.116	0.782 044
	39.7	*Corythoxestis sp.*	0.08	0.782 752
	82.4	*Hemerophila felis*[a]	0.042	0.783 124
	80.6	*Alucita* sp.[a]	0.031	0.783 398
	84.9	*Urodus decens*[a]	0.035	0.783 708
	67.2	*Epicephala relictella*	0.017	0.783 858
	43.6	*Macarostola japonica*	0.006	0.783 912

[a]Outgroups.
SC, sequence data completeness; RBIC, relative bipartition information criterion.

We tested the ability of the molecular data to significantly reject nine previous hypotheses of higher-level gracillariid relationships that disagreed with the molecular tree (Table 1). Five of these previous hypotheses were strongly rejected with $P \leq 0.008$ for nt123 and $P < 0.001$ for degen1, including the sister-group relationship of Gracillariinae to Lithocolletinae ('AGMX+R+L' clade) proposed by Kuznetzov & Stekol'nikov (1987) (Table 1, group 7). Other groups that were rejected by the AU test include the 'AGMX+L' clade, the 'AGMX+O' clade, and the Lithocolletinae + Phyllocnistinae (LP) clade (Table 1). Finally, our AU test results strongly rejected the inclusion of *Leucanthiza* (Clemens) in Gracillariinae (Table 1, group 9), providing further support for its recent transference from Gracillariinae to Lithocolletinae (De Prins & Kawahara, 2012; De Prins *et al.*, 2013).

Discussion

Effect of taxon addition and rogue taxon deletion on node support

The 39-taxon, 21-gene study of Kawahara *et al.* (2011) provided strong support for some nodes but weak support for many others. Direct comparisons between support levels of the largest dataset in the present study (dataset 2; 96 taxa/22 genes) and the largest dataset in the previous study (dataset D; 57 taxa/21 genes; Kawahara *et al.*, 2011) were complicated because of differences in the number of genes, the distribution of missing data, and the phylogenetic programs used. Therefore, we compared the nt123, 39-taxon/10-gene dataset C of Kawahara *et al.* (2011) with the 96-taxon/10-gene dataset from the present study, using the same analytical approaches in GARLI. Adding taxa improved branch support for some but not all nodes (Figure S6). For example, the bootstrap percentage for Marmarinae + Oecophyllembiinae improved (<50–64%), whereas support for Acrocercopinae dropped slightly (100 to 91%). The proportion of branches with strong (BP ≥ 80%) support in the current study was only slightly higher than the comparable analysis from the previous study [68/111 (61.3%) of nodes in this study compared with 32/55 (58.2%) of nodes in Kawahara *et al.*, 2011], suggesting that taxon addition did not have a dramatic effect on branch support of deep gracillariid relationships.

Rogue taxa in a phylogenetic study can substantially reduce bootstrap support for nodes, thereby masking true phylogenetic signal (e.g. Wilkinson, 1995; Liu & Pearl, 2007). Removal of the nine rogue taxa identified by ROGUENAROK resulted in substantial bootstrap increases for up to 13 nodes, most notably for Gracillariinae + Parornichinae, which rose from 60 to 91%.

Removal of just the top two rogue taxa, which had the largest 'raw improvement' values (Table 3), had a similar effect on nodal support. Bootstrap support increased by >3% at 11 nodes, with some nodes, such as the 'LAMPO' clade, increasing by >30% (54–89%).

Rogue taxa are often the result of problems with molecular data quality. However, one rogue taxon in this study, *Callicercops iridocrossa*, has also historically proven difficult to place by morphology. This species is a member of an enigmatic genus with a combination of traits that are both derived and primitive (Vári, 1961). Although *Callicercops* was once tentatively placed in the *Acrocercops* group, *Callicercops* lacks one of the diagnostic features of the *Acrocercops* group, a long intersegmental membrane between A8 and the external genital organs in males (Fig. 4B) (Kumata *et al.*, 1988a,1988b). Its relationship to other Gracillariidae has remained uncertain, as some traits, such as a fourth instar that cuts and rolls the host leaf into a cone, are shared with *Macarostola* and *Calybites* (Fig. 6F). Prior to its removal from the datasets in the present study, *Callicercops* was placed as the sister group to Phyllocnistinae in the 22-gene ML tree, but with weak support (BP = 48%; Figure S7). We therefore do not draw any formal conclusions regarding its phylogenetic placement. For nomenclatural purposes, it is still considered a member of the subfamily Gracillariinae.

Synonymous versus nonsynonymous changes

Multiple studies on the phylogenetic relationships of ditrysian Lepidoptera have demonstrated that the signal from synonymous substitutions, relative to that from nonsynonymous substitutions, can present analytical challenges at particular nodes because of faster divergence in nucleotide composition and accumulation of multiple substitutions per site (e.g. Regier *et al.*, 2009, 2013; Cho *et al.*, 2011). Synonymous substitutions constitute more than 90% of total nucleotide change, and many studies have acknowledged the potentially misleading effects of these changes on higher-level phylogeny inference (e.g. Regier *et al.*, 2009; Cho *et al.*, 2011; Breinholt & Kawahara, 2013). Authors have responded either by removing first and third positions altogether (e.g. Misof *et al.*, 2014) or by removing synonymous change through degeneracy coding (e.g. the degen1 coding of Regier *et al.*, 2009, 2010, 2013; Zwick, 2010). However, for younger divergences (i.e. within a family), the generally faster synonymous change might be particularly informative.

We addressed this question by comparing the levels of bootstrap support for nt123 and degen1 analyses. Table 2 shows that, for 15 historically problematic deep nodes, bootstrap support from nt123 is greater than or equal to that from degen1, without there being any strong conflicts that might imply misleading signal from synonymous change. For the entire tree, nt123 recovers ten more nodes with bootstrap support ≥80% than degen1 with dataset 1, and nine more such nodes with dataset 2. These results suggest that within Gracillariidae, synonymous change contributes most of the signal without introducing significant data conflict.

Relationships within Gracillariidae: deep divergences

This study provides the first large-scale molecular analysis of Gracillariidae, more than doubling the previous taxon sampling. Our discussion focuses on the dataset 2, nt123 tree (Fig. 2) because it includes the greatest amount of character data and provides the most conclusive results. Two main findings emerge from our analyses. First, we find very strong support (BP ≥ 95%) for eight clades corresponding to current subfamilies or previously recognized subclades thereof. These form the basis for a revised subfamily classification, which we present in the next section. Second, we find strong support for some, though not all, groupings subtending multiple subfamilies, allowing partial reconstruction of the earliest divergences within Gracillariidae.

The Gracillariidae divide basally into three strongly supported clades, among which relationships are very weakly and inconsistently supported (node 2; BP = 53%). These three major clades are: (i) Ornixolinae (node 15; BP = 97%, nt123), confirmed here as a subfamily (see later); (ii) the Gracillariinae **sensu n.** and Parornichinae (node 6, BP = 83%, nt123); and (iii) the 'LAMPO' clade (node 3; BP = 89%, nt123), consisting of Lithocolletinae, Acrocercopinae, Marmarinae, Phyllocnistinae and Oecophyllembiinae. The LAMPO clade was not recovered by degen1 analyses, in which the ornixoline genus *Chileoptilia* was placed at the base of the Marmarinae + Oecophyllembiinae, but support for this alternative is very weak (BP = 28%; Figure S1). Exclusion of *Chileoptilia* from the degen1 analysis recovers the LAMPO clade, albeit with weak support (BP = 45%; Figure S8). There are no known morphological synapomorphies for the LAMPO clade, but we predict that new characters will be discovered that are shared among these five subfamilies.

Within the LAMPO clade, the AMPO clade (node 5) and the grouping of three of its subfamilies (Marmarinae, Oecophyllembiinae and Phyllocnistinae; node 8) are weakly supported (BP ≤ 60%). The strongest grouping is that between Marmarinae and Oecophyllembiinae (node 9, BP = 78%, nt123).

None of the weakly supported groupings of subfamilies (nodes 2, 5, 8) display high levels of conflict in single-gene analyses, and no morphological or behavioural characters support or contradict any of the three, with one exception: the subfamilies Marmarinae, Oecophyllembiinae and Phyllocnistinae all have larvae that construct slender, subepidermal, serpentine mines throughout the feeding instars (Davis, 1994; Kumata, 1998; Wagner *et al.*, 2000) (Fig. 6C–E). In general, the lack of robust support for some groups in the combined molecular analysis might be caused by insufficient data, short internal branches or particular taxa that are difficult to place.

Our data provide strong evidence on some but not all previous hypotheses of relationships among gracillariid subfamilies. The inclusion of Phyllocnistinae in the robustly supported LAMPO clade by our analysis conflicts strongly with Kuznetzov & Stekol'nikov's (1987) who postulated that this subfamily is the first to branch off within the family (Fig. 1A). Recent molecular analyses by Regier *et al.* (2013) (Fig. 1B) and Kawahara *et al.* (2011) (Fig. 1C) provided weak support (BP < 50%) for the Phyllocnistinae as the sister group to the Lithocolletinae.

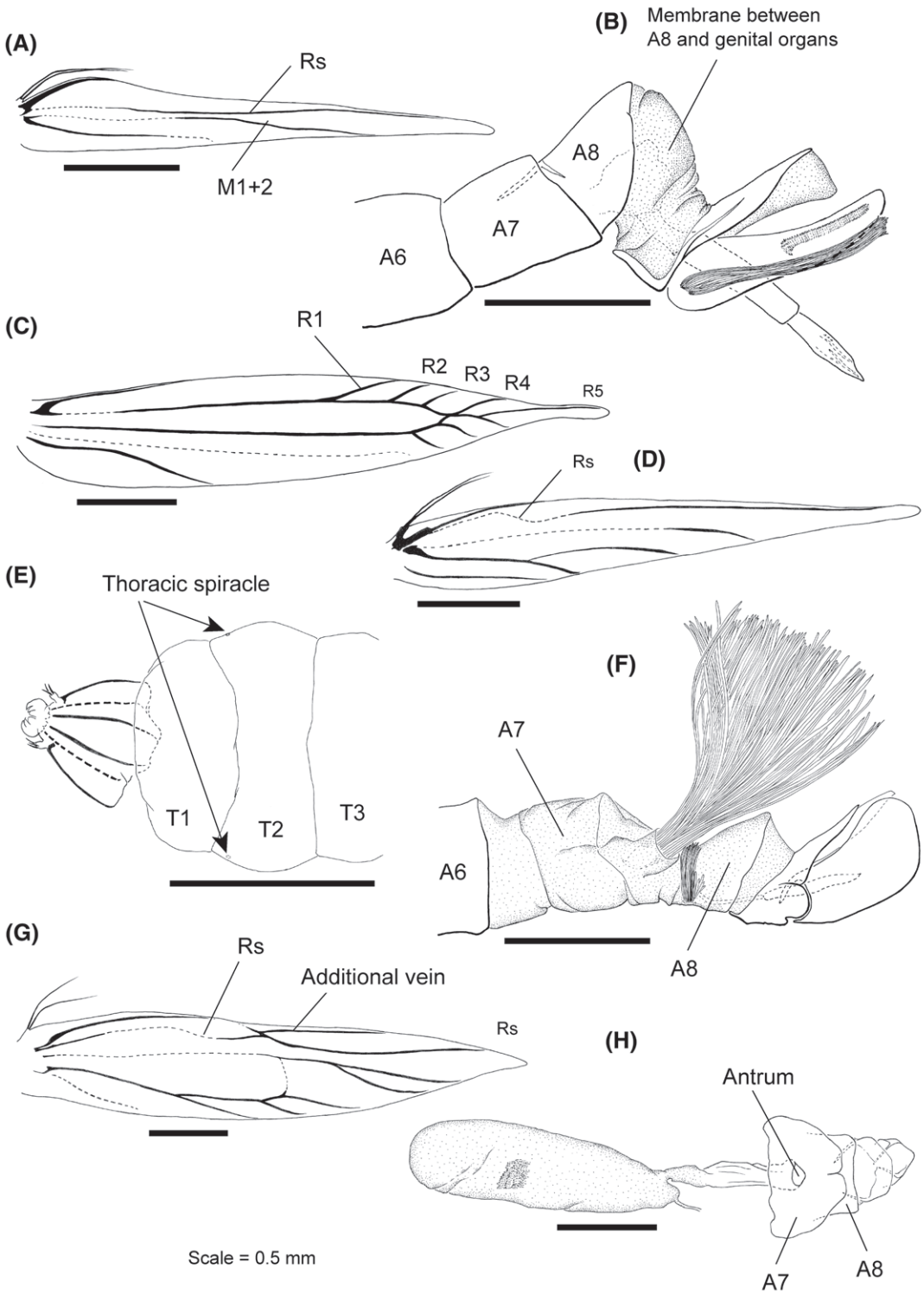

Fig. 4. Key morphological features of each subfamily recognized in this study. Scale bar = 0.5 mm. (A) Lithocolletinae: *Phyllonorycter bicinctella* (Matsumura), hindwing; (B) Acrocercopinae: *Acrocercops transecta* Meyrick, male terminalia; (C) Phyllocnistinae: *Phyllocnistis* sp. associated with *Podocarpus macrophyllus* (Podocarpaceae), forewing; (D) Marmarinae: *Dendrorycter marmaroides* Kumata, hindwing; (E) Oecophyllembiinae: *Metriochroa syringae* Kumata, head and thorax of sap-feeding larva; (F) Gracillariinae: *Caloptilia theivora* (Walsingham), male terminalia; (G) Parornichinae: *Callisto multimaculata* (Matsumura), hindwing; (H) Ornixolinae: *Conopomorpha litchiella* Bradley, female genitalia.

Some authors have treated taxa here placed in the Oecophyllembiinae as Phyllocnistinae, because these two subfamilies share similar larval and adult morphology (e.g. Davis, 1994; Davis & Robinson, 1998; Vári *et al.*, 2002). The sister-group relationship of these two subfamilies was not recovered in any of our analyses, as Oecophyllembiinae was nearly always grouped instead with Marmarinae (Node 9; BP = 78%). The bootstrap value is not quite 80%, and the AU test just misses significance ($P \geq 0.062$, Table 1), but the molecular evidence clearly leans in the direction of nonmonophyly for the Phyllocnistinae sensu lato of Davis & Robinson (1998). Ongoing work using RNA-Seq in Lepidoptera (Bazinet *et al.*, 2013; Kawahara & Breinholt, 2014) and anchored hybrid enrichment (Lemmon *et al.*, 2012) holds promise for conclusive resolution of this and other relationships among subfamilies of Gracillariidae.

Proposed classification and diagnoses of subfamilies

As noted earlier, we find very strong support (BP ≥ 95%) from dataset 2 (94 taxa, 22 genes) for a number of groupings corresponding to subfamilies or divisions thereof proposed by previous authors. In this section we present a new subfamily classification derived from our results. In recognizing the eight subfamilies discussed in the following, we sought as much as possible to use concepts corresponding to pre-existing formal or informal family-group names, and modified circumscription of these only when needed to ensure monophyly as judged from the molecular results. Our goal was to eliminate historically problematic genus-group names and inconsistency in taxonomic rank. For each proposed subfamily we provide both molecular and morphological evidence for (and against) monophyly, and a morphological diagnosis. We begin with a discussion of monophyly for the family, followed by each subfamily in the arrangement shown in Fig. 2.

Gracillariidae Stainton, 1854 (node 1, nt123/degen1 BP = 100/81)

'Gracillariidae' – Stainton, H. T. 1854. Insecta Britannica, Lepidoptera, Tineina: 193. Type genus: *Gracillaria* Haworth, 1828, Lepidoptera Britannica: 527.

Diagnosis. Minet (1986) and Davis (1987) initially proposed a single apomorphy for Gracillariidae: hypermetamorphosis, correlating with a change of habit from sap-drinking to tissue-feeding and/or spinning. However, the recently described Brazilian gracillariid *Spinivalva gaucha* lacks sap-feeding larval instars (Brito *et al.*, 2013). There are several additional apomorphies, including crochet-bearing prolegs on A3–A5 (rarely on A6), the presence of two L setae on T1 (Kobayashi *et al.*, 2011), 14 legs on the final instar, male genitalia with only four pairs of muscles and without a gnathos, females with a short, laterally flattened ovipositor (Kuznetzov & Stekol'nikov, 1987), and antennal flagellomeres in adults of both sexes with two rows of scales: a basal row of large scales covering an apical row of smaller scales (Triberti, 1998).

All of our molecular analyses support monophyly of Gracillariidae, consistent with Kawahara *et al.* (2011). Bootstrap support was strong for degen1 alone and rose sharply when both synonymous and nonsynonymous changes were included (BP = 100%, nt123).

Lithocolletinae Stainton, 1854 (node 4, BP = 100/100, Figs. 5A, 6A)

'Lithocolletidae' – Stainton, H. T. 1854. Insecta Britannica, Lepidoptera: Tineina: 10 (key), 264.

Type genus. Lithocolletis Hübner, [1825], Verzeichnis bekannter Schmetterlinge: 423.

Diagnosis. In the hindwing, the parallel condition of the vein Rs with the vein M1 or M1+2 towards the base (Kumata, 1993) (Fig. 4A) is an apomorphy of this subfamily. Because of its unique hindwing venation, Lithocolletinae has long been recognized to be divergent from other Gracillariidae (Stainton, 1854; Spuler, 1910; Kuznetzov, 1981; Kuznetzov & Stekol'nikov, 1987; De Prins & Kawahara, 2012).

Lithocolletinae was well supported in all molecular analyses. *Leucanthiza* was first described in the family Lithocolletidae (Clemens, 1859), but because of its morphological similarity to *Metriochroa* (Vári, 1961), Davis (1983) moved it to Gracillariinae. All molecular analyses, however, including those in Kawahara *et al.* (2011), strongly support the inclusion of *Leucanthiza* in Lithocolletinae, and the AU test statistically rejects the exclusion of *Leucanthiza* from Lithocolletinae ($P < 0.001$; Table 1). This supports the recent transfer of *Leucanthiza* from Gracillariinae to Lithocolletinae (De Prins *et al.*, 2013). *Phyllonorycter aberrans* has been known to belong in a new genus, and D. Davis will formally describe this genus and its morphological features in a separate publication.

Acrocercopinae Kawahara & Ohshima, new subfamily (node 5, BP = 99/95, Figs. 5B, 6B)

http://zoobank.org/urn:lsid:zoobank.org:act:B23E4AE1-99F0-47BA-8680-45FA8DB9AF45

Type genus. Acrocercops Wallengren, 1881. Entomologisk Tidskrift 1(2): 95, by present designation.

Diagnosis. Apomorphies of Acrocercopinae include a long intersegmental membrane between the A8 and the external genital organs in males (Fig. 4B) (Davis & Wagner, 2005), a curved forewing anal vein and a completely red final larval instar (Kumata *et al.*, 1988a,b). Furthermore, the maxillary palpi of the adult are relatively well developed and four-segmented. The forewing possesses all three medial veins, with two medial veins present in the hindwing; the hindwing is relatively narrow with a maximum width of ~0.10–0.14× that of its length. The final instar possesses two lateral setae on all segments, five pairs of

stemmata (one less than Gracillariinae) and four to five pairs of labral setae.

The *Acrocercops* group *sensu* Kumata (1982) is here elevated to the subfamily rank as Acrocercopinae. Monophyly for the Acrocercopinae was strongly supported, with BP ≥ 95% in all analyses conducted.

Phyllocnistinae Herrich-Schäffer, 1857 (node 13, BP = 100/100, Fig. 5C, 6C)

'Phyllocnistina' – Herrich-Schäffer, G. A. W. 1857. Correspondez-Blatt des zoologisch-minerologischen Vereines in Regensburg 11: 58.

Type genus. Phyllocnistis Zeller, 1848, Linnaea Entomologica 3: 250 (key), 264–266.

Diagnosis. Species in the Phyllocnistinae share an R1 vein arising from the apical half of the discoidal cell in the forewing (Fig. 4C), a putative apomorphic adult character. However, adults are often very small (Fig. 5C) without many other diagnostic characters, so immature stages have also been used to make identifications. Three pupal characters distinguish Phyllocnistinae from other gracillariid subfamilies: (i) frontal process (cocoon cutter) without lateral processes or setae on clypeus; (ii) tergal spines with a pair of dorsal setae and dorsal hooks; and (iii) cremaster with only one pair of caudal processes (Kobayashi *et al.*, 2013). Hindwing venation (M1, M2 and M3 veins do not arise from the vein Rs) and the lack of mesothoracic larval spiracles separate most Phyllocnistinae from Oecophyllembiinae (Kumata, 1998). The adult lacks maxillary palpi and possesses narrow wings with reduced venation. In some species, the apex of the forewing forms a slender lobe. The forewing medial vein has a single branch and the cubital vein has two branches. The maximum width of the hindwing is 0.09–0.10× the length of the wing, and the medial and cubital veins each consist of a single branch. Larval hypermetamorphosis is extreme within this subfamily (as well as in Oecophyllembiinae), with early sap-feeding instars possessing a very flat head with broad, flat mandibles, and no stemmata. The last, nonfeeding instar possesses greatly reduced head morphology, with reduced antennae, no mandibles and no stemmata. The only well-developed head structure is a silk-emitting spinneret, which enables this instar to construct a cocoon for pupation.

Phyllocnistinae was monophyletic and well supported in the present study. It includes one genus, *Phyllocnistis* Zeller, with approximately 100 described species of largely uniform adult morphology (De Prins & Kawahara, 2009; Kawahara *et al.*, 2009; Davis & Wagner, 2011; Kobayashi *et al.*, 2013).

Marmarinae Kawahara & Ohshima, new subfamily (node 11, BP = 100/97, Figs. 5D, 6D)

http://zoobank.org/urn:lsid:zoobank.org:act:4D2D01C8-9E9C-4AEA-85AE-43D8CB04D4E4

Type genus. Marmara Clemens, 1863, Proceedings of the Entomological Society of Philadelphia 2: 6–8, by present designation.

Diagnosis. Putative apomorphies of Marmarinae include the reduced wing venation with an anteriorly arched hindwing radial sector (Fig. 4D) and the wholly concealed quiescent form within the exoskeleton of the preceding instar (the final sap-feeding instar) throughout the pharate phase (Hinton, 1946, 1971). However, *D. marmaroides* lacks an R1 vein in the forewing, an M2 vein in the hindwing and a stalked Cu1 vein in the hindwing. *Dendrorycter* has a male valva that is reduced to a single lamella, but the male valva of *Marmara* is deeply divided, which further distinguishes the two genera (Kumata, 1978). Adult Marmarinae possess a short, three-segmented maxillary palpus with the apical segment elongated. The hindwing is narrow with a maximum width of ∼0.1× that of the length, and with only a single branch remaining in both the M and Cu veins. A single pair of elongate coremata is present on male abdominal segments A6–8. The last spinning instar possesses five to six pairs of stemmata and three pairs of labral setae (Guillen *et al.*, 2001) and a single lateral seta on the mesothorax and metathorax. The head of the pharate phase is very reduced and lacks stemmata and mandibles, and the exuviated head capsule of this phase is usually kept within that of the final sap-feeding instar. The spinneret is also greatly reduced in size but may still be functional.

Marmarinae, newly proposed here, includes two genera thus far: *Marmara* Clemens and *Dendrorycter* Kumata, the latter of which contains the sole species *D. marmaroides* Kumata. Kumata (1978) (Fig. 5D) was the first to note the similarity of *Dendrorycter* to *Marmara*.

Oecophyllembiinae Réal & Balachowsky, 1966 (node 12, BP = 95/80, Figs. 5E, 6E)

'Oecophyllembiinae' – Réal & Balachowsky, 1966. Entomologie appliquée à l'Agriculture 2 (Lépidoptères 1): 333.

Type genus. Oecophyllembius Silvestri, 1908, Bollettino del Laboratorio di Zoologia generale e agraria della R. Scuola Superiore d'Agricoltura de Portici 2: 196–199.

Diagnosis. Oecophyllembiinae is one of three gracillariid subfamilies (along with Phyllocnistinae and Marmarinae) whose larvae lack tissue-feeding instars (Fig. 6E) and, consequently, also lack granular frass; the larvae instead have at least three sap-feeding instars followed by one nonfeeding, highly specialized spinning instar (Davis, 1987). The spinning instar constructs a cocoon within the slightly broader terminus of the mine (Vári, 1961; Davis, 1994). Oecophyllembiinae was recognized as a separate subfamily by Kumata (1998) based on the presence of larval spiracles on the mesothorax (Fig. 4E) and hindwing venation (M1 and M2 veins arise from the vein Rs). We consider these characters putative apomorphies of the subfamily, although mesothoracic larval spiracles are notably lacking

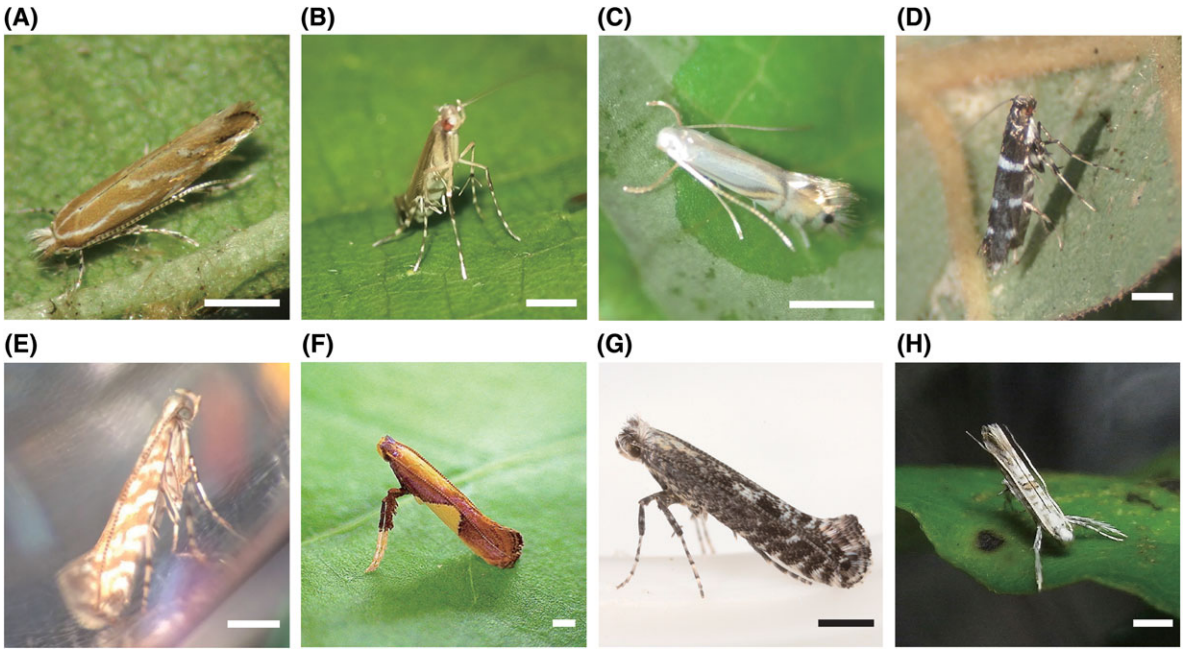

Fig. 5. The adult resting posture of exemplar species from each of the eight gracillariid subfamilies. Adults belonging to the Acrocercopinae, Gracillariinae, Marmarinae, Oecophyllembiinae and Parornichinae typically raise their head and thorax. Adults of Lithocolletinae and Phyllocnistinae either keeps its entire body parallel to the surface or raises its posterior end of its abdomen. In Ornixolinae, the adult raises the posterior end of its abdomen at a steep angle. Scale bar = 1 mm. (A) Lithocolletinae: *Phyllonorycter issikii* (Kumata); (B) Acrocercopinae: *Melanocercops ficuvorella* (Yazaki); (C) Phyllocnistinae: *Phyllocnistis toparcha* Meyrick; (D) Marmarinae: *Dendrorycter marmaroides* Kumata; (E) Oecophyllembiinae: *Eumetriochroa miyatai* Kumata (photograph: M. Kobayashi); (F) Gracillariinae: *Caloptilia cecidophora* Kumata (photograph: A. Hamatani); (G) Parornichinae: *Parornix* sp. associated with *Betula populifolia* (Betulaceae) (photograph: C. Eiseman); (H) Ornixolinae: *Cuphodes wisteriella* Kuroko. [Colour figure can be viewed at wileyonlinelibrary.com].

in the genus *Prophyllocnistis* (Davis 1994). At least three diagnostic pupal characters have been reported for this subfamily: (i) cocoon cutter with unique lateral processes or setae on the clypeus; (ii) tergal spines with only one pair of dorsal setae; and (iii) cremaster with more than two pairs of caudal processes (Kobayashi *et al.*, 2013) except in *Prophyllocnistis* Davis, which has a cremaster with a single large pair of caudal processes (Davis, 1994). The last instar possesses one or two lateral setae on the mesothorax and metathorax. The last instars of Oecophyllembiinae are unusual in the number and placement of abdominal prolegs. For instance, *Metriochroa* and *Prophyllocnistis* possess prolegs on A3–6, and *Cryphiomystis* has prolegs on A2–6. The last instar head of both *Metriochroa* and *Prophyllocnistis* lack stemmata and possess only two to three pairs of labral setae. The adult maxillary palpus can also vary; it is absent in *Prophyllocnistis* and three-segmented in *Metriochroa.* Wing venation is reduced, with R1 absent and only a single branch of M and Cu present in the forewing. The hindwing ranges from narrow to relatively broad, with the maximum width ~0.10–0.18× the length of the wing. The medial vein has two or three branches and the cubital vein has only a single branch.

Oecophyllembiinae originally included four genera: *Cryphiomystis* Meyrick (=*Corythoxestis* Meyrick), *Eumetriochroa* Kumata, *Guttigera* Diakonoff, and *Metriochroa* (Kumata, 1998). *Angelabella* Vargas & Parra was described as an oecophyllembiine genus (Vargas & Parra, 2005), and *Prophyllocnistis* Davis was transferred to Oecophyllembiinae by Kobayashi *et al.* (2013). Four of the six oecophyllembiine genera that we were able to include in our analyses (*Angelabella*, *Corythoxestis*, *Eumetriochroa*, *Metriochroa*) formed a monophyletic group with strong support (BP = 95%).

Gracillariinae Stainton, 1854 (node 10, 83/76, Figs. 5F, 6F)

'Gracillariidae' – Stainton, H. T. 1854. Insecta Britannica, Lepidoptera, Tineina: 193.

Type genus. Gracillaria Haworth, 1828, Lepidoptera Britannica: 527.

Diagnosis. The males of all genera in Gracillarinae *sensu stricto* have a membranous tergite and sternite on A8 (Fig. 4F), a putative apomorphy of the group. The final instars exhibit leaf-folding behaviour (Fig. 6F,G) and possess a full complement of six stemmata, six labral setae, three lateral setae on the mesothorax and metathorax, and two lateral setae on each abdominal segment.

Monophyly of Gracillariinae *sensu* Davis & Robinson (1998), including all subfamilies proposed here, except Lithocolletinae and Phyllocnistinae, remains conceivable, but the molecular evidence clearly leans the other way. In Fig. 2, Acrocercopinae and Marmarinae, as members of the LAMPO clade, are separated

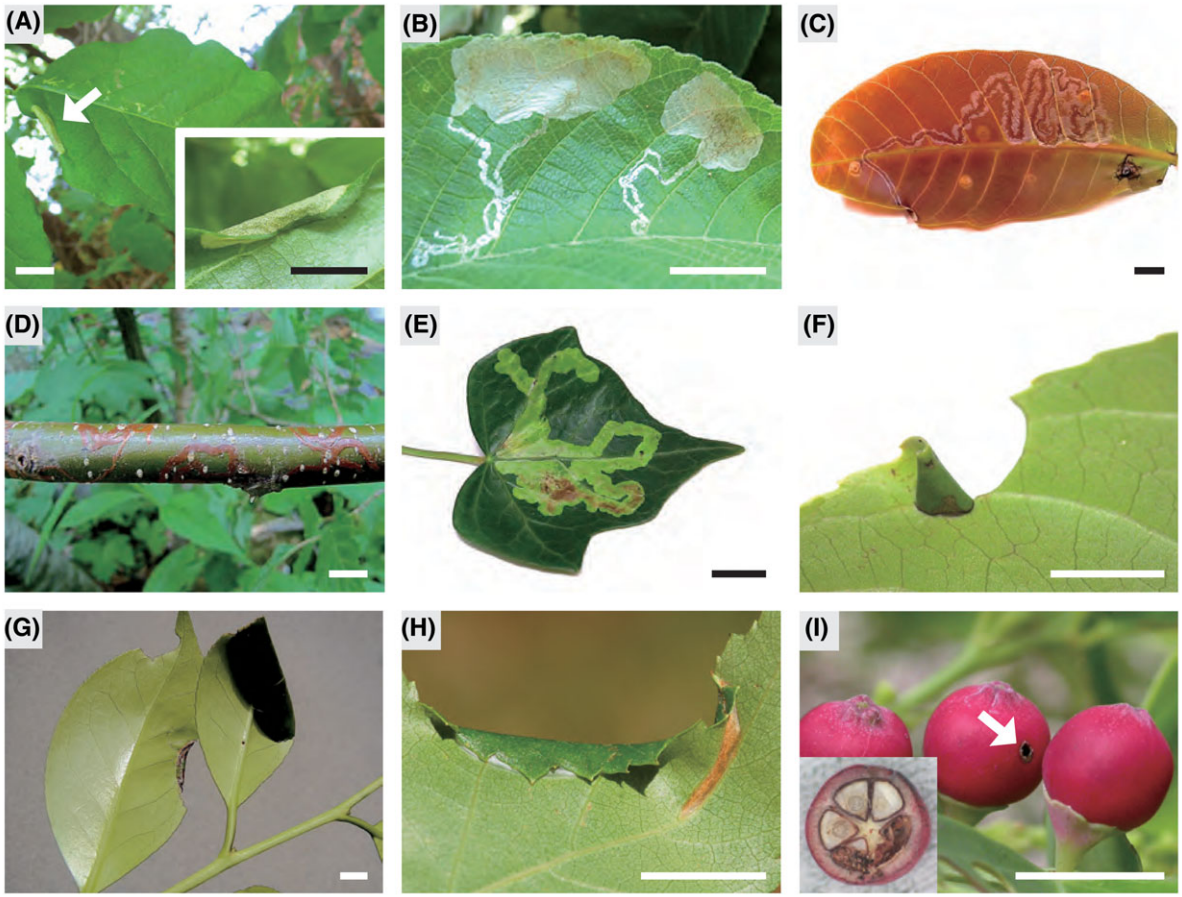

Fig. 6. Larval habitus of exemplar species from each of the eight gracillariid subfamilies. Scale bar = 10 mm. (A) Lithocolletinae: *Phyllonorycter lyoniae* (Kumata), on *Lyonia ovalifolia* (Ericaceae), tentiform blotch-mine; (B) Acrocercopinae: *Acrocercops transecta* Meyrick, on *Juglans mandshurica* (Juglandaceae) – early instars produce serpentine linear mines but third and later instars make blotch-mines; (C) Phyllocnistinae: *Phyllocnistis* sp. associated with *Anacardium occidentale* (Anacardiaceae) – all instars produce serpentine mines; (D) Marmarinae: *Dendrorycter marmaroides* Kumata, stem-mine on *Alnus hirsuta* (Betulaceae) – all instars produce serpentine mines; (E) Oecophyllembiinae: *Eumetriochroa hederae* Kumata, on *Hedera rhombea* (Araliaceae) – all instars produce serpentine mines; (F) Gracillariinae: *Macarostola japonica* Kumata, on *Euscaphis japonica* (Staphyleaceae) – the fourth instar cuts the leaf from the edge towards the midrib and rolls it to form a cone; (G) Gracillariinae: *Caloptilia theivora* (Walsingham), on *Camellia japonica* (Theaceae) – early instars produce mines at the edge of the leaf (left side), while the fourth and later instars roll leaf edges (right side); (H) Parornichinae: *Parornix* sp. associated with *Betula populifolia* (Betulaceae) (photograph: C. Eiseman) – early instars produce blotch-mines, and the fourth and later instars fold the leaf margin and feed inside the resulting shelter; (I) Ornixolinae: *Epicephala vitisidaea* Li, Wang & Zhang, on *Breynia vitis-idaea* (Phyllanthaceae) – larvae bore into fruits and feed on seeds. [Colour figure can be viewed at wileyonlinelibrary.com].

from the remaining Gracillariinae *sensu lato* by a node (node 3) with 89% bootstrap support, although surprisingly the AU test does not significantly reject monophyly of Davis & Robinson's concept ($P > 0.999$; Table 1, group 8). To ensure monophyly of our subfamily concept, we here define Gracillariinae *sensu stricto* as the well-supported clade that includes *Aristaea* Meyrick, *Caloptilia* Hübner, *Calybites* Hübner, *Eucalybites* Kumata, *Euspilapteryx* Stephens, *Gracillaria* Haworth, *Ketapangia* Kumata, *Macarostola* Meyrick and *Systoloneura* Vári (node 10, Fig. 2). We also formally place the genera *Aspilapteryx* Spuler, *Cupedia* Klimesch & Kumata, *Ectropina* Vári, *Neurolipa* Ely, and *Povolnya* Kuznetzov to Gracillariinae *sensu stricto*; these five genera were not in our molecular analysis, but were placed in the *Gracillaria* group by Kumata (1982, 1995) based on morphological evidence. Gracillarinae sensu lato contains some additional gracillariid genera (e.g. *Callicercops*) that are still considered to be in the subfamily Gracillariinae solely for consistency with previous nomenclature (Table 4). There is no evidence that these additional genera belong in Gracillarinae *sensu stricto*, but we do not currently have sufficient information to formally transfer them to an alternate clade.

Caloptilia, a globally distributed genus that includes nearly 300 described species (De Prins & De Prins, 2016), is polyphyletic in the current study. *Caloptilia murtfeldtella* Busck and *C. obliquatella* Matsumura were placed with *Eucalybites*, *Euspilapteryx* and *Gracillaria* with strong support (node 19,

Table 4. List of all valid gracillariid genera, according to De Prins & De Prins (2016), with proposed subfamilial placement.

Acrocercopinae Kawahara & Ohshima, 2016
Acrocercops Wallengren, 1881[a]
Amblyptila Vári, 1961[a]
Artifodina Kumata, 1985[a]
Borboryctis Kumata & Kuroko, 1988[a]
Chilocampyla Busck, 1900[a]
Chrysocercops Kumata & Kuroko, 1988[b]
Corethrovalva Vári, 1961[c]
Cryptolectica Vári, 1961[d]
Dekeidoryxis Kumata, 1989[e]
Deoptilia Kumata & Kuroko, 1988[a]
Dialectica Walsingham, 1897[a]
Eteoryctis Kumata & Kuroko, 1988[a]
Eucosmophora Walsingham, 1897[a]
Gibbovalva Kumata & Kuroko, 1988[a]
Hypectopa Diakonoff, 1955[c]
Lamprolectica Vári, 1961[d]
Leucocercops Vári, 1961[c]
Melanocercops Kumata & Kuroko, 1988[a]
Leucospilapteryx Spuler, 1910[a]
Metacercops Vári, 1961[c]
Monocercops Kumata, 1989[e]
Phodoryctis Kumata & Kuroko, 1988[a]
Psydrocercops Kumata & Kuroko, 1988[a]
Sauterina Kuznetzov, 1979[d]
Schedocercops Vári, 1961[c]
Spulerina Vári, 1961[a]
Telamoptilia Kumata & Kuroko, 1988[a]
Vihualpenia Mundaca, Parra & Vargas, 2013[f]

Gracillariinae Stainton, 1854
Africephala Vári, 1986[g]
Apistoneura Vári, 1961[g]
Aristaea Meyrick, 1907[a]
Aspilapteryx Spuler, 1910[h]
Callicercops Vári, 1961[a,g]
Caloptilia Hübner, 1825[a]
Calybites Hübner, 1822[a]
Cryptologa Fletcher, 1921[g]
Cupedia Klimesch & Kumata, 1973[h]
Dextellia Triberti, 1986[g]
Ectropina Vári, 1961[h]
Epicnistis Meyrick, 1906[g]
Eucalybites Kumata, 1982[a]
Euprophantis Meyrick, 1921[g]
Eurytyla Meyrick, 1893[g]
Euspilapteryx Stephens, 1835[a]
Gracillaria Haworth, 1828[a]
Ketapangia Kumata, 1995[a]
Neurolipa Ely, 1918[h]
Penica Walsingham, 1909–1915[g]
Macarostola Meyrick, 1907[a]
Polymitia Triberti, 1986[g]
Povolnya Kuznetzov, 1979[i]
Synnympha Meyrick, 1915[g]
Systoloneura Vári, 1961[a]

Lithocolletinae Stainton, 1854
Cameraria Chapman, 1902[a]
Chrysaster Kumata, 1961[a]
Cremastobombycia Braun, 1908[a]
Hyloconis Kumata, 1963 [a]
Leucanthiza Clemens, 1859[a]
Macrosaccus Davis & De Prins, 2011[a]
Neolithocolletis Kumata, 1963[a]
Phyllonorycter Hübner, 1822[a]
Porphyrosela Braun, 1908[j]
Protolithocolletis Braun, 1929[j]
Triberta De Prins *et al.*, 2013[j]

Marmarinae Kawahara & Ohshima, 2016
Dendrorycter Kumata, 1978[a]
Marmara Clemens, 1863[a]

Oecophyllembiinae Réal & Balachowsky, 1966
Angelabella Vargas & Parra, 2005[a]
Corythoxestis Meyrick, 1921[a]
Eumetriochroa Kumata, 1998[a]
Guttigera Diakonoff, 1955[k]
Metriochroa Busck, 1900[a]
Prophyllocnistis Davis, 1994[k]

Ornixolinae Kuznetzov & Baryshnikova, 2001
Apophthisis Braun, 1915[c]
Chileoptilia Vargas & Landry, 2005[a]
Conopobathra Vári, 1961[c]
Conopomorpha Meyrick, 1885[a]
Conopomorphina Vári, 1961[c]
Cuphodes Meyrick, 1897[a]
Cyphosticha Meyrick, 1907[c]
Diphtheroptila Vári, 1961[a]
Dysectopa Vári, 1961[c]
Epicephala Meyrick, 1880[a]
Leurocephala Davis & McKay, 2011[a]
Liocrobyla Meyrick, 1916[a]
Micrurapteryx Spuler, 1910[a]
Neurobathra Ely, 1918[a]
Neurostrota Ely, 1918[f]
Oligoneurina Vári, 1961[c]
Ornixola Kuznetzov, 1979[a]
Pareclectis Meyrick, 1937[c]
Parectopa Clemens, 1860[a]
Philodoria Walsingham, 1907[a]
Phrixosceles Meyrick, 1908[c]
Pogonocephala Vári, 1961[c]
Polydema Vári, 1961[c]
Polysoma Vári, 1961[c]
Semnocera Vári, 1961[c]
Spanioptila Walsingham, 1897[f]
Spinivalva Moreira & Vargas, 2013[f]
Stomphastis Meyrick, 1912[a]

Table 4. Continued

Parornichinae Kawahara & Ohshima, 2016	**Phyllocnistinae Herrich-Schäffer, 1857**
Callisto Stephens, 1834[a]	*Phyllocnistis* Zeller, 1848[a]
Graphiocephala Vári, 1961[c]	
Parornix Spuler, 1910[a]	
Pleiomorpha Vári, 1961[c]	

[a]Genus represented in molecular dataset of current study (Fig. 2).
[b]Placement based on intergeneric relationships discussed in Kumata *et al.* (1988b).
[c]Placement based on intergeneric relationships discussed in Vári (1961).
[d]Placement based on intergeneric relationships discussed in Kumata *et al.* (1988a).
[e]Placement based on intergeneric relationships discussed in Kumata (1989).
[f]Placement based on phylogenetic analyses of Lees *et al.* (2014).
[g]No formal taxonomic decision on subfamily placement.
[h]Placement based on intergeneric relationships discussed in Kumata (1995).
[i]Placement based on intergeneric relationships discussed in Kumata (1982).
[j]Placement based on De Prins & De Prins (2016).
[k]Placement based on intergeneric relationships discussed in Kumata (1998).
Bold text are recognized subfamily names. Many genera absent from our molecular dataset are herein formally assigned to new subfamilies, based on previous morphological and/or molecular evidence associating them with genera that are present in our molecular dataset (see footnotes). Some genera are not believed to be part of Gracillariine *sensu stricto*, but were formally assigned to Gracillariinae in previous literature (e.g. Vári, 1961). We refrain from making formal taxonomic decisions on the true placement of these genera (see footnotes), but herein leave them in the 'Gracillariinae' section of the table for nomenclatural consistency.

BP = 99%), while the seven other sampled *Caloptilia* species formed a separate, well-supported monophyletic group (node 18, BP = 99%). Although these results provide reason to narrow the definition of *Caloptilia* and transfer some species to *Gracillaria*, we choose not to do so without further taxon sampling. The name *Caloptilia* represents a diverse group with many pest species and has been used for centuries; a nomenclatural change would introduce much confusion into the literature, and would need thorough justification and documentation.

Parornichinae Kawahara & Ohshima, new subfamily (node 14, 100/99, Figs. 5G, 6G)
http://zoobank.org/urn:lsid:zoobank.org:act:B89C47D0-394D-4AE1-9141-2925A970901D

Type genus. Parornix Spuler, 1910. Die Schmetterlinge Europas. Mit über 3500 Figuren auf 95 Tafeln und 505 Abbildungen im Text. 3. Auflage von Prof. E. Hofmann's Werk: Die Groß-Schmetterlinge Europas. 4 Vols – Vol. 2: 410, by present designation. *Parornix* was established to denote a subgenus of *Ornix* Treitschke, 1833 (Nye & Fletcher, 1991: 266–267). *Ornix* Treitschke is itself a junior synonym and junior homonym of *Ornix* Kollar, 1832, though both have the same type species: *Tinea upupaepennella* (Hübner).

Note: Type species: *Ornix anglicella* Stainton, 1850. Transactions of the Entomological Society of London, N. S. (series 2) 1(3): 92. By subsequent designation by Walsingham (1909–1915). Biologia centrali-americana (Zoology) Lepidoptera-Heterocera 4: 341.

Diagnosis. Although adult Parornichinae possess a four-segmented maxillary palpus and retain all three branches to the forewing medial vein, as do the Acrocercopinae, Gracillariinae and some Ornixolinae, the forewing of Parornichinae only has a single branch present in the cubital vein, thus differing from the two branches present in Acrocercopinae, Gracillariinae and Ornixolinae. The hindwing of Parornichinae has an additional vein arising anteriorly from the Rs vein (Fig. 4G), an apomorphy of the group. The hindwing is the broadest among the eight subfamilies now recognized, with the maximum width varying 0.15–0.2× the wing length. The last instar of Parornichinae has three lateral setae on the abdominal segments as well as on the mesothorax and metathorax. The last instar also possesses a full complement of six stemmata and six labral setae (Kumata, 1965).

The well-supported clade that includes *Callisto* and *Parornix* is treated here as Parornichinae. While species in the clade containing Gracillariinae **sensu n.** and Parornichinae (node 6) could be treated together as Gracillariinae, we chose to recognize the Parornichinae because species in this subfamily have very different morphology compared with the species grouped in the Gracillariinae **sensu n.**

The *Callisto* + *Parornix* clade was originally treated as a subfamily by Kuznetzov & Stekol'nikov (1987), who called it 'Ornichinae', reviving the name Ornichidae Stainton, 1854. However, in order for this clade to retain its identity as a subfamily distinct from Gracillariinae, we must refrain from using the name Ornichinae, due to a previous interpretation by Stainton (1854) and Kuznetzov & Stekol'nikov (1987) of the type genus *Ornix*. Zeller (1839) refers to a revised description of *Ornix* and cites Treitschke (1833) as the previous author of the genus. The type species of *Ornix* Treitschke, 1833 was subsequently designated as *Tinea upupaepennella* (Hübner), which is a junior synonym of the gracillariinae species *Caloptilia stigmatella* (Fabricius). Thus, the type genus of Ornichinae Stainton must be treated as a gracillariine, and the subfamily name Ornichinae must then be treated as a

junior synonym of Gracillariinae Stainton, as per Davis & Robinson (1998). As *Ornix* cannot be the type genus for this new subfamily, we have chosen to designate *Parornix* Spuler as the type genus to reflect Kumata *et al.* (1988a), who previously referred to this clade as the *Parornix* group.

Ornixolinae Kuznetzov & Baryshnikova, 2001 (node 15, BP = 97/75, Figs. 5H, 6H)

'Ornixolinae' – Kuznetzov & Baryshnikova, 2001. Entomologicheskoe Obozrenie 80(1): 99.

Type genus. Ornixola Kuznetzov, 1979. Trudy Zoologicheskogo Instituta, Akademija Nauk SSSR 81: 99.

Diagnosis. Adult Ornixolinae possess moderately long, four-segmented maxillary palpi. The forewing venation is relatively complete for this subfamily, with two to three branches of the medial vein and two branches of the cubital vein. The hindwing is very narrow, with the greatest width varying from 0.09 to 0.12× the length of the wing. The medial vein has retained both branches but the cubital vein has only a single branch. Ornixoline females have an antrum that opens at the seventh sternum (Fig. 4H), a putative apomorphy that is overall an unusual condition for female Lepidoptera (Kumata, 1982). Most ornixoline females also have short, laterally flattened ovipositors, akin to other Gracillariidae, although the genus *Epicephala* exhibits considerable variation in ovipositor morphology, possibly as a result of coevolution with its host plants (Zhang *et al.*, 2012; Kawakita & Kato, 2016). The last instar possesses two lateral setae on the mesothorax and metathorax, five to six stemmata and four to six labral setae, with the greatest reductions in the number of stemmata and labral setae occurring in *Parectopa robiniella* Clemens (Davis *in litt.*).

The *Parectopa* group *sensu* Kumata (1982) is here elevated to the subfamily Ornixolinae **stat. rev.** *Ornixola*, which has previously been assigned a separate family-group name, Ornixolinae (Kuznetzov & Baryshnikova, 2001), is here clearly nested within this clade, subtended by several nodes with very strong support (Fig. 2; BP = 89 and 94%, respectively). Vargas & Landry (2005) described the genus *Chileoptilia* and considered it closely related to the ornixoline *Stomphastis* Meyrick, based on wing pattern (Vargas & Landry, 2005). *Chileoptilia* was not sister to *Stomphastis* in the present study, although the nt123 analysis clearly placed it in the Ornixolinae as a sister taxon to *Philodoria* Walsingham with strong support (BP = 93%). *Chileoptilia* has thus far only been found in Chile, whereas *Philodoria* is endemic to Hawaii (Johns *et al.*, 2016); therefore the phylogenetic placement of these genera has interesting biogeographical implications.

Larval host plant associations

One of the primary purposes of constructing a gracillariid phylogeny is to assess the role of ecological factors in the diversification of leaf-mining insects. Previous hypotheses about the evolution of larval feeding traits in Gracillariidae (e.g. Connor & Taverner, 1997) have been limited, and their evaluation has been hampered by the lack of a robust phylogeny. In this section, we review trends in such features in light of our new phylogeny. We present in Fig. 3, a provisional synopsis of species diversity, host-use patterns, and larval transitions, compiled in Table S3 and superimposed on the phylogeny illustrated in Fig. 2. To provide an initial characterization of the evolution of larval host plant use, we sought to assess the degree of conservatism with respect to the new phylogeny, of mode of leaf mining, host range (diversity of plant taxa used by individual species), host plant growth form, and host plant taxon membership at the family and ordinal level. We also sought to infer the ancestral conditions and evolutionary directionality of these traits when possible.

Mode of leaf mining, perhaps the most slowly evolving trait we examined, is strikingly conserved at the subfamily level or above (Fig. 3; Table S3). In the most phylogenetically widespread condition, seen in Acrocercopinae, some Gracillariinae and Ornixolinae, and most sampled genera of Lithocolletinae, the larva may initially create a narrow linear mine during the sap-feeding early instars, but later, it forms a simple blotch mine, within which it spends the remainder of the larval period (Fig. 6A, B). This condition is inferred in our analysis (Fig. 3) to be ancestral, a conclusion weakened, however, by low support for the basal divergences. Four uniquely derived, apparently independent departures from this putative ancestral state are seen. (i) In many species of Gracillariinae and Parornichinae (Fig. 2, node 6), the blotch-mining larva exits the mine in later instars and completes development in a shelter constructed by rolling the edge of the leaf (Harrison, 2016) (Fig. 6G). (ii) In Phyllocnistinae, Marmarinae and Oecophyllembiinae (the 'MPO' clade), there is no transition to a tissue-feeding form; instead, the sap-feeding larva continues to produce a slender, sub-epidermal serpentine mine throughout development (Fig. 6C–E). (iii) In a subclade of Lithocolletinae that includes *Phyllonorycter* Hübner and *Cremastobombycia* Braun, the sap-feeding instars typically feed instead on the spongy mesophyll, producing a blotch mine on the lower side of the leaf, before returning to feed on the palisade layer of the mine during the tissue-feeding instars. The result is a tent-like blotch mine (Fig. 6A). Tentiform blotch mining has also been observed in species of the acrocercopine genus *Acrocercops*, suggesting multiple independent origins of this mining behaviour (Fig. 3). In *Acrocercops*, this behaviour is correlated with larval preference to mine on a particular side of the leaf: most species mine the adaxial surface and form a normal blotch, but species that mine the abaxial surface (e.g. *A. albinatella*) form a tent. (iv) In Ornixolinae, three of the sampled species feed on flowers or fruit instead of leaves (Fig. 6I), and thus do not produce leaf mines. These taxa do not form a monophyletic clade in our analysis, and one of them (*Conopomorpha cramerella* Snellen) exhibits different feeding behaviour from that of its blotch-mining congener (*Conopomorpha* sp.), providing evidence that flower/fruit-feeding behaviour has independently evolved multiple times in Ornixolinae. (v) A final, unusual type of derivation from simple blotch mining is exemplified by *Caloptilia murtfeldtella*, which is a stem galler (Busck,

1904). In other insects that both mine and gall, galling also appears to be derived from leaf mining (Nyman, 2010).

A second relatively conserved aspect of host use, though seemingly somewhat more labile than mode of leaf mining, is diet breadth. Oligophagy, defined as using plants of a single plant order, is strongly conserved across gracillariids, characterizing 89% of the 82 sampled species for which we have host plant data, and is probably the ancestral condition. We could be underestimating the incidence of polyphagy, defined as using two or more plant orders, because for many species only a single host record exists, and some species in our analysis do not have any host data. However, it appears that oligophagy is considerably more prevalent in Gracillariidae than in Macroheterocera (Powell *et al.*, 1998; Menken *et al.*, 2009). Similar levels of oligophagy are seen in other internal feeding groups such as Nematinae sawflies (Nyman *et al.*, 2006). Unlike the case of host growth form, departures from the prevailing condition show no obvious clustering on the phylogeny: all nine lineages in which polyphagy are inferred to have arisen independently consist of a single species. Polyphagy itself is thus not phylogenetically conserved, but it is possible that the propensity to evolve it varies among lineages. For example, in our sample of the clade consisting of Gracillariinae and Parornichinae (node 6), polyphagy is inferred to have arisen five times among 21 species (24% incidence), whereas for the rest of the family it is inferred to have arisen four times among 73 species (5% incidence). One polyphagous species in Acrocercopinae, *Acrocercops transecta*, comprises two host races associating with distantly related host plants (Ohshima, 2008), and a possibility of host race formation should be assessed for other polyphagous species in future studies.

Host plant growth form shows a phylogenetic pattern somewhat like that of diet breadth. The most common, taxonomically widespread and probably ancestral association is with trees (60% of species in our sample). The species in our sample that feed on shrubs, vines or herbs represent mostly independent origins of each habit, a result consistent with a detailed study of the lithocolletine genus *Phyllonorycter* (Lopez-Vaamonde *et al.*, 2003). Occurrence of these nontree growth forms does not, however, appear entirely random on the phylogeny. Of the 18 lineages in which we infer strict nontree associations to have arisen (i.e. lineages with no host records from trees), five consist of two or more species, and one (*Hyloconis* Kumata + *Neolithocolletis* Kumata) contains four species.

Finally, host taxonomy at the family and ordinal level appears to be the least conserved host-use trait we examined. Although individual species are almost always host-specific, and multiple instances can be found of related species feeding on the same plant family (e.g. the early-diverging lineages of Lithocolletinae on Fabaceae), shifts among host orders/families are very common in this sample of species. Treating polyphagous host lists as ambiguity codes (thus minimizing their contribution to inferred host shift number), a minimum of 58 between-order shifts is required under the parsimony criterion to account for the distribution of host orders across the phylogeny. In other words, at least 62% (58/93) of inferred speciation events are associated with a host-order shift. The comparable fractions for shifts among leaf-mining modes, diet breadth categories and host growth forms are 11% (10/93), 10% (9/93) and 27% (25/93), respectively (when multiple host growth forms are treated as ambiguous). Lability of host taxon association suggests a potentially important role for host shifts in speciation/diversification of Gracillariidae.

The foregoing sketch of evolutionary patterns in gracillariid larval feeding ecology is intended as the first step towards a more detailed study, now in progress. We have identified and categorized several major aspects of larval host use, and provided initial estimates of both their evolutionary history and their relative degree of phylogenetic conservatism. Many questions are now open. One, which we have yet to address, is why Gracillariidae are invariably internal feeders (except sometimes in the last instar), whereas repeated shifts between internal and external feeding occur in the closely related Yponomeutoidea (Sohn *et al.*, 2013).

Conclusions

This study provides the most comprehensive, well-supported gracillariid phylogeny to date. We sampled nearly 100 leaf miner species and analyzed multiple datasets, with our 22-gene, nt123 dataset yielding the best supported ML tree. The main outcomes from this study are as follows:

1. We propose a new subfamily classification based on eight very strongly supported groupings that correspond to previously proposed formal or informal family-group names.
2. All of the subfamilies can be assigned to one of just three strongly supported, nonoverlapping clades, although relationships among these remain weakly supported.
3. An exploratory mapping of larval host-use traits on the new phylogeny shows strong phylogenetic conservation of modes of leaf mining; somewhat greater evolutionary lability of diet breadth and growth form of host plants used; and relatively frequent shifts among host plant orders and families, suggesting that such shifts could play a role in speciation/diversification.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: 10.1111/syen.12210

Figure S1. RAXML tree from the 114-taxon, degen1 analysis of dataset 2, excluding rogue taxon set 1.

Figure S2. RAXML tree from the 107-taxon, nt123 analysis of dataset 2, excluding rogue taxon sets 1 and 2.

Figure S3. RAXML tree from the 114-taxon, degen1 analysis of dataset 1, excluding rogue taxon set 1.

Figure S4. RAXML tree from the 114-taxon, nt123 analysis of dataset 1, excluding rogue taxon set 1.

Figure S5. RAXML tree from the 114-taxon, nt123 analysis of dataset 2, excluding rogue taxon set 1.

Figure S6. GARLI ML tree from the nt123, 10-gene, 114-taxon analysis, analysed with the same parameters as in Kawahara *et al.* (2011).

Figure S7. RAXML tree from the 116-taxon, nt123 analysis of dataset 2.

Table S1. Taxa sampled in the present study, along with GenBank accession numbers, codes, and life stage data. Some species are represented by multiple specimens, and have multiple accession numbers separated by a forward slash.

Table S2. PARTITIONFINDER results showing the preferred model for each partition. Numbers in parentheses refer to one of the three codon positions.

Table S3. Larval host plant data and references for Gracillariidae sampled in this study. References for mining behaviour listed in column E. Host plant information (column G) obtained from De Prins & De Prins (2016).

Acknowledgements

We thank Joaquin Baixeras (University of Valencia, Spain), Soowon Cho (Chungbuk National University, Korea), Willy De Prins (Royal Belgian Institute of Natural Sciences, Belgium), Camiel Doorenweerd (Nationaal Natuurhistorisch Museum Naturalis, the Netherlands), Timothy Friedlander (University of Maryland, USA), Terry Harrison (University of Illinois, USA), Jean-François Landry (Canadian National Collection, Canada), Erik van Nieukerken (Nationaal Natuurhistorisch Museum Naturalis, The Netherlands), Kenji Nishida (University of Costa Rica, Costa Rica) and Kazuhiro Sugisima (Okayama City, Japan) for providing some of the samples used in this study. We also thank Rodolphe Rougerie (Muséum National d'Histoire Naturelle, France) for sending valuable extracts, Kim Mitter (University of Maryland, USA) for assisting with sample preparation, and Charley Eiseman (University of Massachusetts, USA), Keiko Ariyoshi, Akihisa Hamatani, Masayuki Kobayashi, Kei Nariai and Itaru Yamamoto (Kyoto Prefectural University, Japan) for providing adult/larval photos or samples used to illustrate morphological traits. Adam Bazinet, Michael P. Cummings (University of Maryland, USA), Geena Hill (University of Florida, USA), Chris Johns (University of Florida, USA), Conrad Labandeira (National Museum of Natural History, USA), David L. Wagner (University of Connecticut, USA), and Andreas Zwick (CSIRO, Australia) provided constructive comments and assistance throughout this project. We also thank David Lees (University of Cambridge, UK), in particular, for his comments and fruitful discussions on nomenclature. We thank the University of Florida, High Performance Computing Cluster, and the Lattice Project at the University of Maryland for providing the computational support to conduct the analyses included in this study. This paper was funded in part by the National Science Foundation (grant nos DEB-0531639 and DEB-1354585), the National Geographic Society (grant no. 9686-15), and Maryland Agricultural Experiment Station.

References

Aberer, A.J. & Stamatakis, A. (2011) A simple and accurate method for rogue taxon identification. *Institute of Electrical and Electronics Engineers, International Conference on Bioinformatics and Biomedicine, 2001*. Atlanta, pp. 118–122.

Aberer, A.J., Krompass, D. & Stamatakis, A. (2013) Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Systematic Biology*, **62**, 162–166.

Bazinet, A.L., Cummings, M.P., Mitter, K.T. & Mitter, C.W. (2013) Can RNA-Seq resolve the rapid radiation of advanced moths and butterflies (Hexapoda: Lepidoptera: Apoditrysia)? An exploratory study. *PLoS ONE*, **8**, e82615.

Body, M., Burlat, V. & Giron, D. (2015) Hypermetamorphosis in a leaf-miner allows insects to cope with a confined nutritional space. *Arthropod-Plant Interactions*, **9**, 75–84.

Braun, A.F. (1908) Revision of the North American species of the genus *Lithocolletis* Hübner. *Transactions of the American Entomological Society*, **34**, 269–357, pl. XX–XXIV.

Breinholt, J.W. & Kawahara, A.Y. (2013) Phylotranscriptomics: saturated third codon positions radically influence the estimation of trees based on next-gen data. *Genome Biology and Evolution*, **5**, 2082–2092.

Brito, R., Gonçalves, G.L., Vargas, H.A. & Moreira, G.R.P. (2013) A new Brazilian *Passiflora* leafminer: *Spinivalva gaucha*, gen. n., sp n. (Lepidoptera, Gracillariidae, Gracillariinae), the first gracillariid without a sap-feeding instar. *ZooKeys*, **291**, 1–26.

Busck, A. (1904) Tineid moths from British Columbia, with descriptions of new species. *Proceedings of the United States National Museum*, **27**, 745–778.

Cho, S., Zwick, A., Regier, J.C. *et al.* (2011) Can deliberately incomplete gene sample augmentation improve a phylogeny estimate for the advanced moths and butterflies (Hexapoda: Lepidoptera)? *Systematic Biology*, **60**, 782–796.

Clemens, B. (1859) Contribution to American Lepidopterology. No. 2. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **11**, 317–328.

Connor, E.F. & Taverner, M.P. (1997) The evolution and adaptive significance of the leaf-mining habit. *Oikos*, **79**, 6–25.

Davis, D.R. (1983) Acanthopteroctetidae, Dalceridae, Eriocraniidae, Gracillariidae, Heliozelidae, Hepialidae, Incurvariidae, Lyonetiidae, Micropterigidae, Ochsenheimeriidae, Opostegidae, Psychidae, Tineidae, Tischeriidae. *Check List of the Lepidoptera of America North of Mexico* (ed. by R.W. Hodges, T. Dominick, D.R. Davis, *et al.*), pp. 2–11. E.W. Classey Ltd. and The Wedge Entomological Research Foundation, London.

Davis, D.R. (1987) Gracillariidae. *Immature Insects* (ed. by F.W. Stehr), pp. 372–374. Kendall/Hunt, Dubuque, Iowa.

Davis, D.R. (1994) New leaf-mining moths from Chile, with remarks on the history and composition of Phyllocnistinae (Lepidoptera: Gracillariidae). *Tropical Lepidoptera*, **5**, 65–75.

Davis, D.R. & De Prins, J. (2011) Systematics and biology of the new genus *Macrosaccus* with descriptions of two new species (Lepidoptera, Gracillariidae). *ZooKeys*, **98**, 29–82.

Davis, D.R. & Robinson, G.S. (1998) The Tineoidea and Gracillarioidea. *Lepidoptera, Moths and Butterflies: Evolution, Systematics and Biogeography, Handbook of Zoology*, Vols. 1, 4, Part 35 (ed. by N.P. Kristensen), pp. 91–117. De Gruyter, Berlin.

Davis, D.R. & Wagner, D.L. (2005) Biology and systematics of the Neotropical leafminer genus *Eucosmophora* (Lepidoptera: Gracillariidae). *Tropical Lepidoptera*, **13**, 1–40.

Davis, D.R. & Wagner, D.L. (2011) Biology and systematics of the New World *Phyllocnistis* leafminers of the avocado genus *Persea* (Lepidoptera: Gracillariidae). *ZooKeys*, **97**, 39–73.

Davis, D.R., McKay, F., Oleiro, M., Vitorino, M.D. & Wheeler, G.S. (2011) Biology and systematics of the leafmining Gracillariidae of Brazilian pepper tree, *Schinus terebinthifolius raddi*, with descriptions of a new genus and four new species. *The Journal of the Lepidopterists' Society*, **65**, 61–93.

De La Torre-bárcena, J.E., Kolokotronis, S.O., Lee, E.K. *et al.* (2009) The impact of outgroup choice and missing data on major seed plant phylogenetics using genome-wide EST data. *PLoS ONE*, **4**, e5764.

De Prins, J. & De Prins, W. (2016) *Global Taxonomic Database of Gracillariidae (Lepidoptera)*. Belgian Biodiversity Platform, Brussels. [WWW document]. URL http://www.gracillariidae.net [accessed on 25 March 2016].

De Prins, J. & Kawahara, A.Y. (2009) On the taxonomic history of *Phyllocnistis* Zeller 1848 (Lepidoptera: Gracillariidae). *Nota Lepidopterologica*, **32**, 27–35.

De Prins, J. & Kawahara, A.Y. (2012) Systematics, revisionary taxonomy, and biodiversity of Afrotropical Lithocolletinae (Lepidoptera: Gracillariidae). *Zootaxa*, **3594**, 1–283.

De Prins, J., Davis, D.R., De Coninck, E., Sohn, J.-C. & Triberti, P. (2013) Systematics, phylogeny and biology of a new genus of Lithocolletinae (Lepidoptera: Gracillariidae) associated with Cistaceae. *Zootaxa*, **3741**, 201–227.

Gilbert, M., Guichard, S., Freise, J. *et al.* (2005) Forecasting *Cameraria ohridella* invasion dynamics in recently invaded countries: from validation to prediction. *Journal of Applied Ecology*, **42**, 805–813.

Giron, D., Kaiser, W., Imbault, N. & Casas, J. (2007) Cytokinin-mediated leaf manipulation by a leafminer caterpillar. *Biology Letters*, **3**, 340–343.

Guillen, M., Davis, D.R. & Heraty, J.M. (2001) Systematics and biology of a new, polyphagous species of *Marmara* (Lepidoptera: Gracillariidae) infesting grapefruit in the southwestern United States. *Proceedings of the Entomological Society of Washington*, **103**, 636–654.

Gutzwiller, F., Dedeine, F., Kaiser, W., Giron, D. & Lopez-Vaamonde, C. (2015) Correlation between the green-island phenotype and *Wolbachia* infections during the evolutionary diversification of Gracillariidae leaf-mining moths. *Ecology and Evolution*, **5**, 2045–7758.

Harrison, T. (2016) *Microleps.org*. [WWW document]. URL http://www.microleps.org [accessed on 30 July 2015].

Heppner, J.B. (1993) Citrus leafminer, *Phyllocnistis citrella*, in Florida (Lepidoptera: Gracillariidae: Phyllocnistinae). *Tropical Lepidoptera*, **4**, 49–64.

Hinton, H.E. (1946) Concealed phases in the metamorphosis of insects. *Nature*, **157**, 552–553.

Hinton, H.E. (1971) Some neglected phases in metamorphosis. *Proceedings of the Royal Entomological Society of London, Series C*, **35**, 55–64.

Johns, C.A., Moore, M.R., Kawahara, A.Y. (2016) Molecular phylogeny, revised higher classification, and implications for conservation of endangered Hawaiian leaf-mining moths (Lepidoptera: Gracillariidae: *Philodoria*). *Pacific Science*, **70**, 361–372.

Kaiser, W., Huguet, E., Casas, J., Commin, C. & Giron, D. (2010) Plant green-island phenotype induced by leaf-miners is mediated by bacterial symbionts. *Proceedings of the Royal Society of London, Series B*, **277**, 2311–2319.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Moleculare Biology and Evolution*, **30**, 772–780.

Kawahara, A.Y. & Breinholt, J.W. (2014) Phylogenomics provides strong evidence for relationships of butterflies and moths. *Proceedings of the Royal Society of London, Series B*, **281**, 20140970.

Kawahara, A.Y., Nishida, K. & Davis, D.R. (2009) Systematics, host plants, and life histories of three new *Phyllocnistis* species from the central highlands of Costa Rica (Lepidoptera, Gracillariidae, Phyllocnistinae). *ZooKeys*, **27**, 7–30.

Kawahara, A.Y., Ohshima, I., Kawakita, A. *et al.* (2011) Increased gene sampling strengthens support for higher-level groups within leaf-mining moths and relatives (Lepidoptera: Gracillariidae). *BMC Evolutionary Biology*, **11**, 182.

Kawakita, A. & Kato, M. (2016) Revision of the Japanese species of *Epicephala* Meyrick with descriptions of seven new species (Lepidoptera, Gracillariidae). *ZooKeys*, **568**, 87–118.

Kawakita, A., Takimura, A., Terachi, T., Sota, T. & Kato, M. (2004) Cospeciation analysis of an obligate pollination mutualism: have *Glochidion* trees (Euphorbiaceae) and pollinating *Epicephala* moths (Gracillariidae) diversified in parallel? *Evolution*, **10**, 2201–2214.

Kawakita, A., Okamoto, T., Goto, R. & Kato, M. (2010) Mutualism favours higher host specificity than does antagonism in plant-herbivore interaction. *Proceedings of the Royal Society of London, Series B*, **277**, 2765–2774.

Kobayashi, S., Sakamoto, Y., Jinbo, U., Nakamura, A. & Hirowatari, T. (2011) A new willow leaf blotch miner of the genus *Phyllocnistis* (Lepidoptera: Gracillariidae: Phyllocnistinae) from Japan. *Lepidoptera Science*, **62**, 75–93.

Kobayashi, S., Huang, G.H., Nakamura, A. & Hirowatari, T. (2013) Four new species of Gracillariidae (Lepidoptera) from China and Japan, and description of the pupal morphology of the genera *Corythoxestis, Eumetriochroa, Guttigera*, and *Metriochroa*. *Zootaxa*, **3619**, 101–129.

Kumata, T. (1963) Taxonomic studies on the Lithocolletinae of Japan (Lepidoptera: Gracillariidae) part 3. *Insecta Matsumurana*, **26**, 69–88.

Kumata, T. (1965) On the species of the genera *Parornix* Spuler and *Callisto* Stephens occuring in Japan, with descriptions of two new species (Lepidoptera: Gracillariidae). *Insecta Matsumurana*, **28**, 62–68.

Kumata, T. (1978) A new stem-miner of alder in Japan, with a review of the larval transformation in the Gracillariidae (Lepidoptera). *Insecta Matsumurana, New Series*, **13**, 1–27.

Kumata, T. (1982) A taxonomic revision of the *Gracillaria* group occurring in Japan (Lepidoptera: Gracillariidae). *Insecta Matsumurana, New Series*, **26**, 1–186.

Kumata, T. (1985) A new genus of Gracillariidae, with three new species from Asia (Lepidoptera). *Insecta Matsumurana, New Series*, **32**, 109–137.

Kumata, T. (1989) Two new genera related to *Acrocercops* (Gracillariidae, Lepidoptera), with five new species from the Oriental region. *Insecta Matsumurana, New Series*, **42**, 47–82.

Kumata, T. (1993) A contribution to the knowledge of the Malaysian Lithocolletinae (Gracillariidae, Lepidoptera), with a revision of Indian *Cameraria* associated with Leguminosae. *Insecta Matsumurana, New Series*, **48**, 1–85.

Kumata, T. (1995) *Ketapangia*, a new genus for *Macarostola leucochorda* and *Acrocercops regulifera* (Gracillariidae, Lepidoptera). *Insecta Matsumurana, New Series*, **52**, 133–148.

Kumata, T. (1998) Japanese species of the subfamily Oecophyllembiinae Réal and Balachowsky (Lepidoptera: Gracillariidae), with descriptions of a new genus and eight new species. *Insecta Matsumurana, New Series*, **54**, 77–131.

Kumata, T., Kuroko, H. & Ermolaev, V.P. (1988a) Japanese species of the *Acrocercops*-group (Lepidoptera: Gracillariidae). Part I. *Insecta Matsumurana, New Series*, **38**, 1–111.

Kumata, T., Kuroko, H. & Ermolaev, V.P. (1988b) Japanese species of the *Acrocercops*-group (Lepidoptera: Gracillariidae). Part II. *Insecta Matsumurana, New Series*, **40**, 1–133.

Kuznetzov, V.I. (1979) New subgenera and species of moths (Lepidoptera, Gelechiidae, Gracillariidae) from Siberia and the Far East. *New Species of Insects from Siberia and the Far East – Trudy Zoologicheskogo Instituta*, Vol. 81, pp. 87–102. Akademija Nauk SSSR, St. Petersburg.

Kuznetzov, V.I. (1981) Gracillariidae (Lithocolletidae). *Keys of the Insects of the European Part of the USSR*, Vol. IV (ed. by G.S. Medvedev), pp. 314–410. E. J. Brill, Leiden, New York, New York, København, Köln.

Kuznetzov, V.I. & Baryshnikova, S.V. (2001) Review of the Palaearctic genera of the family Gracillariidae (Lepidoptera) with description of a new subfamily Ornixolinae Kuznetzov et Baryshnikova, subfam. n. *Entomological Review*, **81**, 26–46.

Kuznetzov, V.I. & Stekol'nikov, A.A. (1987) Functional morphology of the male genitalia and notes on the classification and phylogenetic relationships of mining moths of the superfamily Gracillarioidea (Lepidoptera). *Entomological Review*, **66**, 16–30.

Kuznetzov, V.I. & Stekol'nikov, A.A. (2001) *New Approaches to the System of Lepidoptera of World Fauna (on the base of the functional morphology of abdomen)* (ed. by E.R. Nartshuk), 462 pp. Nauka, St. Petersburg.

Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Lees, D.C., Kawahara, A.Y., Rougerie, R. *et al.* (2014) DNA barcoding reveals a largely unknown fauna of Gracillariidae leaf-mining moths in the Neotropics. *Molecular Ecology Resources*, **14**, 286–296.

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 727–744.

Liu, L. & Pearl, D.K. (2007) Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic Biology*, **56**, 504–514.

Lopez-Vaamonde, C., Charles, H., Godfray, H.C.J. & Cook, J.M. (2003) Evolutionary dynamics of host-plant use in a genus of leaf-mining moths. *Evolution*, **57**, 1804–1821.

Lopez-Vaamonde, C., Wikstrom, N., Labandeira, C., Godfray, H.C.J., Goodman, S.J. & Cook, J.M. (2006) Fossil-calibrated molecular phylogenies reveal that leaf-mining moths radiated millions of years after their host plants. *Journal of Evolutionary Biology*, **19**, 1314–1326.

Lopez-Vaamonde, C., Agassiz, D.V.L., Augustin, S. *et al.* (2010) Lepidoptera. Chapter 11*Alien Terrestrial Arthropods of Europe* (ed. by A. Roques *et al.*). *BioRisk*, **4**, 603–668.

Maddison, W.P. & Maddison, D.R. (2014) *Mesquite: A Modular System for Evolutionary Analysis. Version 3.01* [WWW document]. URL http://mesquiteproject.org [accessed on 17 August 2014].

Menken, S.B.J., Boomsma, J.J. & Van Nieukerken, E.J. (2009) Large-scale evolutionary patterns of host plant associations in the Lepidoptera. *Evolution*, **64**, 1098–1119.

Minet, J. (1986) Ébauche d'une classification moderne de l'ordre des Lépidoptères. *Alexanor*, **14**, 291–313.

Misof, B., Liu, S.L., Meusemann, K. *et al.* (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, **346**, 763–767.

van Nieukerken, E.J., Kaila, L., Kitching, I.J. *et al.* (2011) Order Lepidoptera Linnaeus, 1758. *Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness* (ed. by Z.-Q. Zhang). *Zootaxa*, **3148**, 212–221.

Nye, I.W.B. & Fletcher, D.S. (1991) *Microlepidoptera. The Generic Names of the Moths of the World*, Vol. 6. Natural History Museum Publications, London.

Nyman, T. (2010) To speciate, or not to speciate? Resource heterogeneity, the subjectivity of similarity, and the macroevolutionary consequences of niche-width shifts in plant-feeding insects. *Biological Reviews*, **85**, 393–411.

Nyman, T., Farrell, B.D., Zinovjev, A.G. & Vikberg, V. (2006) Larval habits, host-plant associations, and speciation in Nematine sawflies (Hymenoptera: Tenthredinidae). *Evolution*, **60**, 1622–1637.

Ohshima, I. (2008) Host race formation in the leaf-mining moth *Acrocercops transecta* (Lepidoptera: Gracillariidae). *Biological Journal of the Linnean Society*, **93**, 135–145.

Ohshima, I. (2012) Genetic mechanisms preventing the fusion of ecotypes even in the face of gene flow. *Scientific Reports*, **2**, 506. DOI: 10.1038/srep00506.

Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Worheide, G. & Baurain, D. (2011) Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*, **9**, 402.

Powell, J.A., Mitter, C. & Farrell, B.D. (1998) Evolution of larval food preferences in Lepidoptera. *Lepidoptera, Moths and Butterflies: Evolution, Systematics and Biogeography, Handbook of Zoology*, Vols. 1, 4, Part 35 (ed. by N.P. Kristensen), pp. 403–422. De Gruyter, Berlin.

Rambaut, A. (2014) *FigTree Version 1.4.2*. University of Edinburgh, Edinburgh. [WWW document]. URL http://tree.bio.ed.ac.uk/ [accessed on 22 August 2015].

Regier, J.C. (2008) *Protocols, Concepts, and Reagents for Preparing DNA Sequencing Templates, Version 12 April 2008* [WWW document]. URL http://www.umbi.umd.edu/users/jcrlab/PCR_primers.pdf [accessed on 20 August 2015]

Regier, J.C., Shultz, J.W., Ganley, A.R. *et al.* (2008) Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Systematic Biology*, **57**, 920–938.

Regier, J., Zwick, A., Cummings, M. *et al.* (2009) Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. *BMC Evolutionary Biology*, **9**, 280.

Regier, J.C., Shultz, J.W., Zwick, A. *et al.* (2010) Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature*, **463**, 1079–1083.

Regier, J.C., Mitter, C., Zwick, A. *et al.* (2013) A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS ONE*, **8**, e58568.

Roure, B., Baurain, D. & Philippe, H. (2013) Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, **30**, 197–214.

Sanderson, M.J. & Shaffer, H.B. (2002) Troubleshooting molecular phylogenetic analyses. *Annual Review of Ecology and Systematics*, **33**, 49–72.

Shapiro, L.H., Scheffer, S.J., Maisin, N. *et al.* (2008) *Conopomorpha cramerella* (Lepidoptera: Gracillariidae) in the Malay Archipelago: genetic signature of a bottlenecked population? *Annals of the Entomological Society of America*, **101**, 930–938.

Shimodaira, H. (2002) An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, **51**, 492–508.

Shimodaira, H. & Hasegawa, M. (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, **17**, 1246–1247.

Sohn, J.C., Regier, J.C., Mitter, C., Davis, D., Landry, J.F., Zwick, A. & Cummings, M.P. (2013) A molecular phylogeny for Yponomeutoidea (Insecta, Lepidoptera, Ditrysia) and its implications for classification,

biogeography and the evolution of host plant use. *PLoS ONE*, **8**, e55066.

Spuler, A. (1910) *Die Schmetterlinge Europas. II*. Schweizerbarthsche Verlagshandlung, Stuttgart.

Stainton, H.T. (1854) *Lepidoptera: Tineina*. Lovell Reeve, London.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Treitschke, F. (1833) *Die Schmetterlinge von Europa*, Vol. 9 (2). Ernst Fleischer, Leipzig.

Triberti, P. (1998) Remarks on the phylogeny of the genera *Parornix* Spuler and *Callisto* Stephens (Lepidoptera Gracillariidae). *Bolletino Museo Civico di Storia naturale, Verona*, **22**, 175–197.

Vargas, H.A. & Landry, B. (2005) A new genus and species of Gracillariidae (Lepidoptera) feeding on flowers of *Acacia macracantha* Willd. (Mimosaceae) in Chile. *Acta Entomológica Chileana*, **29**, 47–57.

Vargas, H.A. & Parra, L.E. (2005) Un nuevo género y una nueva especie de Oecophyllembiinae (Lepidoptera: Gracillariidae) de Chile. *Neotropical Entomology*, **34**, 227–233.

Vári, L. (1961) South African Lepidoptera. Vol. 1. Lithocolletidae. *Transvaal Museum Memoires*, **12**, 1–238.

Vári, L., Kroon, D.M. & Krüger, M. (2002) *Classification and Checklist of the Species of Lepidoptera Recorded in Southern Africa*. Simple Solutions, Chatswood.

Wagner, D.L., Loose, J.L., Fitzgerald, T.D., De Benedictis, J.A. & Davis, D.R. (2000) A hidden past: the hypermetamorphic development of *Marmara arbutiella* (Lepidoptera: Gracillariidae). *Annals of the Entomological Society of America*, **93**, 59–64.

Walsingham, L.T. de G. (1909–1915) Insecta. Lepidoptera-Heterocera.Tineina, Pterophorina, Orneodina, and Pyralidina and Hepialina (part.). *Biologia Centrali-Americana, Lepidoptera-Heterocera*, Vol. 4, pp. 225–392, 9–10 pls. Published for the editors by R. H. Porter, London, U.K.

Wilkinson, M. (1995) More on reduced consensus methods. *Systematic Biology*, **44**, 435–439.

Zeller, P.C. (1839) Versuch einer naturgemässen Eintheilung der Schaben. *Isis von Oken*, **3**, 168–220.

Zimmerman, E.C. (1978) *Insects of Hawaii*, *Volume 9. Microlepidoptera, Part 1. Monotrysia, Tineoidea, Tortricoidea, Gracillarioidea, Yponomeutoidea, and Alucitoidea*. University of Hawaii Press, Honolulu, Hawaii.

Zhang, J., Wang, S., Li, H., Hu, B., Yang, X. & Wang, Z. (2012) Diffuse coevolution between two *Epicephala* species (Gracillariidae) and two *Breynia* species (Phyllanthaceae). *PLoS ONE*, **7**, e41657.

Zwick, A. (2010) *Degeneracy Coding Web Service. PhyloTools.* 10 February 2010 [WWW document]. URL http://www.phylotools.com/ptdegen1webservice.htm [accessed on 22 August 2015]

Zwickl, D.J. (2006) *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. PhD Thesis, The University of Texas, Austin, Texas.

Accepted 8 May 2016
First published online 13 October 2016