



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Phylogenomic analysis of protein-coding genes resolves complex gall wasp relationships

Citation for published version:

Hearn, J, Gobbo, E, Nieves-Aldrey, JL, Branca, A, Nicholls, JA, Koutsovoulos, G, Lartillot, N, Stone, GN & Ronquist, F 2023, 'Phylogenomic analysis of protein-coding genes resolves complex gall wasp relationships', *Systematic entomology*. <https://doi.org/10.1101/2022.06.20.496719>, <https://doi.org/10.1111/syen.12611>

Digital Object Identifier (DOI):

[10.1101/2022.06.20.496719](https://doi.org/10.1101/2022.06.20.496719)
<https://doi.org/10.1111/syen.12611>

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Systematic entomology

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Phylogenomic analysis of protein-coding genes resolves complex gall wasp relationships

Jack Hearn¹  | Erik Gobbo^{2,3}  | José Luis Nieves-Aldrey⁴  |
Antoine Branca⁵  | James A. Nicholls⁶  | Georgios Koutsovoulos⁷  |
Nicolas Lartillot⁸  | Graham N. Stone⁷  | Fredrik Ronquist^{2,3} 

¹Centre for Epidemiology and Planetary Health, Scotland's Rural College, Inverness, UK

²Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

³Department of Zoology, Stockholm University, Stockholm, Sweden

⁴Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), Madrid, Spain

⁵UMR Évolution, Génomes, Comportement et Écologie, IRD, CNRS, Université Paris-Saclay, Gif-sur-Yvette, France

⁶Australian National Insect Collection, CSIRO, Canberra, Australia

⁷Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, UK

⁸Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard Lyon 1, Lyon, France

Correspondence

Fredrik Ronquist, Department of Zoology,
Stockholm University, Stockholm, Sweden.
Email: fredrik.ronquist@nrm.se

Funding information

Agence Nationale de la Recherche,
Grant/Award Numbers: ANR-
15-CE12-0010-01/DASIRE, ANR-
19-CE02-0008; H2020 Marie Skłodowska-
Curie Actions, Grant/Award Number: 642241;
Ministerio de Economía y Competitividad/
FEDER, Grant/Award Number: UE
CGL2015-66571-P; Natural Environment
Research Council, Grant/Award Numbers: NE/
J010499, NBAF375; Vetenskapsrådet,
Grant/Award Numbers: 2018-05973,
2018-04620; National Genomics
Infrastructure and Science for Life Laboratory,
Grant/Award Number: P14912; Swedish
National Infrastructure for Computing at
UPPMAX, Grant/Award Numbers: 2017097,
2017/7-283; NSC, Grant/Award Numbers:
2021/23-157, 2021/5-118

Abstract

Gall wasps (Hymenoptera: Cynipidae) comprise 13 distinct tribes whose interrelationships remain incompletely understood. Recent analyses of ultra-conserved elements (UCEs) represent the first attempt at resolving these relationships using phylogenomics. Here, we present the first analysis based on protein-coding sequences from genome and transcriptome assemblies. Unlike UCEs, these data allow more sophisticated substitution models, which can potentially resolve issues with long-branch attraction. We include data for 37 cynipoid species, including two tribes missing in the UCE analysis: Aylacini (s. str.) and Qwaqwaiini. Our results confirm the UCE result that Cynipidae are not monophyletic. Specifically, the Paraulacini and Diplolepidini + Pediaspidini fall outside a core clade (Cynipidae s. str.), which is more closely related to the insect-parasitic Figitidae, and this result is robust to the exclusion of long-branch taxa that could mislead the analysis. Given this, we here divide the Cynipidae into three families: the Paraulacidae stat. prom., Diplolepididae stat. prom. and Cynipidae (s. str.). Our results suggest that the Eschatocerini are the sister group of the remaining Cynipidae (s. str.). Within the Cynipidae (s. str.), the Aylacini (s. str.) are more closely related to oak gall wasps (Cynipini) and some of their inquiline (Ceroptresini) than to other herb gallers (Aulacideini and Phanacidini), and the Qwaqwaiini likely form a clade together with Synergini (s. str.) and Rhoophilini. Several alternative scenarios for the

Jack Hearn and Erik Gobbo contributed equally to this study.

Graham N. Stone and Fredrik Ronquist also contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Systematic Entomology* published by John Wiley & Sons Ltd on behalf of Royal Entomological Society.

evolution of cynipid life histories are compatible with the relationships suggested by our analysis, but all are complex and require multiple shifts among parasitoids, inquilines and gall inducers.

KEYWORDS

gall inducers, gall wasps, inquilines, life history transitions, parasitoids, phylogenomics, protein-coding genes

INTRODUCTION

Gall wasps (Hymenoptera: Cynipidae) induce the development of highly modified plant tissues, termed galls, in which their immature stages develop (Melika & Abrahamson, 2002; Stone et al., 2002). The cynipid larva is enclosed inside a gall chamber lined with specialised nutritive cells formed by the plant in response to signals released by the gall wasp egg and larva (Harper et al., 2004; Hearn et al., 2019; Stone & Schönrogge, 2003). While all cynipids that appear induce the development of such nutritive tissues, several lineages—termed inquilines—can only induce nutritive tissue development within galls initiated by other species (Sanver & Hawkins, 2000). The inquilines can thus be seen as cynipids that induce a ‘gall within a gall’. The presence of inquilines can negatively affect the fitness of the primary gall inducer, in many cases killing it (László & Tóthmérész, 2006). Female *Periclistus* Förster inquilines have been reported to kill the larva of the inducing *Diplolepis* gall wasp by stabbing it with their ovipositor, potentially injecting harmful substances in the process (Shorthouse, 1973).

Several hypotheses on the mechanism of cynipid gall induction have been advanced, partly inspired by knowledge of other gall-inducing organisms: secretion of auxins (Tooker & Helms, 2014), injection of virus-like particles (Cambier et al., 2019; Cornell, 1983), manipulation of plant nodulation factors (signalling molecules inducing nodules in legume roots; Hearn et al., 2019) or involvement of bacterial or fungal symbionts (Hearn et al., 2019). However, in contrast to some other gall induction systems (Harris & Pitzschke, 2020), there is no conclusive evidence for any of these hypotheses in cynipids, even though our understanding of the changes in gene expression patterns in early gall development is quickly increasing (Cambier et al., 2019; Hearn et al., 2019; Martinson et al., 2022).

Our understanding of the evolutionary origin of cynipid gall inducers and inquilines is equally poor. It has generally been assumed that the phytophagous forms constitute a monophyletic lineage, the family Cynipidae, although it has been surprisingly difficult to find morphological characters supporting their monophyly (Liljeblad & Ronquist, 1998; Ronquist, 1999; Ronquist et al., 2015). The Cynipidae are deeply nested within the insect-parasitic Apocrita (Blaimer et al., 2023; Heraty et al., 2011; Klopstein et al., 2013; Peters et al., 2017; Ronquist, 1995, 1999; Sharkey et al., 2011), and all other members of the superfamily Cynipoidea are insect parasitoids, so it has long been clear that the phytophagous gall inducers and inquilines must have evolved from insect-parasitic ancestors. Except for the Cynipidae, the superfamily Cynipoidea comprises the families

Austrocynipidae, Ibalidae, Liopteridae and Figitidae (Ronquist, 1995, 1999). The life history of several species of ibalids and figitids is well-studied (Ronquist, 1999, and references cited therein). They are all koinobiont endoparasitoids in the early larval instars. Towards the end of their development, they emerge and consume the remains of the moribund host as ectoparasitoids. The most diverse lineage is the Figitidae, which has appeared as the sister group of the Cynipidae in most previous analyses (Buffington et al., 2007, 2012; Ronquist, 1995, 1999; Ronquist et al., 2015; but see Blaimer et al., 2020).

The origin of the Cynipidae appears to be linked to that of several lineages of gall-associated Figitidae, which appear to form early-diverging lineages in the family (Blaimer et al., 2020; Buffington et al., 2007, 2012; Ronquist, 1995, 1999; Ronquist et al., 2015). These gall-associated figitids include the Parnipinae (Ronquist & Nieves-Aldrey, 2001), Plectocynipinae (Buffington & Nieves-Aldrey, 2011; Ros Farre & Pujade-Villar, 2007), Thrasorinae (Buffington, 2008; Paretas-Martínez et al., 2011), Mikeiinae (Paretas-Martínez et al., 2011) and Euceroptresinae (Buffington & Liljeblad, 2008; note that the subfamily name should be Euceroptresinae and not Euceroptrinae). There is fairly strong evidence that the Parnipinae are koinobiont early-internal-late-external parasitoids of cynipid gall inducers in the genera *Barbotinia* Nieves-Aldrey and *Iraella* Nieves-Aldrey (Ronquist et al., 2018). The life history of the other lineages remains unclear, although they are generally assumed to be parasitoids of other inhabitants in the cynipid and chalcidoid galls from which they have been reared.

Ashmead (1903) divided the Cynipidae into six tribes: the Cynipini, Diplolepidini (or Rhoditini), Pediaispini, Eschatocerini, Aylacini (s. lat.) and Synergini (s. lat.). Of these, all but the Eschatocerini are Holarctic. The Cynipini, which induce galls on oaks (*Quercus* L.) and other Fagaceae, are one of the largest radiations of insect gall inducers with more than 1000 described species (Stone et al., 2002). The Diplolepidini induce galls on roses (*Rosa* L.) and include the well-known bedeguar gall wasp, *Diplolepis rosae* (L.). The Pediaispini and Eschatocerini are two small tribes, originally including a single genus each: *Pediaispis* Tischbein, a European genus inducing galls on maples (*Acer* L.), and *Eschatocerus* Mayr, a South American genus associated with galls on *Vachellia* Wight & Arn. (previously *Acacia*; Maslin et al., 2003) and other woody Fabaceae. The inquilines are grouped in this system into the Synergini, and the remaining gall inducers, mostly associated with herbaceous host plants, in the Aylacini.

Early phylogenetic analyses based on morphology suggested that the Aylacini (s. lat.) form a paraphyletic assemblage of early-diverging cynipid lineages (Liljeblad & Ronquist, 1998; Ronquist, 1994),

consistent with ideas presented over a century ago by cynipidologist and later famous sexologist Alfred Kinsey (Kinsey, 1920). They also placed the genus *Himalocynips*, a cynipid from Nepal with unknown biology and originally placed in a separate subfamily (Yoshimoto, 1970), in Pediapidini (Liljeblad & Ronquist, 1998).

Subsequent analyses of molecular data and combined molecular, morphological and life history data (Nylander et al., 2004; Ronquist et al., 2015) have revealed that the Synergini (s. lat.) are not monophyletic either. Specifically, *Periclistus* (inquilines in *Diplolepis* Geoffroy galls on *Rosa*) and *Synophromorpha* Ashmead (inquilines in *Diastrophus* Hartig galls on *Rubus* L.) are closely related to the Aylacini (s. lat.) genera *Diastrophus* and *Xestophanes* Förster, forming a lineage (now recognised as the Diastrophini) of gall inducers and inquilines associated with herbaceous and woody hosts in the Rosaceae. The remaining Aylacini (s. lat.) fall into three distinct lineages: (1) the Aylacini (s. str.), gall inducers associated with poppies (*Papaver* L.); (2) the Aulacideini, gall inducers mostly associated with Asteraceae and Lamiaceae, but also with a few other families, including Papaveraceae; and (3) the Phanacidini, gall inducers mainly associated with Asteraceae and Lamiaceae, and often inducing stem galls. The remaining inquilines fall into two distinct monophyletic lineages: (1) the Ceroptresini, including the single genus *Ceroptres* Hartig associated with Cynipini oak galls, and (2) and a lineage comprising the remaining inquilines of Cynipini oak galls and *Rhoophilus* Mayr, a South African inquiline in lepidopteran galls on species of *Searsia* F. A. Barkley (Anacardiaceae) (van Noort et al., 2007). Several analyses have supported a sister-group relationship between *Rhoophilus* and the remaining members of the clade (Ide et al., 2018; Liljeblad & Ronquist, 1998; Ronquist et al., 2015), and recently, it was proposed to recognise separate tribes for *Rhoophilus*, the Rhoophilini (Lobato-Vila et al., 2022) and the

remaining members, the Synergini (s. str.) (Table 1). Recent work has also shown that the latter, previously considered to consist entirely of inquilines, includes at least one deeply nested lineage comprising *Synergus itoensis* Abe, Ide & Wachi and related species that induce galls de novo inside acorns (Abe et al., 2011; Gobbo et al., 2020; Ide et al., 2018).

In recent years, two additional lineages associated with woody host plants have been added: (1) the tribe Qwaqwaiini, based on a newly discovered gall inducer on *Scolopia* Schreb. (Salicaceae) in South Africa (Liljeblad et al., 2011), and (2) the Paraulacini, a lineage of temperate South American cynipids associated with galls on southern beeches (*Nothofagus* Blume; Nothofagaceae) (Nieves-Aldrey et al., 2009). Somewhat surprisingly, a member of the Paraulacini was recently shown to be a parasitoid (Rasplus et al., 2022; see also *Discussion* below).

In summary, current classification of the Cynipidae comprises 13 tribes (Table 1). The most recent 'old-school' analysis of these relationships based on combined data from five molecular markers, morphology and life history data (Ronquist et al., 2015) failed to resolve relationships among these tribes, with three notable exceptions (Figure 1a): (1) Diplolepidini and Pediapidini, both gallers of woody-rosid host plants, are sister groups; (2) the two major lineages of herb gallers, the Aulacideini and Phanacidini, are sister groups; and (3) the Rhoophilini and Synergini (s. str.) are sister groups, as mentioned above. Many of the tribes appear to represent isolated lineages, with no close relatives among the other tribes (Nylander et al., 2004; Ronquist et al., 2015).

Recently, Blaimer et al. (2020) used phylogenomic analysis of ultra-conserved element (UCE) DNA sequence data (Faircloth et al., 2012) to infer relationships among a phylogenetically diverse

TABLE 1 Life history of the 13 tribes of Cynipidae recognised currently (Lobato-Vila et al., 2022; Ronquist et al., 2015).

Tribe	Distribution	Life history	Host
Aulacideini	Holarctic	Gall inducers	Herbs in the families Asteraceae, Lamiaceae and others
Aylacini (s. str.)	Holarctic	Gall inducers	<i>Papaver</i> L. (Papaveraceae)
Ceroptresini	Holarctic	Inquilines	Galls of Cynipini on oaks
Cynipini	Holarctic	Gall inducers	<i>Quercus</i> L., occasionally related genera (Fagaceae)
Diastrophini	Holarctic	Gall inducers and inquilines	Gall inducers on Rosaceae (<i>Rubus</i> L. and <i>Potentilla</i> L.), and inquilines in cynipid galls on Rosaceae (<i>Rubus</i> and <i>Rosa</i> L.)
Diplolepidini	Holarctic	Gall inducers	<i>Rosa</i> L. (Rosaceae)
Eschatocerini	Neotropical	Gall inducers (see text)	<i>Vachellia</i> Wight & Arn., <i>Prosopis</i> Fabricius (Fabaceae)
Paraulacini	Neotropical	Parasitoids (see text)	<i>Aditrochus</i> Rübsaamen (Pteromalidae) gall inducers on <i>Nothofagus</i> Blume (Nothofagaceae, Fagales)
Pediapidini	Palaeartic	Gall inducers	<i>Acer</i> L. (Sapindaceae)
Phanacidini	Holarctic	Gall inducers	Herbs, mostly in the families Asteraceae and Lamiaceae
Qwaqwaiini	Afrotropical	Gall inducers	<i>Scolopia</i> Schreb. (Salicaceae)
Rhoophilini	Afrotropical	Inquilines	Lepidoptera galls on <i>Searsia</i> Parr (Anacardiaceae)
Synergini (s. str.)	Holarctic	Inquilines; a few gall inducers	Galls of Cynipini; a few are true gall inducers on <i>Quercus</i> (Fagaceae)

set of cynipoids. Their taxon set includes representatives of all families except the Austrocynipidae and spanned a significant amount of the known diversity within each family. Several surprising results emerged from this UCE analysis (Figure 1b). First, the Liopteridae and Ibalidae were placed within the Figitidae (s. lat.), among the early-diverging gall-associated lineages. Second, the Paraulacini and the Diplolepidini + Pediaspidini were placed outside the clade formed by the Figitidae and the remaining Cynipidae (the Cynipidae s. str.), a relationship first hinted at in Hymenoptera-wide analyses (Peters et al., 2017; see also Blaimer et al., 2023) and indicating that herbivorous cynipids do not form a monophyletic clade. Finally, the analysis suggested that the Eschatocerini may be the sister group of the Figitidae, although the evidence for this was weak and alternative placements appeared under some analysis settings. Although these results seemingly conflict substantially with those of Ronquist et al. (2015), all robustly resolved relationships among major lineages are actually consistent between the studies, except for the different rooting of the cynipoid tree and a minor difference in the position of Euceroptresinae (Figure 1). The major advances in the UCE analysis are the drastically different rooting and the increased resolution of relationships among major lineages of figitids and cynipids.

The analysis we present here is the first phylogenomic analysis based on genome and transcriptome assemblies, and it allows a largely independent test of the results from the UCE analysis. In contrast to the UCE analysis, our taxon sampling is focused on cynipids. It lacks ibaliids and liopterids, and is relatively sparse with respect to figitids. However, it includes representatives of all cynipid tribes except the recently recognised Rhoophilini. Importantly, it includes the Qwaqwaiini and Aylacini (s. str.), both of which were missing from the UCE analysis. Blaimer et al. (2020) used *Aylax salviae* Giraud as their only representative of the tribe Aylacini (s. str.). However, this species has long been placed in the genus *Neaylax* Nieves-Aldrey (Nieves-Aldrey, 1994, 2001), which does not belong to the Aylacini (s. str.) but instead is deeply nested inside the Aulacideini (Ronquist et al., 2015). Specifically, *Neaylax salviae* belongs to a clade of Aulacideini gallers of Lamiaceae related to the genus *Antistrophus* Walsh (Ronquist et al., 2015), and this is entirely consistent with the placement of 'Aylax' *salviae* in the UCE analysis (Blaimer et al., 2020). Another key taxon represented in our analysis but not in the UCE analysis is the single species in the recently described genus *Protobalandricus* Nicholls, Stone & Melika, *P. spectabilis* (Kinsey), which represents a divergent sister group to all other sampled Cynipini (Nicholls et al., 2018).

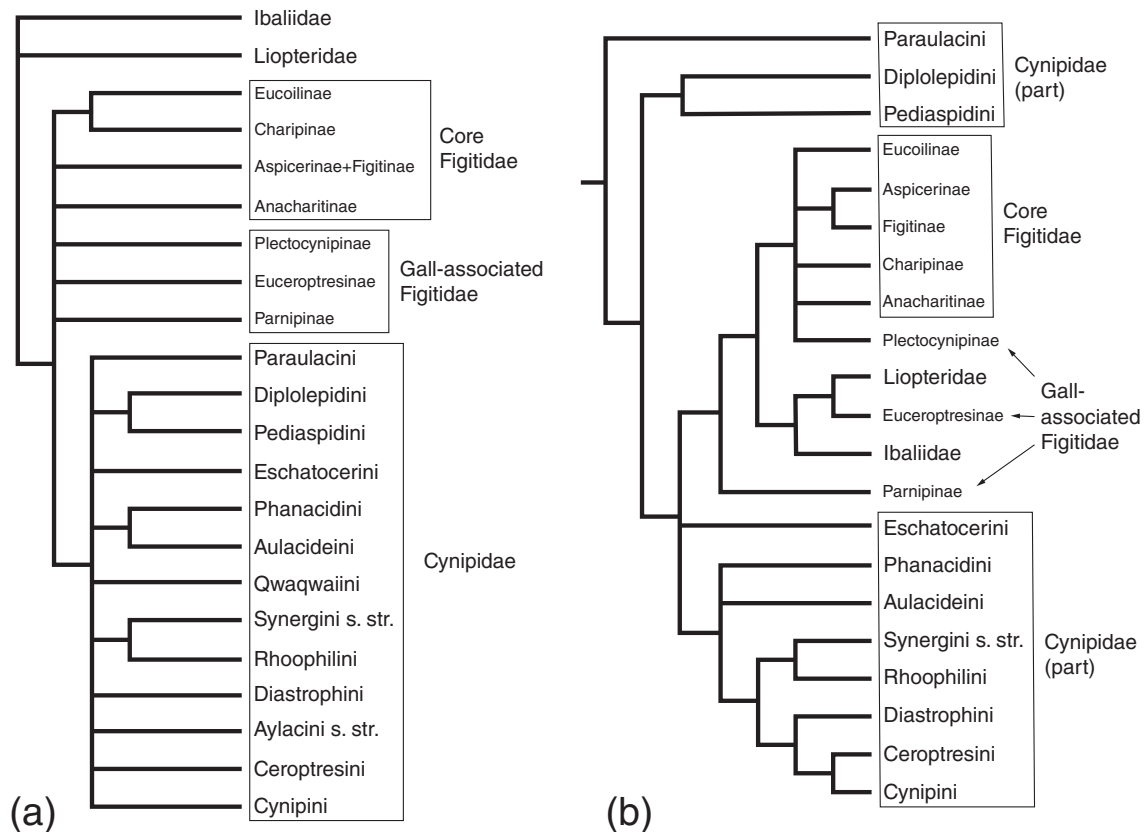


FIGURE 1 Recent hypotheses of cynipid relationships. (a) Combined analysis of data from five molecular markers, morphology and life history (Ronquist et al., 2015: fig. 2). The tree was rooted on Ibalidae; the analysis lacked more distant outgroups. (b) Phylogenomic analysis of 1147 UCE loci (Blaimer et al., 2020: fig. 1). The tree was rooted using multiple outgroups (indicated by the stalk). Some figitid lineages lacking in Ronquist et al. (2015) removed to facilitate comparison. In both cases, we only show clades with at least 95% support (posterior probability or bootstrap support). Figitid lineages are shown in smaller font.

TABLE 2 Overview of the species included in this study.

Higher taxon	Species	Life history	Type
Cynipidae			
Aulacideini	<i>Aulacidea tavakolii</i> Melika in Karimpour, Tavakoli & Melika	Gall inducer on <i>Tragopogon</i> L. (Asteraceae)	NG
	<i>Fumariphilus hypecoi</i> (Trotter)	Gall inducer on <i>Hypecoum</i> L. (Papaveraceae)	NG
	<i>Hedickiana levantina</i> (Hedicke)	Gall inducer on <i>Salvia</i> L. (Lamiaceae)	NG
	<i>Isocolus centaureae</i> Dyakonchuk	Gall inducer on <i>Centaurea</i> L. (Asteraceae)	NG
Ayalacini (s. str.)	<i>Iraella hispanica</i> Nieves-Aldrey	Gall inducer on <i>Papaver</i> L. (Papaveraceae)	NG
Ceroptresini	<i>Ceroptres masudai</i> Abe	Inquiline in <i>Andricus</i> Hartig galls on <i>Quercus</i> L.	NG
Cynipini	<i>Andricus curvator</i> Hartig	Gall inducer on <i>Quercus</i> (Fagaceae)	NG
	<i>A. grossulariae</i> Giraud	Gall inducer on <i>Quercus</i> (Fagaceae)	G
	<i>A. quercusramuli</i> (L.)	Gall inducer on <i>Quercus</i> (Fagaceae)	G
	<i>Belonocnema kinseyi</i> Weld	Gall inducer on <i>Quercus</i> (Fagaceae)	NG
	<i>Biorhiza pallida</i> (Olivier)	Gall inducer on <i>Quercus</i> (Fagaceae)	G
	<i>Druon quercuslanigerum</i> (Ashmead)	Gall inducer on <i>Quercus</i> (Fagaceae)	T
	<i>Neuroterus valhalla</i> Brandão-Dias, Zhang, Pirro, Vinson, Weinersmith, Ward, Forbes & Egan	Gall inducer on <i>Quercus</i> (Fagaceae)	G
	<i>Protobalandricus spectabilis</i> (Kinsey)	Gall inducer on <i>Quercus</i> (Fagaceae)	NG
Diastrophini	<i>Diastrophus kincaidii</i> Gillette	Gall inducer on <i>Rubus</i> L. (Rosaceae)	NG
	<i>Periclistus</i> Förster sp.	Inquiline in <i>Diplolepis</i> Geoffroy galls on <i>Rosa</i> L.	NG
Diplolepidini	<i>Diplolepis spinosa</i> (Ashmead)	Gall inducer on <i>Rosa</i> (Rosaceae)	NG
Eschatocerini	<i>Eschatocerus acacia</i> Mayr	Gall inducer on <i>Vachellia</i> Wight & Arn. (Fabaceae)	NG
Paraulacini	<i>Cecinothofagus ibarrai</i> Nieves-Aldrey & Liljeblad	Parasitoid of <i>Aditrochus</i> Rübsaamen (Chalcidoidea) in galls on <i>Nothofagus</i> Blume (Nothofagaceae)	NG
Pediaspidini	<i>Pediaspis aceris</i> (Gmelin)	Galls on <i>Acer</i> L. (Aceraceae)	NG
Phanacidini	<i>Phanacis</i> Förster sp.	Gall inducer on Asteraceae	NG
Qwaqwaiini	<i>Qwaqwaia scolopiae</i> Liljeblad, Nieves-Aldrey & Melika	Gall inducer on <i>Scolopia</i> Schreb. (Salicaceae)	NG
Synergini	<i>Synergus gifuensis</i> Ashmead	Inquiline in <i>Andricus</i> galls on <i>Quercus</i>	G
	<i>S. itoensis</i> Abe, Ide & Wachi	Gall inducer on <i>Quercus</i> (Fagaceae)	G
	<i>S. japonicus</i> Walker	Inquiline in <i>Andricus</i> galls on <i>Quercus</i>	G
	<i>S. umbraculus</i> (Olivier)	Inquiline in <i>Andricus</i> galls on <i>Quercus</i>	G
Figitidae			
Aspicerinae	<i>Callaspidia notata</i> (Reinhard)	Parasitoid of syrphid larvae (Diptera) feeding on aphids	NG
Charipinae	<i>Alloxysta arcuate</i> (Kieffer)	Hyperparasitoid of aphidiine braconids in aphids	NG
	<i>Phaenoglyphis villosa</i> (Hartig)	Hyperparasitoid of aphidiine braconids in aphids	NG
Eucoilinae	<i>Ganaspis</i> Förster sp.	Parasitoid of Diptera larvae	T
	<i>Leptopilina bouldardi</i> Barbotin, Carton & Keiner-Pillault	Parasitoid of <i>Drosophila</i> Fallén larvae (Diptera)	G
	<i>L. clavipes</i> (Hartig)	Parasitoid of <i>Drosophila</i> larvae (Diptera)	G
	<i>L. heterotoma</i> (Thomson)	Parasitoid of <i>Drosophila</i> larvae (Diptera)	G
Parnipinae	<i>Parnips nigripes</i> (Barbotin)	Parasitoid of <i>Barbotinia</i> Nieves-Aldrey and <i>Iraella</i> Nieves-Aldrey (Aylacini s. str.) in galls on <i>Papaver</i> (Papaveraceae)	NG
Outgroups			
Braconidae:	<i>Microplitis demolitor</i> Wilkinson	Parasitoid of Lepidoptera larvae	G
Microgastrinae			
Chalcidoidea:	<i>Nasonia vitripennis</i> (Walker)	Parasitoid of Calliphoridae and Sarcophagidae larvae (Diptera)	G
Pteromalidae			
Orussidae	<i>Orussus abietinus</i> (Scopoli)	Parasitoid of Coleoptera and Hymenoptera larvae in wood	G

Note: For full data on the genome and transcriptome assemblies, see Table S1. *Neuroterus valhalla* reference was recorded as *Callirhytis* sp. in the NCBI assembly, and *Belonocnema kinseyi* was recorded under the previous name *B. treatae* Mayr (Zhang et al., 2021). See Table S1, for accession numbers, references and assembly quality statistics.

Abbreviations: G, previously published genome assembly; NG, new genome assembly generated for this study; T, previously published transcriptome.

Importantly, by focusing on data from protein-coding genes, we can use sophisticated substitution models that accommodate variation in amino acid profiles across sites. These models are known to resolve some issues with long-branch attraction that can affect analyses under standard models, such as those used in the UCE analysis (Kapli & Telford, 2020). The most surprising UCE results do involve the placement of long, isolated lineages—the Paraulacini, Eschatocerini and Diplolepidini + Pediaspidini—and the tree is rooted with distant outgroups. It is therefore possible that long-branch attraction may be an issue. Based on the results of our analysis, which largely confirm and complement the UCE results, we propose a new family-level classification of the Cynipidae. We also discuss the implications of our findings for inference of the evolutionary origin of cynipoid gall inducers and inquilines.

MATERIALS AND METHODS

Taxon sampling

Thirty seven species were chosen to represent all of the currently recognised tribes of cynipid gall wasps except Rhoophilini, and to include as much of the phylogenetic diversity within each lineage as possible (Table 2). New genomic data were generated for 21 of the 37 species (Table 2). Our Cynipini selection included *Protobalandricus spectabilis*, inducing galls on *Quercus* section *Protobalanus* (Trelease) A. Camus oaks in California and seven other species from diverse Cynipini genera: *Andricus* Hartig, *Belonocnema* Mayr, *Biorhiza* Westwood, *Druon* Kinsey and *Neuroterus* Hartig. In the Aulacideini, we included two gallers of Asteraceae (*Isocolus centaureae* Dyakonchuk and *Aulacidea tavakolii* Melika), one galler of Lamiaceae (*Hedickiana levantina* (Hedicke)) and one galler of Papaveraceae (*Fumariphilus hypecoi* (Trotter)), thus covering much of the diversity in host plant preferences in the group. The last species, *F. hypecoi*, was originally placed in the genus *Aylax* Hartig, but has been known for some time to belong to the Aulacideini (Nieves-Aldrey, 2022; Ronquist et al., 2015) rather than the Aylacini (s. str). Our Diastrophini selection included one inquiline (*Periclistus*) and one gall inducer (*Diastrophus*), covering both of the major life history strategies in the tribe. Our Synergini (s. str.) selection was unfortunately restricted to the most species-rich genus, *Synergus* Hartig, but it included both a gall inducer (*S. itoensis*) and three inquilines (*S. gifuensis* Ashmead, *S. japonicus* Walker and *S. umbraculus* (Olivier)). The remaining eight tribes were represented by single species; most of these tribes include few species and have uniform life histories (Table 1). The selection of exemplars for this study was completed before the appearance of the recent UCE study (Blaimer et al., 2020), but it does cover all major cynipid lineages detected in that analysis.

Our sampling included eight species covering the entire diversity of Figitidae (Table 2). Importantly, our selection included *Parnips nigripes* (Barbotin) (Parnipinae), the only gall-associated figitid whose life history is known in some detail (Ronquist et al., 2018). The Parnipinae have appeared in previous analyses as the sister group

of the remaining Figitidae, or even as the sister group of the Cynipidae (Blaimer et al., 2020; Buffington et al., 2007; Ronquist, 1999; Ronquist et al., 2015). We also included three more distant outgroups: *Orussus abietinus* (Scopoli) (Orussidae), *Nasonia vitripennis* (Walker) (Chalcidoidea: Pteromalidae) and *Microplitis demolitor* Wilkinson (Braconidae: Microgastrinae) (Table 2). Of those, *O. abietinus* is the most distant (Heraty et al., 2011; Klopstein et al., 2013; Peters et al., 2017; Sharkey et al., 2011) and was used for rooting the trees generated in our analyses.

Genome and transcriptome data

Our analysis included 21 de novo genome assemblies generated for this study (marked by NG in Table 2). Two publicly available transcriptomes were included, for the oak gall wasp *Biorhiza pallida* (Olivier) (Hearn et al., 2019) and the figitid *Ganaspis* Förster species 1 (Mortimer et al., 2013). Genome assemblies for the oak gall wasps *Andricus grossulariae* Giraud, *Belonocnema kinseyi* Weld, *Druon quercuslanigerum* (Ashmead) and *Neuroterus valhalla* Brandão-Dias, Zhang, Pirro, Vinson, Weinersmith, Ward, Forbes & Egan (Brandão-Dias et al., 2022), four species of *Synergus* (Gobbo et al., 2020), three species of the figitid genus *Leptopilina* Förster and three outgroups (*Nasonia vitripennis*, *Microplitis demolitor* and *Orussus abietinus*) were downloaded from NCBI. References to all genome and transcriptome assemblies are provided in the Table S1.

De novo genome assemblies

Two protocols were followed (Table S1). For the *Andricus curvator* Hartig and *A. quercusramuli* (L.) assemblies, DNA was extracted from single adults using the Thermo Scientific KingFisher Cell and Tissue DNA Kit and the KingFisher Duo magnetic particle processor. Genomes were sequenced by the Swedish National Genomic Infrastructure from ChromiumX libraries (Zheng et al., 2016) on a NovaSeq6000 (NovaSeq Control Software 1.6.0/RTA version 3.4.4) with a 2 × 151 set-up using ‘NovaSeqXp’ workflow in ‘S4’ mode flow cell. The Bcl to FastQ conversion was performed using bcl2fastq_v2.20.0.422 from the CASAVA software suite. Filtering and assembly were conducted by running 10X Genomics’ Supernova version 2.1.0. The remaining genomes were assembled as follows. Single individuals were chosen per species, with preference for males when available, whose haploid status facilitates assembly. Paired-end sequencing libraries targeting 300 base pair (bp) insert sizes were prepared using the Illumina Nextera protocol. Libraries were quality checked using an Agilent bioanalyzer and sequenced as 150 bp paired-end reads on the Illumina Hi-seq platform by Edinburgh Genomics, United Kingdom. Sequencing for *Protobalandricus spectabilis* and additional sequencing for *Parnips nigripes* using Qiagen UltraLow Input libraries on an Illumina NextSeq mid-output 300-cycle run was performed at the ACRF Biomolecular Resource Facility, The John Curtin School of Medical Research, Australian National University. Raw reads were quality filtered and overlapping pairs merged in fastp (version 0.20.1) (Chen et al., 2018) with default settings, and output fastq files were visually assessed for remaining adapters and other issues with

Fastqc (version 0.11.9) (Andrews, 2010). Most genome assemblies were constructed using SPAdes (version 3.14.0) (Bankevich et al., 2012), with most species run in isolate mode with coverage cut-off estimated automatically and default k-mers. Exceptions to this were *Cecinothofagus ibarrai* Nieves-Aldrey & Liljeblad, *Callaspidia notata* (Fonscolombe), *Periclistus* spJH-2016 and *Phaenoglyphis villosa* (Hartig), which were assembled without a coverage cut-off and *Fumariphilus hypocoeci* and *Eschatocerus acaciae* Mayr, which were both assembled with an additional k-mer of 99. Data for several species were first published in Hearn et al. (2019), but were re-assembled as described here for consistency (Table S1, assembly origin column). *Synergus* species genomes and the *Biorhiza pallida* and *Ganaspis* species 1 transcriptomes were not re-assembled here. Quality statistics for all genomes and transcriptomes are given in Table S1.

Gene finding

To find conserved genes suitable for phylogenetic analysis, we predicted Hymenoptera and Eukaryota BUSCOs for each genome using BUSCO version 4.0.6 and OrthoDB version 10 (Simão et al., 2015) for each genome and transcriptome. The Hymenoptera dataset consisted of 5991 BUSCO groups predicted from 40 species (Table S1). Lineage-specific BUSCO datasets are composed of genes present almost universally as single-copy genes, although duplications within test datasets can occur (Simão et al., 2015).

Only sequences classified as complete single-copy BUSCOs were used in our analysis. A predicted BUSCO is defined as complete if its length is within two standard deviations of that BUSCO group's mean length, that is, within 95% of its expected length (Simão et al., 2015; Waterhouse et al., 2018). BUSCOs were divided into categories based on the number of species in which the gene was retrieved in a complete, single copy. In total, we found 5890 complete single-copy BUSCOs in at least one of the 37 genomes/transcriptomes (Table 3). Our phylogenetic analyses focused

TABLE 3 BUSCO gene sets and the number of species in which they were found.

No. species	No. genes (cumulative)	Total length, kb (cumulative)
37	31 (31)	61.7 (61.7)
36	92 (123)	224 (285)
35	173 (296)	350 (635)
34	246 (542)	602 (1240)
33	273 (815)	634 (1870)
32	300 (1115)	702 (2570)
31	380 (1495)	928 (3500)
30	404 (1898)	1020 (4520)
<30	3991 (5890)	12,200 (16,800)

Note: The total sequence length is given in nucleotide base pair equivalents; the number of amino acid sites is one third of this number. The numbers in parentheses represent the cumulative number of genes and the total sequence length of all data from the top of the table down to the corresponding row.

on the 523 genes that were present in 34 or more of the 37 taxa (representing a total of 1.24 Mb of nucleotide sequence data after alignment) and subsets of this dataset. We also ran one analysis using a quality-filtered subset of the 1898 genes present in 30 or more taxa. The completeness of each genome/transcriptome assembly is given in Table S1.

Alignment and quality scoring

Sequences were aligned using ClustalOmega version 1.2.4 (Sievers et al., 2011), and the alignments were filtered using Gblocks version 0.91b (Talavera & Castresana, 2007), with default parameters except for gap treatment, which was set to 'all' to retain more phylogenetic information (Kück et al., 2010). For the purpose of phylogenetic reconstruction based on multiple genes, custom scripts were used to concatenate the desired alignments.

The putative quality of alignments was scored using the fraction of the total alignment length retained after Gblocks filtering. As alternative quality filtering and scoring options, we used HmmCleaner version 0.180750 (Di Franco et al., 2019) and OD-Seq version 1.22.0 (Jehl et al., 2015), in the former case with and without previous Gblocks filtering, and in the latter after previous Gblocks filtering. HmmCleaner was used with default settings, OD-Seq with settings: distance_metric = 'affine', B = 1000, threshold = 0.025.

Phylogenetic analysis

Phylogenetic analysis was performed using IQ-Tree version 1.6.12 (Nguyen et al., 2015) and PhyloBayes version 1.8 (Lartillot & Philippe, 2004) using models that accommodate site-specific amino acid profiles. Specific settings for each program are given below.

IQ-Tree

We used IQ-Tree for maximum-likelihood analyses based on the C60 + I + G5 model. The C60 option specifies a fast approximation of an amino acid profile mixture model with 60 profile categories estimated from reference data (Wang et al., 2018). We modelled rate variation across sites using a mixture of invariable sites and a discrete approximation of a gamma distribution with five categories (i.e., I + G5). Support values were estimated using the ultrafast bootstrap (Hoang et al., 2018; Minh et al., 2013) with 2000 replicates per analysis. For each inference problem, we ran two independent analyses to confirm that phylogenetic relationships and support values were consistent. All runs used 32 CPU cores.

PhyloBayes

In PhyloBayes, we used the CAT F81 model. The CAT model infers the amino acid profile for each site from the data, assuming that the

profiles come from a Dirichlet process mixture. We assumed that the exchangeability rates were the same (F81) rather than trying to estimate them from the data (the GTR option). Estimating the exchangeability rates was too computationally complex for the analyses we attempted, and it is not obvious that the results would be more accurate, as rare changes can be explained both by unusual amino acid profiles and by low exchangeability rates under the CAT-GTR model, creating an identifiability problem that is potentially problematic. Rate variation across sites was modelled using a discrete approximation of the gamma distribution with four categories. For each inference problem, we ran two independent PhyloBayes analyses for 72 h using the MPI version on 32 CPU cores. Convergence diagnostics and consensus trees were generated for each pair of analyses using the *bpcomp* program in the PhyloBayes package, retaining every tenth sample and using a burn-in of 25% of samples. In all cases, the mean difference in split frequencies was less than 0.005, usually much less. The maximum difference in split frequencies was 0.09 for the analysis of the problematic 36-taxon dataset (see below), but was below 0.05 for all other analyses.

Individual gene tree analysis

As a complement to the analyses based on concatenated gene data, we also assessed node support using metrics summarising the information for individual gene trees. We used as the species tree the one inferred in PhyloBayes under the CAT-F81 model and based on the best third of the alignments that include at least 34 of the 37 taxa. Each individual gene tree was reconstructed using maximum likelihood with the best fit substitution model automatically selected by ModelFinder. First, using IQtree2 (Minh, Schmidt, et al., 2020), we calculated the gene concordance factor (gCF), which reflects the proportion of genes supporting a node while correcting for potential uneven taxon sampling per gene, and the site concordance factor (sCF), which estimates the concordance at the level of individual sites (Minh, Hahn, & Lanfear, 2020; Mo et al., 2022). Second, using RAxML version 8.2.12, we calculated internode certainty (IC), which informs about the certainty of a bipartition by considering its occurrence in a set of gene trees relative to the occurrence of the second-best bipartition (Salichos & Rokas, 2013).

UCE analysis

To directly test our results against the previous UCE analysis by Blaimer et al. (2020) and extend this analysis, we compiled a dataset combining data from both studies. We first blasted the UCES in the previous study to our Gblocks-filtered data for genes present in at least 34 of the 37 taxa in our analysis. We saved all matches with an e -value $<10^{-30}$, and then we removed UCES with multiple hits in the same genomes and UCES that were too close (less than 50 bp on either side) to the edge of a contig/scaffold. For the phylogenetic analysis, we followed Blaimer et al. (2020). Specifically, we adopted the SWSC-EN partitioning scheme, then used PartitionFinder2 (Lanfear et al., 2017) to find the optimal partitioning scheme, and

finally ran an ML search in IQ-Tree under default settings for partitioned data, that is, using ModelFinder (Kalyaanamoorthy et al., 2017) to find the best nucleotide model for each partition, and then running an ML search for the best tree. We used 1000 ultrafast bootstrap replications in IQ-Tree to assess branch support in the resulting tree.

Tree figures

Illustrations of phylogenetic trees were generated using the R package *ggtree* version 3.2.1 (Yu et al., 2017) running under R version 4.1.1.

RESULTS

Alignment quality and phylogenetic signal

We first explored the data by analysing the dataset that contained the 31 genes that were present in all 37 taxa (Table 2). We will refer to this as the T37-G31 dataset, for 37 taxa and 31 genes. We then successively expanded the amount of data by analysing the T36-G123 dataset (123 genes present in 36 or more taxa), T35-G296 dataset (296 genes present in 35 or more taxa) and T34-G542 datasets (542 genes present in 34 or more taxa). This series represents a trade-off between completeness in terms of taxa and amount of genomic data included.

When these four datasets were analysed with IQ-Tree and a model accommodating site-specific amino acid profiles (C60 + I + G5), we discovered striking differences in topology (Figure 2). In the smallest dataset (T37-G31; Figure 2a), including only the complete alignments, *Eschatocerus* (*Eschatocerini*) diverges early in the tree. However, in the next smallest dataset (T36-G123; Figure 2b), *Eschatocerus* is instead grouped inside the core cynipid lineages, in a clade together with *Protobalandricus* (*Cynipini*), *Phanacis* Förster (*Phanacidini*) and *Iraella* (*Aylacini* s. str.). This is a somewhat surprising result, as it breaks the monophyly of the oak gall wasps (*Cynipini*), long presumed to be a monophyletic group. It also moves *Phanacis* (*Phanacidini*)—representing one major herb-galling clade—from a sister-group relationship with the other major herb-galling clade (*Aulacideini*) to a position within a heterogeneous collection of lineages. As more genes (and more gaps) are added, *Eschatocerus* changes again to an early-diverging position (Figure 2c,d), but *Protobalandricus* remains outside the *Cynipini*, even though the support for this is quite poor in the largest dataset (Figure 2d).

In trying to understand these results, we noted that *Eschatocerus* is a long-branch taxon, and that the three other taxa that group with *Eschatocerus* in the next smallest dataset (Figure 2b) have three of the five most incomplete genome assemblies in terms of the number of retrieved genes (Table S2). This suggests that the clade consisting of *Eschatocerus*, *Protobalandricus*, *Phanacis* and *Iraella* may be spurious and caused by long-branch attraction and/or poor or misleading gene alignments.

We looked at long-branch attraction first. The C60 model in IQ-Tree is an approximation of the CAT model in PhyloBayes and may

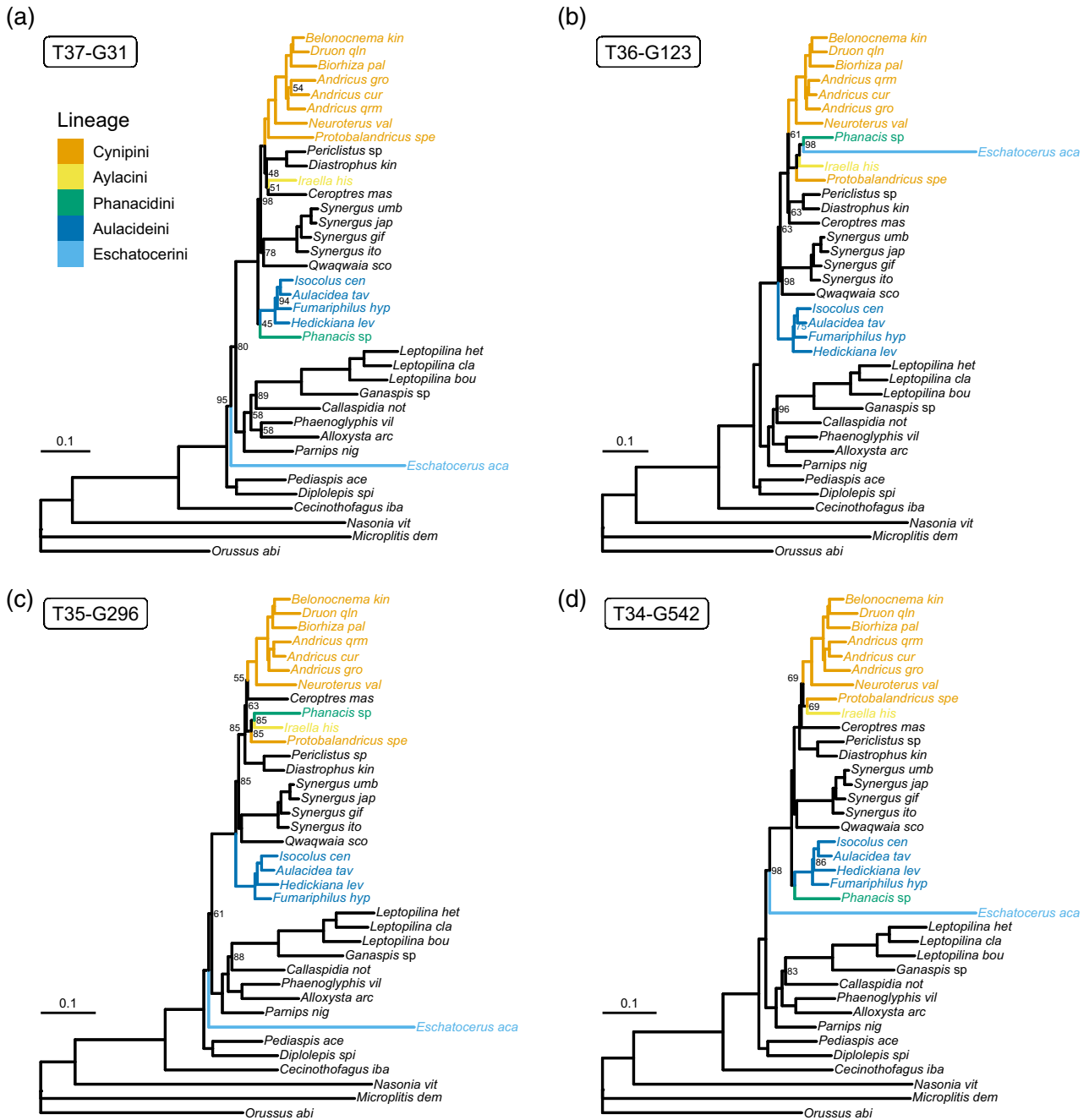


FIGURE 2 Phylogenetic relationships according to IQ-Tree analyses of different data subsets under the C60 + I + G5 substitution model. Support values (ultrafast bootstrap method in %) are shown on branches only if they are less than 100%. (a) Analysis of the 31 genes present in all 37 taxa (dataset T37-G31). (b) Analysis of the 123 genes present in 36 or more taxa (T36-G123). (c) Analysis of the 296 genes present in 35 or more taxa (T35-G296). (d) Analysis of the 542 genes present in 34 or more taxa (T34-G542). Note that the smallest dataset (a) results in strong support for Cynipini monophyly (blue clade). In the second smallest dataset (b), this is not the case because *Protobalandricus*, the most basal lineage in Cynipini, groups strongly with *Iraella* (Aylacini), *Phanacis* (Phanacidiini) and *Eschatocerus* (Eschatocerini), the latter of which sits on a long branch. When even more data are added, the support for this assemblage successively weakens (c, d), until there is only insignificant evidence (69% bootstrap support) against Cynipini monophyly in the largest dataset (d).

not accurately represent site-specific amino acid profiles in cynipoids. Such deviations could potentially cause problems with long-branch attraction in our analysis. To check this possibility, we repeated the analysis of the two smallest datasets (the others were too large) in PhyloBayes using the CAT-F81 model (Figure 3). The

results were identical with those obtained with IQ-Tree, suggesting that the topological changes are not caused by problems with the C60 approximation.

Next, we turned our attention to alignment quality. We noted that even the relatively unrestrictive Gblocks filtering we used

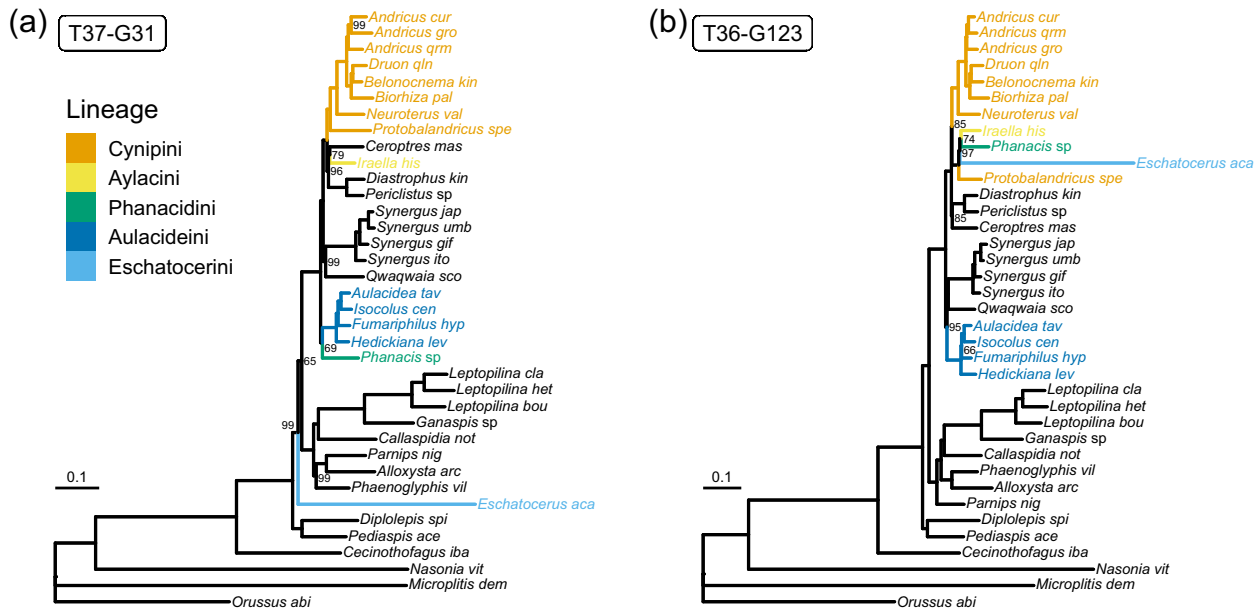


FIGURE 3 Phylogenetic relationships according to PhyloBayes analyses of the two smallest data subsets under the CAT-F81 model. Support values (posterior probability in %) are shown on branches only if they are less than 1.0. (a) Analysis of the 31 genes present in all 37 taxa (dataset T37-G31). (b) Analysis of the 123 genes present in 36 or more taxa (T36-G123). Despite the more sophisticated CAT-F81 model, which learns the amino acid profiles from the data, the results are virtually identical to the corresponding results of IQ-Tree (Figure 2a,b).

sometimes removed substantial portions of the alignments. If a substantial portion of an alignment is unreliable, then the part that remains after filtering could also be potentially of doubtful quality. To examine this possibility, we divided the T34-G542 dataset into six approximately equal gene subsets based on the proportion of the alignments removed by Gblocks. When analysed with IQ-Tree under the C60 + I + G5 model, the three best data subsets (that is, the three subsets with the smallest proportion of site columns removed by Gblocks) resulted in trees (Figure 4a–c) that were identical to each other and to the tree from the no-gaps dataset T37-G31 (Figure 2a), except for a few minor details, most of which were not well supported. Notably, *Eschatocerus* always diverged early, Cynipini was monophyletic and *Phanacis* grouped with Aulacideini in all these trees.

The results for the three worst data subsets (Figure 4d–f) differed among themselves and from the results of the no-gaps dataset (T37-G31) in several respects, often involving contrasting placements of the four problematic taxa mentioned previously—*Eschatocerus*, *Protobalandricus*, *Phanacis* and *Iraella*—or unusual arrangements of more basal branching events, but with low support. Thus, it appears that the phylogenetic signal is consistent in the best gene alignments, that is, those that contain only small portions that are detected as problematic by the Gblocks filter.

To further test the effect of alignment quality, we also explored partitions of the T34-G542 dataset generated using other filtering and scoring methods. Specifically, we tried HmmCleaner, Gblocks + HmmCleaner and Gblocks + OD-Seq, and then divided the gene alignments into subsets based on how many sites were removed (HmmCleaner and Gblocks + HmmCleaner) or how many sequences (Gblock + OD-Seq) were removed by each of these pipelines. These

tools represent different approaches to data cleaning: Gblocks removes site columns, OD-Seq removes outlier sequences and HmmCleaner removes sequence fragments. In all cases, the IQ-Tree analyses of the highest quality data subset or subsets resulted in trees that were identical or almost identical to the tree from the no-gaps analysis (Figures S1–S3). The quality scores for OD-Seq are few, and it was difficult to devise criteria that generated partitions of equal size. We therefore ended up with the best partition (no sequences removed) being much smaller than the other ones (approximately 2900 sites vs. 25,600–56,800 sites), and resulting in a poorly resolved tree with some unusual features (Figure S3A). Analysis of the next best OD-Seq partition, however, retrieved a tree that was highly similar to the no-gaps tree (compare Figure 2a to Figure S3B).

Based on these results, we conclude that it is mainly poor-quality alignments that generate the somewhat unexpected placements of *Eschatocerus*, *Protobalandricus*, *Phanacis* and *Iraella* in analyses of the T36-G123, T35-G296 and T34-G542 datasets.

Long-branch attraction

The tree on which all analyses of high-quality alignments converge (e.g., Figure 4a–c) supports many previous notions of cynipid relationships. For instance, the cynipid tribes Cynipini, Diastrophini, Synergini (s. str.) and Aulacideini are all monophyletic, as is the family Figitidae, the figitid subfamilies Eucoilinae and Charipinae, and the Cynipoidea as a whole. However, they also support one of the novel conclusions from the UCE analysis (Blaimer et al., 2020), namely that gall wasps themselves (Cynipidae) are not monophyletic. The

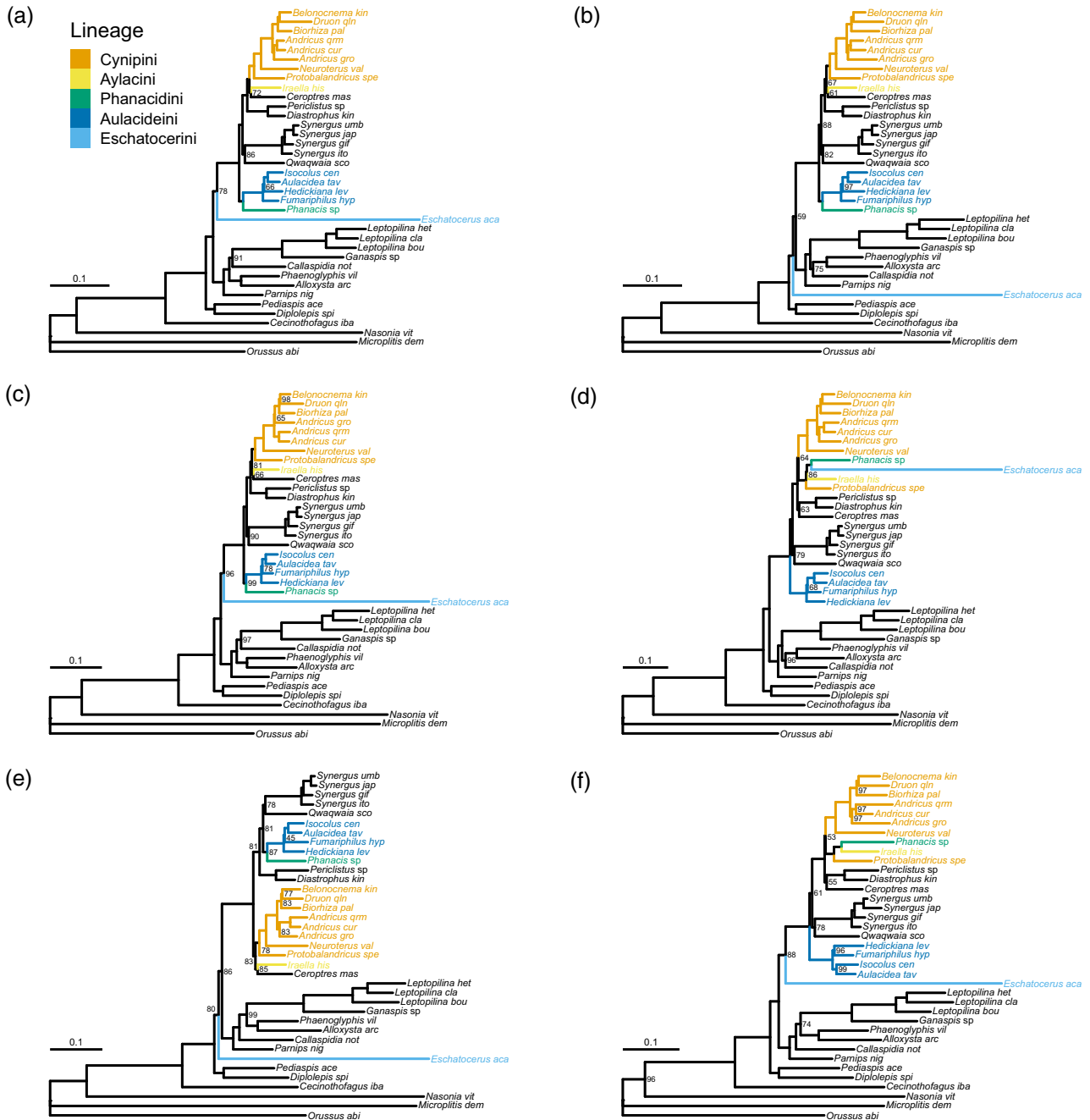


FIGURE 4 Phylogenetic results for six equally-sized, quality-ranked subsets of the T34-G542 dataset, analysed using IQ-Tree under the C60 + I + G5 model. The raw alignments were subjected to filtering and quality ranking by Gblocks. Support values (ultrafast bootstrap in %) are shown on branches only if they are less than 100%. (a) Less than 13% of sites filtered out (best quality). (b) From 13% to 26% filtered out. (c) From 26% to 37% filtered out. (d) From 37% to 47% filtered out. (e) From 47% to 59% filtered out. (f) More than 59% filtered out (worst quality). The three best subsets (a-c) yield congruent results except for the position of *Eschatocerus*, which varies slightly but without strong conflict in support values. All have monophyletic Cynipini (blue lineages), and none of them group Phanacidini, Aylacini and Eschatocerini with each other or with *Protobalandricus*, as seen in some of the poor-quality data subsets (d, f).

Cynipidae tribes Diplolepidini + Pediaspini (represented by *Diplolepis* and *Pediaspis*) and Paraulacini (represented by *Cecinothofagus* Nieves-Aldrey & Liljeblad) lineages both fall outside a core cynipid clade that apparently constitutes the sister group of the Figitidae. In some analyses, the Eschatocerini (represented by *Eschatocerus*) also fall outside this clade.

The putative cynipid lineages that place outside the core cynipid clade all represent long branches in the tree, as do the outgroup taxa. Could the non-monophyly of Cynipidae be the result of long-branch attraction, pulling isolated cynipid lineages towards the outgroups? To examine this question, we focused on a dataset consisting of the two best subsets of the T34-G542 dataset according to the Gblocks

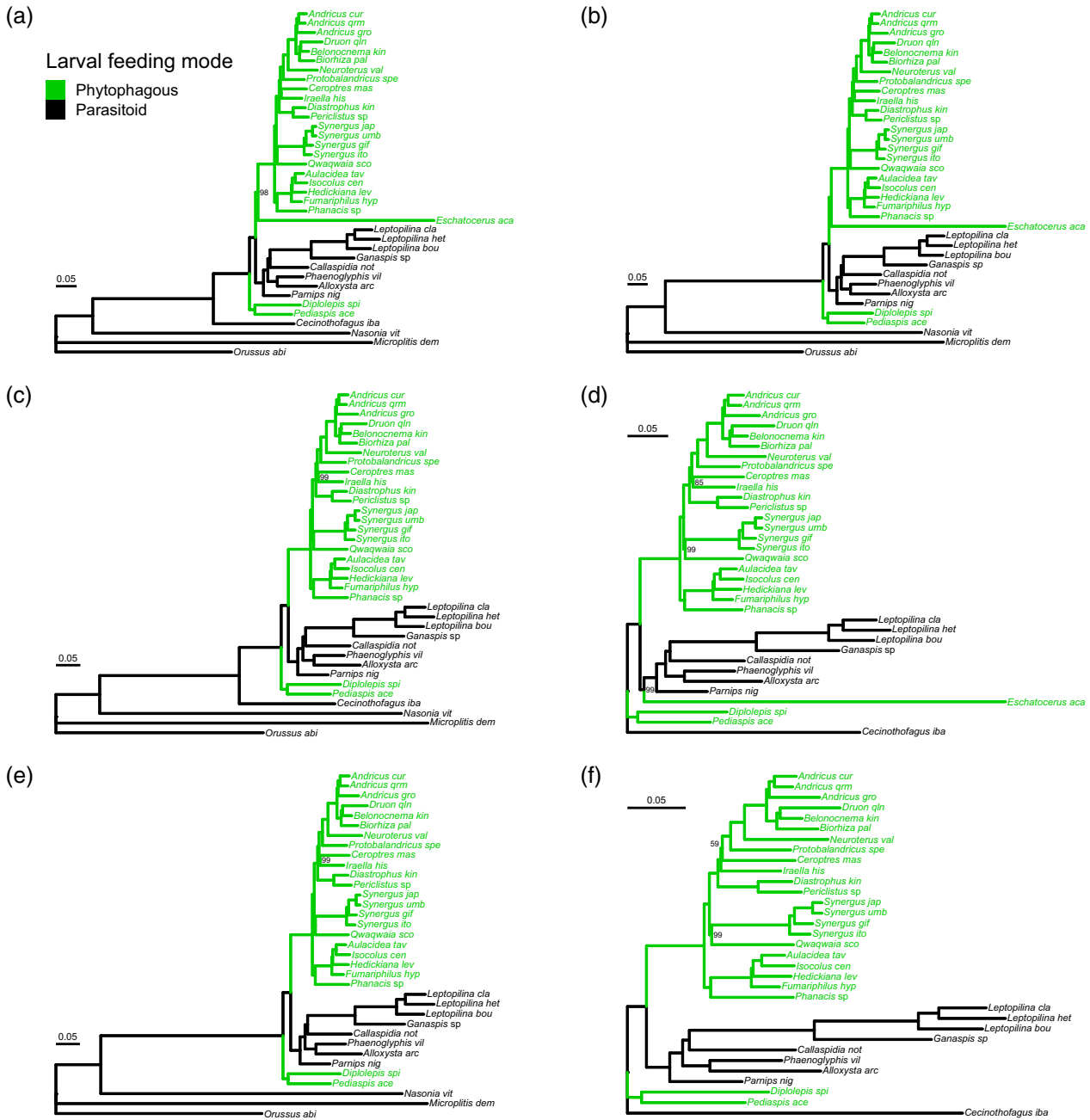


FIGURE 5 Testing the potential effect of long-branch taxa on phylogenetic results. For these analyses, we used the best third of the T34-G542 alignments, that is, the alignments where Gblocks filtered out 26% or less of the sites (see Figure 3). For the best possibility of detecting model-related long-branch attraction effects, we used PhyloBayes and the CAT-F81 model. Branch support values (posterior probability in %) are only shown if they are less than 1.0. (a) Analysis of the full taxon set. (b) *Cecinothofagus* excluded. (c) *Eschatocerus* excluded. (d) Outgroups excluded. (e) *Cecinothofagus* and *Eschatocerus* excluded. (f) *Eschatocerus* and outgroups excluded. Regardless of taxon exclusion, the relationships among the included lineages remain identical to those in the full analyses, except for a slight variation in the position of *Eschatocerus* when outgroups are excluded (d). Notably, the phytophagous groups (gall inducers and inquilines, green) remain diphyletic in all analyses with respect to the parasitoid lineages (black).

criterion, and we used PhyloBayes as the best approach for detecting long-branch attraction. The analysis of the complete taxon set resulted in the tree with non-monophyletic Cynipidae (Figure 5a). From this dataset, we then removed in turn *Cecinothofagus*, *Eschatocerus*, outgroups, *Cecinothofagus* + *Eschatocerus*, and *Eschatocerus* + outgroups. These were the only removals of long-branch taxa that

left a sufficient number of remaining lineages to test non-monophyly of Cynipidae. In all cases, the support for non-monophyletic Cynipidae remained at 100% (Figure 5b–f). The results were almost identical when the same datasets were analysed with IQ-Tree (Figure S4). In conclusion, we find no evidence that the non-monophyly of Cynipidae is caused by long-branch taxa causing problems with the analysis.

Gene tree analysis

The gene tree concordance analysis shows that there is consistent signal across gene trees for the deep splits in the superfamily, that is, between *Cecinothofagus* and the remaining taxa, and between *Diplolepis* + *Pediaspis* on one hand and the remaining Cynipidae and Figitidae on the other (Figure S5). This is reflected both by a positive IC and a gCF > 40%. However, the relationships among *Eschatocerus*, remaining Cynipidae (s. str.) and Figitidae are not consistently resolved across gene trees. Similarly, this analysis indicates a fair amount of inconsistency across gene trees concerning tribal relationships within the Cynipidae (s. str.) excluding *Eschatocerus*. This could be because errors in the assemblies, errors in gene tree inference due to lack of data or biases in the simplified model used, or reflect genuine topological variation among the gene trees. However, we did not pursue this further.

Trade-off between taxon completeness and data quality filtering

To further explore the trade-off between taxon completeness and alignment quality filtering, we ran an IQ-Tree analysis under the C60 model also for the best gene alignments (those where Gblocks removed at most 10% of alignment columns) that had data for 30 or more of the taxa. The size of the filtered dataset (811,373 amino acid sites for 37 taxa) was close to the limit of what was computationally feasible with IQ-Tree under the C60 model on the cluster we used. This analysis is more relaxed with respect to taxon completeness and stricter with respect to alignment quality compared with the high-quality T34-G542 analysis (Figure 5a). The resulting tree (Figure S6) differed only in minor details from the T34-G542 tree but was less resolved, indicating that this analysis provided a less favourable balance between signal and noise.

Extended UCE analysis

The combined UCE analysis of our data and the data of Blaimer et al. (2020) resulted in a tree (Figure S7) that is entirely consistent with the high-quality T34-G542 tree (Figure 5a). Notably, the clade consisting of *Iraella*, Cynipini and Ceroptresini remained strongly supported, confirming our conclusion that the Aylacini (s. str.) are not closely related to the Aulacideini + Phanacidini. The *Eschatocerini* were placed as sister to the remaining Cynipidae (s. str.) with strong support, in accordance with our high-quality T34-G542 tree and with some analyses of Blaimer et al. (2020). Interestingly, this analysis placed the Qwaqwaiini as sister group to the Rhoophilini, with strong support. Importantly, the extended UCE analysis suggests that the phylogenomic results presented here and in Blaimer et al. (2020) are robust to the choice of different types of data, analytical approaches and sampling of ingroup and outgroup taxa.

Phylogenetic relationships

As our best phylogenomic estimate of relationships, we present the PhyloBayes (CAT-F81) analysis of the two best subsets of the T34-G542 dataset according to the Gblocks criterion (Figure 6; see also Figures 5a and S4A), as it appears to represent the best trade-off between signal and noise as judged by clade credibility values, and uses the most sophisticated of the analytical approaches explored here. The tree is also largely congruent with (and never conflicts strongly with) the results from any of the other analyses of quality-filtered data. It is also entirely consistent with the results of the extended UCE analysis (Figure S7). On the tree, we have indicated the currently recognised cynipid tribes and a proposed reclassification of the family Cynipidae into three family-level taxa: the Cynipidae (s. str.) for the core cynipid clade, including *Eschatocerini*; the Diplolepididae stat. prom. for *Diplolepidini* + *Pediaspidini*; and the Paraulaciidae stat. prom. for the Paraulacini.

Our results suggest that the two major tribes of herb gallers, Phanacidini and Aulacideini, form a natural group at the base of the Cynipidae (s. str.). The third tribe of herb gallers (Aylacini s. str.), represented in our analysis by *Iraella*, is apparently more closely related to the oak gallers (Cynipini) and the oak inquiline in the tribe Ceroptresini (represented by *Ceroptres*) than to the other herb gallers. The Aylacini (s. str.) are all associated with plants in the family Papaveraceae. The Phanacidini and Aulacideini are most commonly associated with Asteraceae and Lamiaceae, but there is one species in the Aulacideini, *Fumariphilus hypecoi*, associated with Papaveraceae. Our results confirm that this species is not a member of Aylacini (s. str.), consistent with previous analyses (Ronquist et al., 2015; see also Nieves-Aldrey, 2022).

The Diastrophini, represented in our analysis by *Periclistus* and *Diastrophus*, form the sister group of the clade including Cynipini + Ceroptresini + Aylacini (s. str.). It is a tribe that includes both inquilines and gall inducers associated with host plants in the family Rosaceae, mostly bushes of the genera *Rubus* and *Rosa* but also herbs in the genus *Potentilla* L.

The tribe Qwaqwaiini, represented in our analysis by the only described species, *Qwaqwaia scolopiae* Liljeblad, Nieves-Aldrey & Melika, appears to form a clade together with Synergini (s. str.) and Rhoophilini; specifically, the extended UCE analysis (Figure S7) suggests that the Qwaqwaiini forms the sister group of the Rhoophilini. Members of the Qwaqwaiini + Rhoophilini + Synergini (s. str.) clade are all inquilines in Cynipini galls and a few other insect galls with two notable exceptions: *Qwaqwaia* is assumed to be a gall inducer, and one of the species we analysed, *Synergus itoensis*, represents a small subgroup within the Synergini (s. str.) of true gall inducers on oaks. The latter subgroup appears to be the sister group of the rest of the Synergini (s. str.) in our analysis only because several early-diverging representatives are missing (Ide et al., 2018; Lobato-Vila et al., 2022; Ronquist et al., 2015). The Qwaqwaiini + Rhoophilini + Synergini (s. str.) seem to represent the sister group of Diastrophini + Cynipini + Ceroptresini + Aylacini (s. str.).

The *Eschatocerini*, represented in our analysis by the single genus *Eschatocerus*, was placed as the sister group of all other Cynipidae

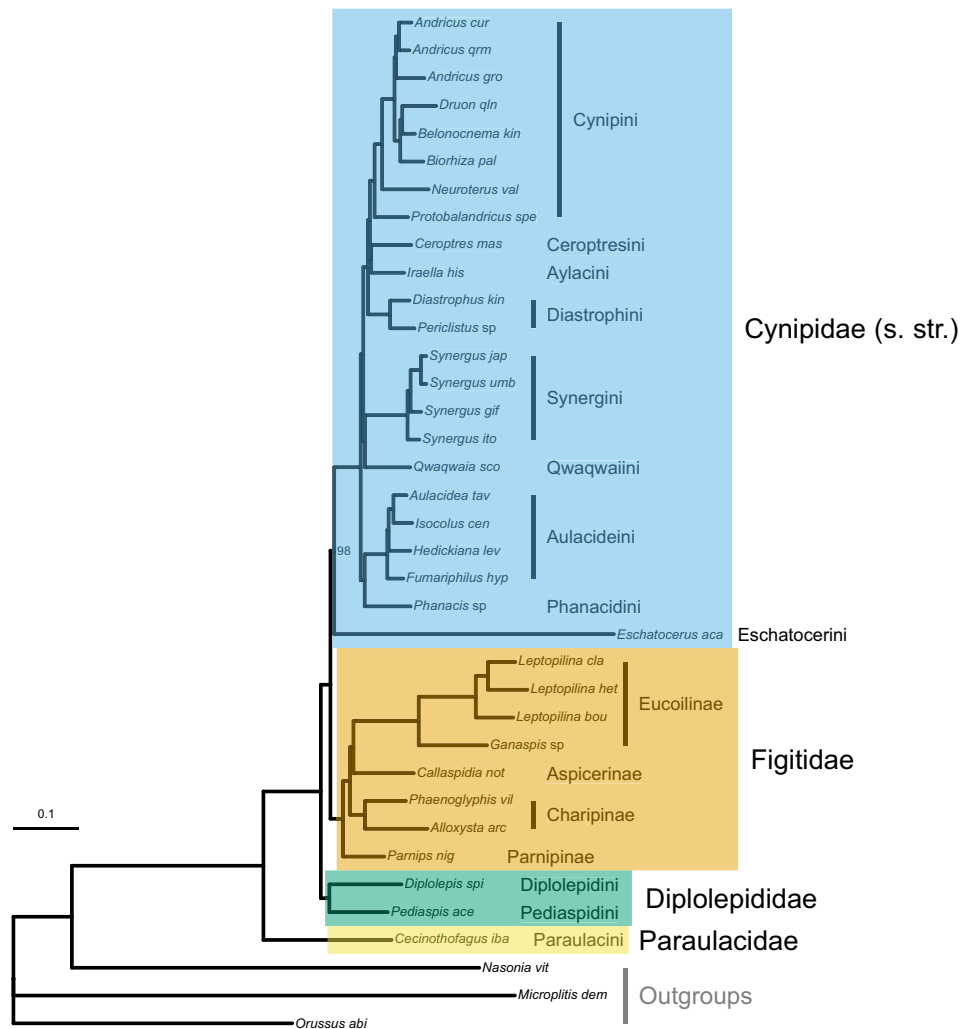


FIGURE 6 Main conclusions on phylogenetic relationships. The tree is based on the best third of the alignments that include at least 34 of the 37 taxa (T34-G542 dataset, Gblocks filtering removed less than 26% of sites), analysed using PhyloBayes under the CAT-F81 model (the same analysis shown in Figure 5a). Current cynipid tribes and figitid subfamilies are indicated, together with the proposed new classification of cynipid lineages into three distinct families.

(s. str.) in our analysis, a conclusion also supported by the extended UCE analysis (Figure S7). Occasionally, we retrieved *Eschatocerus* as the sister group of remaining Cynipidae (s. str.) + Figitidae (Figures 4b, S2A,B and S3B) or only Figitidae (Figure S1B), although with unconvincing support values. Thus, we conclude that the tribe Eschatocerini likely belongs to the Cynipidae (s. str.).

The Figitidae in our analyses form a strongly supported monophyletic group. The subfamily Parnipinae, represented in our analysis by the single genus *Parnips*, appears as the sister group of the remaining lineages. It is a parasitoid of cynipid gall inducers in the tribe Aylacini (s. str.). The Charipinae, represented by *Phaenoglyphis* Förster and *Alloxysta* Förster in our analysis, form a monophyletic group. They are hyperparasitoids of other parasitic wasps attacking aphids. The remaining Figitidae apparently form a monophyletic group, falling into two subgroups: the Aspicerinae (*Callaspidia* Dahlbom) and the Eucoilinae (the remaining species). Both subfamilies are parasitoids of Diptera larvae.

Among the more early-diverging cynipoid lineages, the Diplolepidini, represented by *Diplolepis*, and the Pediaspidini, represented by *Pediaspis*, form a strongly supported clade, which appears to be the sister group of Figitidae + Cynipidae (s. str.). We propose here that this clade be recognised as a separate family, the Diplolepididae stat. prom. (Figure 6). Previous studies suggest that the Diplolepidini and Pediaspidini are reciprocally monophyletic and that each lineage contains two distinct genera (Liljeblad et al., 2008; Zhang et al., 2020). Given this, we suggest that these tribes are recognised as separate subfamilies, the Diplolepidinae stat. prom. and the Pediaspidinae stat. prom., within the Diplolepididae stat. prom. (Figure 6).

Finally, our results support the conclusion of Blaimer et al. (2020) that the Paraulacini (represented by *Cecinothofagus*) form the sister group of the remaining cynipoid lineages. We propose here that also this clade be recognised as a separate family, the Paraulacidae stat. prom. (Figure 6). The new classification is discussed in more detail in the Taxonomy section below.

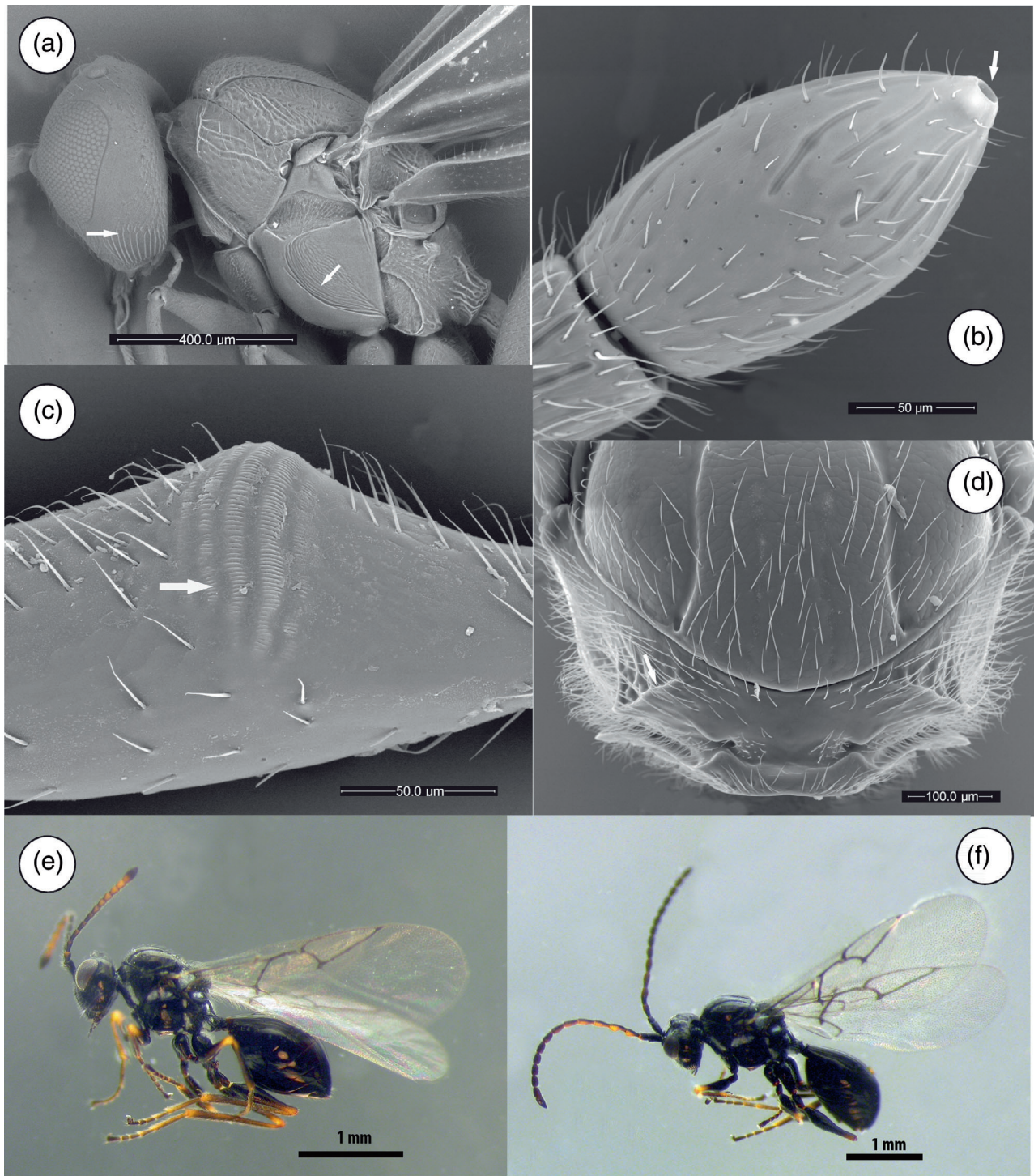


FIGURE 7 Morphological characters and habitus of exemplar species of Paraulacidae stat. prom. (a) *Paraulax queulensis* Nieves-Aldrey & Liljeblad. Head and mesosoma lateral view. (b–d) *Cecinothofagus gallaecoihue* Nieves-Aldrey & Liljeblad. (b) Antennal clava, female. (c) Profemur. (d) Pronotum. (e, f) Habitus of *Cecinothofagus ibarrae* Nieves-Aldrey & Liljeblad. (e) Female. (f) Male.

TAXONOMY

Taxonomy

Given that our results provide solid and independent confirmation of the results from the UCE analysis (Blaimer et al., 2020) regarding the

non-monophyly of Cynipidae, we find it appropriate to revise the family-level classification to reflect these findings here. As the circumscription of the 13 cynipid tribes and potential apomorphies characterising each of them have been discussed at length previously (Lobato-Vila et al., 2022; Ronquist et al., 2015), we just give a brief formal synopsis of the proposed family classification here. The

synopsis does not include the fossil family-level taxa, the phylogenetic position of which must be carefully re-evaluated in light of the phylogenomic findings. We refrain from revising the classification of the non-cynipid family-level taxa in the Cynipoidea, as the results of the UCE analysis on Liopteridae, Ibalidae and some Figitidae lineages still await independent confirmation.

Paraulacidae Nieves-Aldrey and Liljeblad, stat. prom. [urn:lsid:zoobank.org:act:00B72CB3-E455-468E-AAC0-3A7637B4BC8A]

Type genus *Paraulax* Kieffer, 1904: 568

Paraulacini Nieves-Aldrey and Liljeblad, 2009

Habitus female and male (Figure 7e,f)

Diagnosis: Gena with 5–9 vertical carinae in the ventral region (Figure 7a). Genal part of occipital carina present. Ventral part of clypeus not or only slightly projecting over mandibles. Female antenna with 10 flagellomeres, F10 clavate with presence of a large volcano sensilla at flagellum apex (Figure 7b). Modified flagellomere of male antenna always F2, F3 or both. Dorsolateral margin of pronotal plate projecting laterally (Figure 7d). Lateral pronotal carina present. Scutellar foveae always shallow or indistinct; round. Mesopleural impression present (Figure 7a). Profemur with the basal third swollen and carrying a structure of 4–5 rows of sharp, closely spaced and deep costulae (Figure 1c).

Comments: A set of unique morphological features allow easy differentiation of Paraulacidae from Cynipidae and other families in Cynipoidea (Ronquist et al., 2015). Three unique autapomorphies can be emphasised: the presence of 5–9 vertical carinae in the ventral region of the gena; the profemur with the basal third swollen and carrying a structure of 4–5 rows of sharp, closely spaced and deep costulae; and the presence of volcano sensilla on the apical flagellomere (Polidori & Nieves-Aldrey, 2014). The Paraulacidae appear to be parasitoids of gall-inducing chalcidoids of the genus *Aditrochus* Rübsaamen on species of *Nothofagus* (Nothofagaceae) (Rasplus et al., 2022).

Circumscription: The family includes the genera *Paraulax* and *Cecinothofagus*, each with three species. Southern South America, found only in the temperate *Nothofagus* forests of Chile and Argentina.

Diplolepididae Latreille, 1802, stat. prom. [urn:lsid:zoobank.org:act:8D49E26F-4FBB-46DA-B240-D362E7A1DB9A]. Type genus *Diplolepis* Geoffroy, 1762. Conserved (see Kerzhner, 1991)

Dipolepariae Latreille, 1802. Corrected to Diplolepididae

Rhoditini Hartig, 1840

Diplolepidini Latreille (Ronquist, 1999)

Diagnosis: Distinctive morphological and biological features that separate this family from the Cynipidae (see key to tribes in Ronquist et al., 2015) are: female antenna with 12 or more flagellomeres and male antenna without modified F1 (Figure 8e); mesopleuron with a mesopleural longitudinal impression (Figure 8b); scutellar foveae faint or absent (Figure 8a); metatarsal claws simple; and hypopygium either ploughshare-shaped (Figure 8c) or hypopygial spine short. Gall inducers on *Rosa* or *Acer*.

Comments: Potential apomorphies of the Diplolepididae include the mesopleuron with a mesopleural longitudinal impression, faint or absent scutellar foveae and the female and male antenna having 12 or

more flagellomeres (Ronquist et al., 2015: couplet 3 in the key to cynipid tribes). A quantitative analysis of the available morphological and biological evidence for Diplolepididae monophyly is still missing. Before such an analysis is attempted, however, it would be valuable to reassess the morphological evidence in the light of the phylogenomic results. It is clear that such an analysis would have to span all major cynipoid lineages, and not be restricted to the former cynipid groups.

Circumscription: The family includes two subfamilies, Diplolepidinae and Pediaspidinae, corresponding to the tribes Diplolepidini and Pediaspidini in Ronquist (1999), here elevated to subfamily status.

Diplolepidinae Latreille, 1802, stat. prom. [urn:lsid:zoobank.org:act:DB27C5E2-8A03-47 AD-A381-4307D0E60EDD]. Type genus *Diplolepis* Geoffroy, 1762

Diplolepidini Latreille, 1802

Habitus female and male (Figure 8f,g)

Diagnosis: Pronotum short dorsomedially. Pronotal plate not marked (Figure 8d). Scutellar foveae faint or absent. Mesopleuron with a broad, crenulate mesopleural impression (Figure 8b). Lateral propodeal carinae indistinct (Figure 8d). Metanotal trough broad, apically truncate. 2r of fore wing with a prominent median vein stump projecting anterolaterally. Nucha dorsally short. Hypopygium ploughshare-shaped (Figure 2c). Ovipositor articulation present as a weak flexion point or a distinct articulation.

Comments: Putative morphological apomorphies for the Diplolepidini include the ploughshare-shaped hypopygium, the broad and crenulate mesopleural impression, and the lack of lateral propodeal carinae (Ronquist et al., 2015). However, quantitative analyses have struggled to identify unique or distinct apomorphies for the subfamily, partly because of variation among the constituent taxa, and partly because of the previous difficulties in resolving relationships among cynipid tribes (Ronquist et al., 2015). There is some uncertainty surrounding the interpretation of the hypopygial character. Vyrzhikovskaja (1963) claims that the hypopygium of *Liebelia* is straight and not ploughshare-shaped as in *Diplolepis*, a claim that is repeated elsewhere without reference to the original source (Melika, 2006; Pujade-Villar et al., 2020). Nonetheless, the hypopygium of *L. magna* Vyrzhikovskaja clearly appears to be ploughshare-shaped in the SEM illustration provided by Liljeblad et al. (2008), albeit less extremely so than the hypopygium of *Diplolepis*. The hypopygium of *L. dzhungarica* also appears to have some distinct similarities in shape to that of *Diplolepis* (Abe, Melika & Stone 2007). Given this uncertainty, the character would clearly warrant more detailed study in these and related taxa.

Circumscription: Two genera, *Diplolepis* Geoffroy with 51 species and *Liebelia* Kieffer with 10 species. Holarctic. Gall inducers on *Rosa* spp. (Rosaceae).

Pediaspidinae Ashmead, 1903, stat. prom. [urn:lsid:zoobank.org:act:EC6E5132-F1FB-4D37-B1CD-0042226B515A]. Type genus *Pediaspis* Tischbein, 1852

Pediaspidini Ashmead, 1903. *Psyche* (Cambridge Mass.), 10: 147.

Himalocynipinae Yoshimoto, 1970. *Can. Entomol.*, 102: 1583.

Habitus female and male (Figure 9d–f).

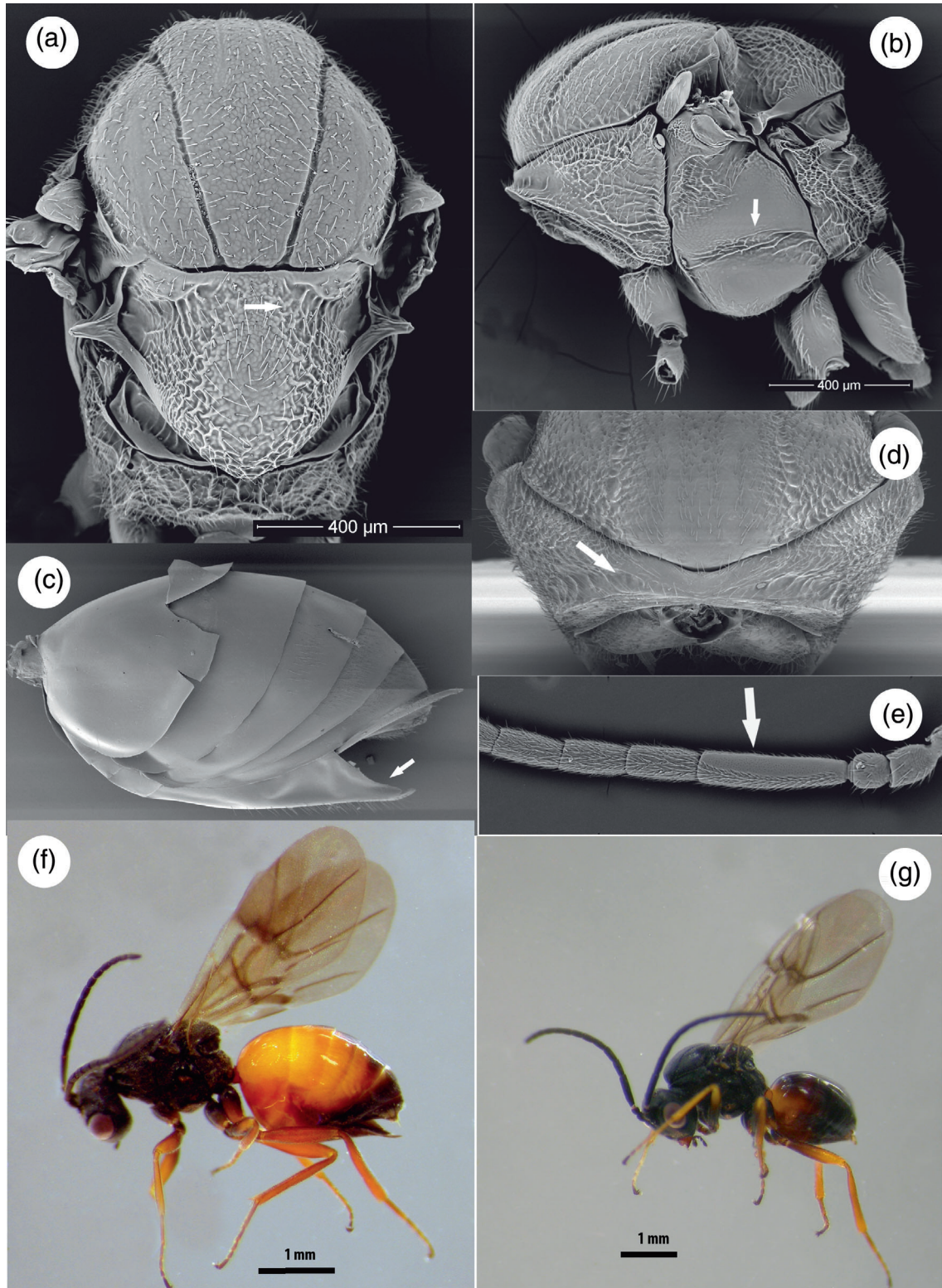


FIGURE 8 Morphological characters and habitus of exemplar species of Dipolepididae stat. prom., Dipolepidinae stat. prom. (a, b) *Dipolepis mayri* (Schlechtendal). (a) Mesosoma in dorsal view. (b) Mesosoma in lateral view. (c, d) *Dipolepis triforma* Shorthouse & Ritchie. (c) Metasoma in anterior view. (d) Pronotum in frontal view. (e) Male antenna of *D. mayri*. (f, g) Habitus of *Dipolepis mayri*. (f) Female. (g) Male.

Diagnosis: Facial strigae radiating from clypeus distinct but not reaching past 0.6 distance to compound eye. Sculpture on vertex dorsad compound eye more or less erased. Ventral area of gena with

smooth sculpture, without vertical carinae. Ventral part of clypeus broadly projecting over mandibles. Female antenna with 12 or more flagellomeres; last flagellomere not wider than the penultimate. Male

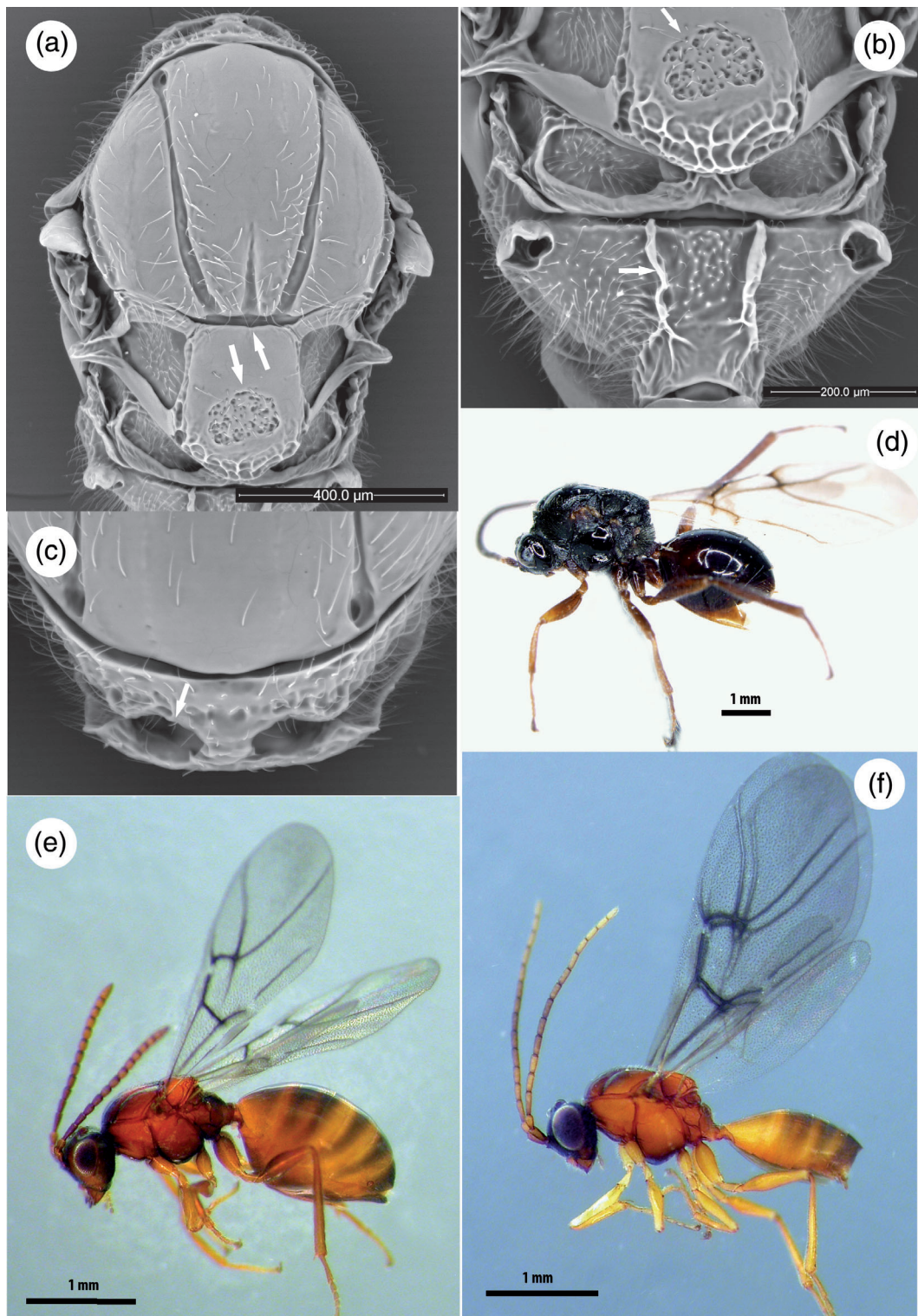


FIGURE 9 Morphological characters and habitus of exemplar species of Diplolepididae stat. prom., Peditaspidinae stat. prom. (a–c) *Peditaspis aceris* (Gmelin) sexual generation. (a) Mesosoma in dorsal view. (b) Mesoscutellum and propodeum. (c) Pronotum in frontal view. (d) Habitus of female of *Himalocynips vigintilis* Yoshimoto. (e, f) Habitus of *Peditaspis aceris* (Gmelin) sexual generation. (e) Female. (f) Male.

antenna without modified F1. Dorsolateral margin of pronotal plate not projecting laterad; admedian depressions of pronotum deep and widely separated (Figure 9c); area posterior to transscutal fissure flat or convex (Figure 9a). Scutellar foveae absent; a round, distinctly

margined posteromedian scutellar impression present (Figure 9a,b). Mesopleural impression linear and not sculptured. Lateral carinae of propodeum distinct (Figure 9b). Profemur not modified. Mesocoxa with a hump present laterobasally.

Comments: The Pediaspidini are characterised by several unique or distinct apomorphies, among them the posteromedian scutellar impression (Ronquist et al., 2015). *Himalocynips*, a genus with a single species that was described within its own subfamily (Himalocynipinae Yoshimoto, 1970) was included within the Pediaspidini by Ronquist (1999). Its phylogenetic proximity to *Pediaspis* was later supported by a morphological phylogenetic analysis (Liljeblad et al., 2008). The biology and host plant of this species are however unknown, although it may (as for *Pediaspis*) be a galler on *Acer* (Sapindaceae). We were unable to include this rare and poorly studied species in our analysis, and a molecular confirmation of its placement within the Pediaspidini and the Diplolepididae is an obvious priority for future studies.

Circumscription: Two genera, *Himalocynips* and *Pediaspis*, with one species each. Palaearctic.

One genus gall inducer on *Acer* spp. (Sapindaceae); the other with unknown biology.

Cynipidae (s. str.)

Cynipsera Latreille, 1802. Corrected to Cynipidae. Type genus: *Cynips Linnaeus*, 1758.

Circumscription: As here proposed, the family includes the tribes Eschatocerini, Phanacidini, Aulacideini, Qwaqwaiini, Synergini, Rhooophilini, Diastrophini, Aylacini, Ceroptresini and Cynipini.

Comments: The position of the Eschatocerini is still highly uncertain, and its life history is also poorly studied. Future studies will have to show whether it truly belongs to the Cynipidae (s. str.), or elsewhere in the Cynipoidea, probably then as a separate family. The potential apomorphies of each of the remaining tribes have been analysed previously (Ronquist et al., 2015), although it would be valuable to reassess the morphological and biological evidence and re-analyse it in the light of the new phylogenomic results. The same applies to potential apomorphies for the Cynipidae (s. str.). In the latter case, there are no known apomorphies at present.

Although there is growing evidence that the Phanacidini and Aulacideini are sister groups, we prefer to keep them as separate tribes (in contrast to Blaimer et al., 2020), as there are distinct biological and morphological differences between the groups. The Phanacidini tend to be small and elongate species, and most of them are stem gallers. The Aulacideini tend to be larger and their body form is more rounded. They usually induce galls in other plant parts.

Although one could argue for the grouping of tribes within the Cynipidae (s. str.) into subfamilies, we consider it premature to do so at the current time. In particular, it would be advantageous if the position of the Eschatocerini could be determined unambiguously before further refinement of the classification is considered. Thus, for now, we argue that all extant tribes of the family should remain in a single subfamily, the Cynipinae (s. str.).

DISCUSSION

Alignment quality and phylogenetic signal

Assembling genomes or transcriptomes from short sequence reads and finding single-copy orthologs in those assemblies are challenging

tasks. Thus, one might expect some variation in the quality of the resulting gene datasets. There is a plethora of tools for aligning the gene sequences in those datasets, and for filtering out alignment sites or sequences that may provide noisy or misleading phylogenetic signal. Nevertheless, it may be difficult to eliminate such data issues. Our phylogenetic results varied depending on which gene alignments were included but were consistent for the high-quality alignments, regardless of method used to identify the latter (alignment completeness, Gblocks results, HmmCleaner results, Gblocks + HmmCleaner results, or Gblocks + OD-Seq results). This suggests that we had problems with misleading phylogenetic signal in poor alignments, rather than true conflict between different gene trees. This is also supported by the fact that the four taxa that were apparently incorrectly grouped together in analyses including poor alignments (*Eschatocerus*, *Iraella*, *Phanacis* and *Protobalandricus*) also were represented by some of the most incomplete genome assemblies. It is interesting to note that the taxa represented by transcriptomes (*Biorhiza* and *Ganaspis*) were not affected by similar problems with unstable phylogenetic placements, despite the rather incomplete representation of the genome in these transcriptomes (Table S2). This, in addition, supports the conclusion that some alignments included misleading phylogenetic signal from poor-quality genome assemblies and gives some confidence in the tree resulting from analysis of the high-quality alignments.

Interestingly, our results also suggest that quality filtering tools, such as the ones we tested (Gblocks, HmmCleaner and OD-Seq), are better at identifying problem alignments than they are at filtering out erroneous or misleading sites and sequences. None of these tools were able to remove the misleading phylogenetic signal from the poor alignments, although they might have had some positive effect.

The ultimate cause of the discordant phylogenetic signal remains unclear. The four problematic taxa may group together in some analyses simply because they share divergent or incorrect sequences for some genes. The signal could be entirely erroneous, for example, through sharing of specific gene pairs that can easily be merged into chimeric sequences in challenging genome assemblies, resulting in positively misleading phylogenetic signal that groups them together.

As several alternative approaches to filtering out poor gene alignments gave consistent end results, we are fairly confident that our phylogenetic analysis is not misled by erroneous genome assemblies. It is more difficult to exclude the possibility of shortcomings in the substitution model used for probabilistic inference resulting in artificial long-branch attraction. Resolving cynipoid relationships involves determining the branching order of several long-branch taxa (i.e., groups linked to other members of the taxon set by a long, non-dividing branch inserting deep in the phylogeny), including the Eschatocerini, Paraulacidae and Diplolepididae. The problem is accentuated by the long evolutionary distance between known cynipoid and out-group genomes. Using models that accommodate among-site variation in amino acid profiles, we applied some of the best available tools for resolving long-branch attraction due to model shortcomings (Kapli & Telford, 2020). We also note that removal of long-branch taxa in various combinations revealed no sign of an alternative phylogenetic signal obscured by long-branch attraction effects (Figure 5).

Phylogenetic relationships

The phylogenetic results from our analysis are largely congruent with and complement those of the earlier UCE analysis (Blaimer et al., 2020). Here we highlight the major agreements and disagreements between these two phylogenomic studies.

- i. *Division of Cynipidae into three families and placement of Eschatocerini*: Both studies support division of the Cynipidae into three separate lineages—recognised here as the families Paraulacidae, Diplolepididae and Cynipidae (s. str.). However, the evidence on the placement of Eschatocerini is slightly different. Our analysis suggests that the Eschatocerini belong to the Cynipidae (s. str.), forming the sister group of the remaining lineages in that clade, while the UCE analysis favoured a sister-group relationship between the Eschatocerini and Figitidae. Interestingly, the extended UCE analysis presented here (Figure S7) strongly supported the *Eschatocerus* position favoured in our main analysis. However, as the *Eschatocerus* genome assembly is one of the least complete in our study, further genomic sequencing of this taxon would be highly desirable.
- ii. *Rejection of monophyly of cynipid herb galls*: Our study provides even stronger support for the conclusion of the UCE analysis (Blaimer et al., 2020) that the herb-galling clade of Aulacideini + Phanacidini is monophyletic. Both analyses are consistent with previous studies suggesting that these two tribes are reciprocally monophyletic (Liljeblad & Ronquist, 1998; Ronquist et al., 2015). Unlike Blaimer et al. (2020), we prefer to keep the tribes Aulacideini and Phanacidini separate until there is evidence that this would conflict with phylogenetic relationships. Blaimer et al. (2020) also concluded that the cynipid herb galls (apart from a few species in the tribe Diastrophini) form a monophyletic clade, Aylacini (s. lat.), which is the sister group of all remaining Cynipidae (s. str.). As mentioned above, this interpretation is based on the incorrect assumption that *Nealyx salviae* (which they name *Aylax salviae*) belongs to the Aylacini (s. str.). In fact, this species belongs to a clade of Lamiaceae galls in the Aulacideini (Ronquist et al., 2015) and is unrelated to the true Aylacini (s. str.), all known species of which are associated with poppies (Papaveraceae). Our study is the first phylogenomic analysis to include a true representative of the Aylacini (s. str.), *Iraella hispanica* Nieves-Aldrey, and our analysis clearly shows that herb galls in Aylacini (s. str.) and in Aulacideini + Phanacidini are phylogenetically divergent. Instead, Aylacini (s. str.) is deeply nested within the sister-group of Aulacideini + Phanacidini, a clade that is dominated by inquilines and gall inducers associated with woody rosids (the only exception being a few species of Diastrophini that are galls of herbs in the genus *Potentilla*). Thus, galling of herbs in the family Papaveraceae by the Aylacini (s. str.) appears to be secondary. Our results are consistent with several earlier analyses suggesting that the Aylacini (s. str.) form a lineage that is distinct from that of the Aulacideini and Phanacidini (Liljeblad & Ronquist, 1998;

Nylander et al., 2004; Ronquist et al., 2015), and they agree with preliminary analyses of a recent genome assembly of *Aylax minor* Hartig, another member of the Aylacini (s. str.) (AB, unpublished data).

- iii. *Phylogenetic placement of the Qwaqwaiini*: Ours is the first phylogenomic analysis to include the Qwaqwaiini. Our analysis places *Qwaqwaiia scolopiae*, the only known species in the Qwaqwaiini, as the sister group of the inquiline clade consisting of Synergini (s. str.). The extended UCE analysis (Figure S7) further suggests that the Qwaqwaiini forms the sister group of Rhoophilini, and that these two groups together form the sister group of the Synergini (s. str.). To date, the Qwaqwaiini and Rhoophilini are the only two indigenous lineages of Cynipidae (s. str.) known from the afro-tropical zone. This could potentially indicate a deep vicariance event involving the split between this afro-tropical group and the Holarctic Synergini (s. str.).
- iv. *Relationships in Figitidae*: Our sampling of Figitidae is not as extensive as Blaimer et al.'s (2020) UCE analysis, but our results are entirely consistent for all taxa that overlap. In both analyses, the Parnipinae is the sister group to all other Figitidae (s. lat., that is, including Liopteridae and Ibalidae), the Charipinae (*Phaenoglyphis* and *Alloxysta* in our analysis) are monophyletic, the Diptera-associated lineages (Aspicerinae (*Callaspidia*) and Eucoilinae (*Ganaspis* and *Leptopilina*) in our analysis) form a monophyletic group, and the Eucoilinae are monophyletic. As our analysis did not include any representatives of Liopteridae and Ibalidae, we cannot comment on their placement. Neither our analysis nor any other molecular phylogenetic analysis has yet included representatives of the rare Australian Austrocynipidae, which is assumed to be the sister group of all other cynipoids (Ronquist, 1995, 1999).

Evolutionary implications

We end by discussing some alternative scenarios for the origin of cynipoid gall inducers and inquilines in the new light shed on this problem by the phylogenomic analyses and by recent findings on the life history of key groups. We aim to identify the most critical knowledge gaps and the most promising avenues towards making further progress in the quest of understanding the early evolution of cynipoids.

Transitions between phytophagous and parasitoid life cycles

The phylogenomic results suggest two alternative scenarios for the origin of gall inducers and inquilines from insect-parasitic ancestors, both appearing among the reconstructions presented by Blaimer et al. (2020). One scenario (independent phytophagy) assumes that the parasitoid life history traces back to the ancestral cynipoid (Figure 10a). If so, then herbivory (inquilinism/gall induction) must have originated twice from such ancestors (in Diplolepididae and Cynipidae s. str.). The other scenario (parasitoid reversal) assumes instead that it is the

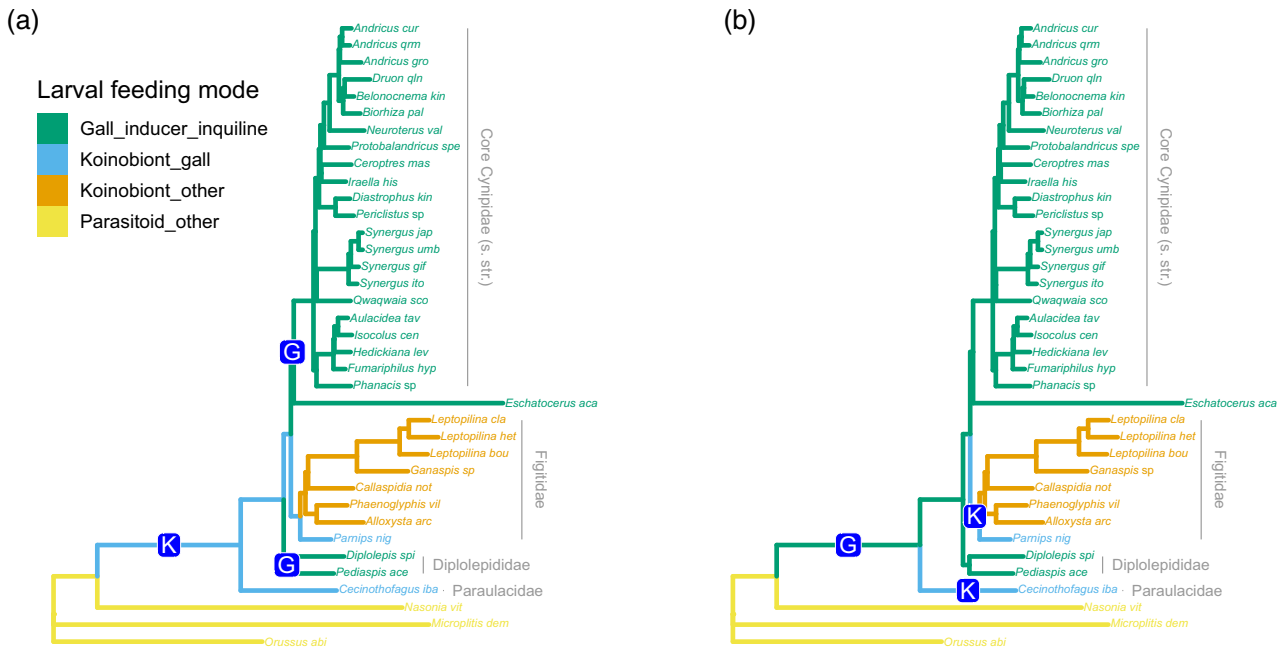


FIGURE 10 Two possible scenarios for the origin of major life history types in the Cynipoidea. (a) *Independent phytophagy scenario*. The ancestor of cynipoids was a koinobiont endoparasitoid (at least in early instars) of gall-inducing insects (orange lineages, origin of koinobionts marked with “K”). The ancestral life history persists today in the Paraulacidae and basal lineages of Figitidae (s. lat.), like the Parnipinae. Gall inducers and inquilines originated twice from these koinobionts of gall insects (“G”). (b) *Parasitoid reversal scenario*. The koinobiont endoparasitoids of gall insects (“K”) evolved independently in the Paraulacidae and Figitidae (s. lat.), possibly in both cases from phytophagous gall inducers and inquilines (“G”). In both scenarios, advanced figitid lineages (in red) remained koinobiont parasitoids of insects but colonised hosts in other environments.

phytophagous habit that traces back to the cynipoid ancestor (Figure 10b). If this scenario is correct, then parasitoids must have evolved twice independently from phytophagous ancestors (in Paraulacidae and Figitidae s. lat.).

The evidence for or against these scenarios hinges critically on the life history and phylogenetic placement of two lineages, the Paraulacidae and Eschatocerini. The Paraulacidae are associated with *Nothofagus* galls induced by the chalcidoid genus *Aditrochus*, but it has been unclear whether they are phytophagous inquilines or parasitoids of some other gall inhabitant (Nieves-Aldrey et al., 2009; Ronquist et al., 2015), which has added considerable uncertainty to the reconstruction of evolutionary scenarios (Blaimer et al., 2020; Ronquist et al., 2015). Interestingly, recent genetic analyses have provided a case where the genome of a paraulacid, *Cecinothofagus ibarrai*, was retrieved from a larva of *Aditrochus coihuensis* Ovruski, together with the *Aditrochus coihuensis* genome (Rasplus et al., 2022). This suggests that the Paraulacidae are not only parasitoids but they are also likely koinobiont endoparasitoids in early larval instars, like all other insect-parasitic cynipoids in the sister group of Cynipidae (s. str.) including Parnipinae, Ibalidae and numerous Figitidae (Ronquist, 1999, and references cited therein; Ronquist et al., 2018). Thus, if the parasitoid reversal scenario is correct, a complex koinobiont life history must have evolved twice independently in the Cynipoidea from phytophagous ancestors, which appears unlikely if not entirely impossible.

Another important piece in the puzzle is the life history of the Eschatocerini. They have been reared from galls on *Prosopis* L. and *Vachellia* (formerly *Acacia* Mill.) collected in Argentina and Chile, and have generally been assumed to be gall inducers (Aranda-Rickett et al., 2017; Nieves-Aldrey & San, 2015). However, like the *Nothofagus* galls hosting the Paraulacidae, these galls also produce a number of other insects that could potentially be gall inducers. These include *Allorhogas prosopidis* (Kieffer & Jörgensen) (Braconidae), a genus of phytophagous braconids that may be inquilines or gall inducers (Samacá-Sáenz et al., 2020), and the chalcidoid *Tanaostigmodes coeruleus* (Kieffer) (Chalcidoidea, Tanaostigmatidae), which belongs to a genus that is known to include phytophagous species (either inquilines or gall inducers). The galls are also inhabited by members of several genera of eurytomids, namely *Proseurytoma* Kieffer, *Sycophila* Walker and *Eurytoma* Illiger. Other members of these genera include true gall inducers, such as *Proseurytoma gallarum* Kieffer, a gall inducer on *Geoffroea decorticans* (Gillies ex Hook. & Arn.) Burkart (another Fabaceae sharing habitats with *Prosopis* and *Acacia*). Preliminary data available to one of us (JLNA) suggest that *Allorhogas* Gahan and *Tanaostigmodes* Ashmead are both inquilines in *Eschatocerus* galls, which is at least consistent with the common assumption (also adopted in this paper) that *Eschatocerus* is the true gall inducer. However, additional evidence supporting this conclusion would be highly desirable. The possibility of Eschatocerini being parasitoids cannot be

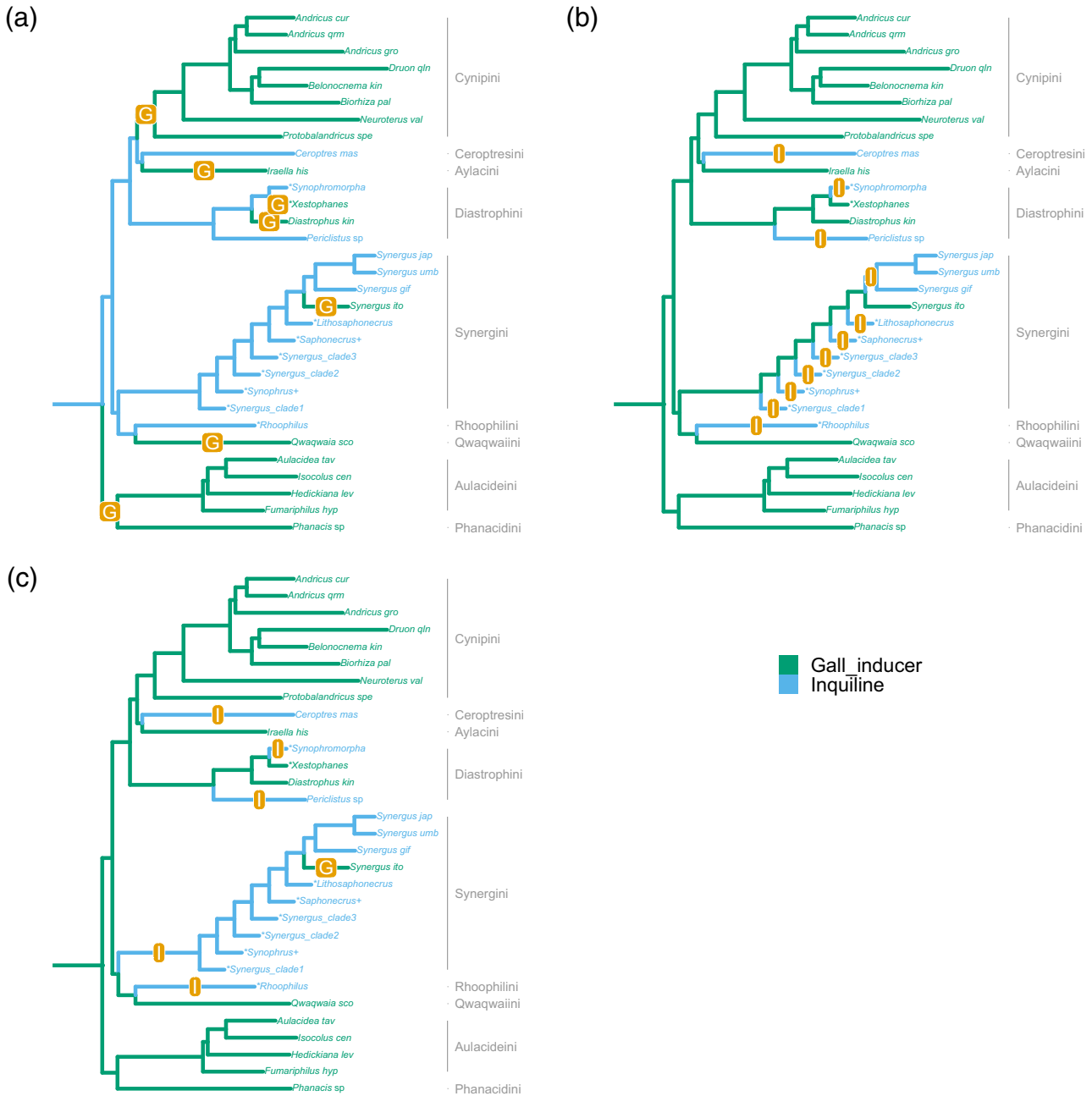


FIGURE 11 Three possible scenarios for the origin of gall inducers (green lineages) and inquiline (blue lineages) in the Cynipidae (s. str.). The tree is based on the analysis presented here, augmented with additional taxa (preceded with “**”) based on the extended UCE analysis (Figure S7) and other recent analyses (Blaimer et al., 2020; Lobato-Vila et al., 2022; Ronquist et al., 2015). *Eschatocerus* was excluded from the tree because of uncertainty concerning its life history. Added clades with members from more than one genus marked with ‘+’. (a) *Inquiline-first scenario*. Gall inducers evolved repeatedly from inquiline, which always represent an intermediate stage in the origin of true gall inducers. (b) *Gallers-first scenario*. In this scenario, inquiline always represent gall inducers that have lost the ability to initiate galls. (c) *Parsimonious galler-first scenario*. One of the most parsimonious reconstructions, weighting transitions equally, largely agrees with the gallers-first scenario but suggests that *Synergus itoensis* and related species regained the ability to induce galls on their own.

excluded, in which case the independent phytophagy scenario (Figure 10a) would gain additional support.

Interestingly, both Dipolepididae and Cynipidae (s. str.) include species whose genomes encode plant cell wall degrading enzyme genes (Hearn et al., 2019). These may have been acquired from an herbivorous shared common ancestor, or alternatively they may be

essential components of cynipid herbivory that have been acquired convergently during independent evolution of galling lifestyles. Analyses of whether such complex genomic features associated with the two different life histories are likely to have a shared history or separate origins provides one of the most promising ways of distinguishing between the two possible scenarios.

Transitions between gall-inducing and inquiline life cycles

The evolution of gall inducers and inquilines is clearly more complex than the origin of phytophagy, even though inquilines are only known from the Cynipidae (s. str.). If the transitions between gall inducers and inquilines were always in one direction, there would be two alternative scenarios (see also Blaimer et al., 2020; Ronquist et al., 2015). If gall inducers always evolved from inquilines (inquilines-first scenario), then gall inducers must have evolved at least seven times independently, taking all available phylogenetic evidence into account (Figure 11a). At the other extreme, if all inquilines evolved from gall inducers (gallers-first scenario), there would have been at least 11 independent origins of inquiline cynipids from gall inducers (Figure 11b).

Thus, if we assume that transitions have always been in one direction, then the inquilines-first scenario appears more likely (Figure 11a). However, the only transition that appears to be clearly supported by phylogenetic evidence at this point is the origin of gall induction by *Synergus itoensis* and close relatives from inquiline ancestry within the Synergini (s. str.), because the gall inducers are so deeply nested within inquiline lineages (Ide et al., 2018). If we are willing to assume that this represents a reversal from an inquiline life cycle, then the gallers-first scenario provides a more parsimonious explanation of the remaining transitions (Figure 11c; note that there are alternative reconstructions with the same total number of changes but with more independent origins of gall inducers).

Which hypothesis is better supported depends crucially on the relative ease (in evolutionary terms) or weight (in terms of inferred state changes) of transitions between the alternative states of gall induction and inquiline life cycles (Stone & French, 2003). While both gall inducers and inquiline cynipids can cause the development of nutritive gall tissues on which the larvae feed, only true inducers can cause the development of gall tissues de novo, and the development of the structurally complex outer gall tissues that characterise many cynipid galls. If it is easier to transition from full gall induction to a simpler inquiline life history than vice versa, then a gallers-first scenario may be more likely a priori. Alternatively, it might be a relatively minor step in evolutionary terms for cynipids to transition from inquiline to becoming a gall inducer. We currently know too little about the differences between these alternative life histories to provide any clear weighting of transition probabilities between them, beyond suspecting that unweighted parsimony may potentially be an unreliable guide. While the evolution of gall induction in *Synergus itoensis* shows that gall induction can evolve from inquiline, the galls they induce consist only of nutritive tissues and lack morphologically complex non-nutritive tissues. Some Synergini inquilines do modify the complex gall morphology of host galls usurped at a very early stage in their development (Pérez et al., 2009), but no case is yet known of a shift from inquiline to gall induction that also includes ability to induce complex gall phenotypes.

Again, the life history of some key taxa is important in weighing these alternative scenarios. Demonstration that Eschatocerini are

inquilines would strengthen the inquilines-first scenario, while demonstration that they are true gall inducers would strengthen support for the gallers-first scenario. The Ceroptresini are commonly assumed to be inquilines, but detailed studies of their life history are sorely lacking, and there is one report claiming that members of *Ceroptres* are true gall inducers (Blair, 1949). The Qwaqwaiini is another taxon for which more detailed life history information would be valuable. According to the only existing report, it is a gall inducer (Liljeblad et al., 2011), but it remains possible that it could be an inquiline, like most members of the Synergini (s. str.) and Rhoophilini. Such a demonstration would strengthen support for the inquilines-first scenario by removing one of the independent origins of gall inducers. That inquilines can easily originate from gall inducers is suggested by observations of facultative intraspecific inquiline in *Diastrophus* (Diastrophini) (Pujade-Villar, 1984), and it has also been suggested that the remarkable parallelisms between the Aulacideini and Phanacidiini in the evolution of host plant preferences could be due to facultative or obligate inquiline among some cynipid herb gallers (Nieves-Aldrey et al., 2004; Ronquist & Liljeblad, 2001). Finally, we note that an ancestral state for inquiline in Cynipidae (s. str.) requires that the ancestral host was not itself a cynipid gall inducer. While rare examples of inquiline cynipids developing in non-cynipid galls are known (Askew, 1999; van Noort et al., 2007), it is notable that the vast majority of inquiline cynipids develop in cynipid galls.

Transitions between strikingly different life histories, such as those between koinobiont endoparasitoids, gall inducers and inquilines in cynipoids, should have major effects on genomes. For instance, transitions to or from a koinobiont endoparasitic life history should involve recruitment or loss of a swathe of genes or gene functions associated, for example, with suppressing or evading host immune systems, maintaining basic physiological functions within a host body and adjusting larval development and feeding patterns so that the host larva survives and develops normally as long as possible. This should be noticeable as an unusual number of protein-coding genes with markedly increased or decreased rates of non-synonymous rates of evolution along branches of the phylogeny involving life history changes. Similarly, the genes undergoing unusual amounts of change should also belong to particular functional categories. Transitions between gall inducers and inquilines may be less dramatic but should nevertheless leave similar genomic signatures. A recent study suggests that this is indeed the case for the transition from inquilines to gall inducers in species related to *Synergus itoensis* (Gobbo et al., 2020). Whether such genomic signatures of life history transitions can be detected deeper down in the cynipoid tree remains unclear. However, this is clearly a possibility that is well worth investigating, and the genomic data reported here represent a first step in supporting such a line of research.

AUTHOR CONTRIBUTIONS

Fredrik Ronquist: Funding acquisition; writing – original draft; supervision; resources; project administration; software; formal analysis; visualization; writing – review and editing; methodology; data curation. **Jack Hearn:** Conceptualization; investigation; writing – original

draft; validation; methodology; software; formal analysis; data curation; writing – review and editing. **Erik Gobbo**: Investigation; writing – original draft; methodology; software; formal analysis; validation; data curation; writing – review and editing. **José Luis Nieves-Aldrey**: Investigation; validation; writing – review and editing; resources. **Antoine Branca**: Methodology; visualization; formal analysis; software; writing – review and editing. **James A. Nicholls**: Investigation; writing – review and editing; methodology; software; formal analysis. **Georgios Koutsovoulos**: Investigation; writing – review and editing; methodology; software; formal analysis. **Nicolas Lartillot**: Methodology; software; formal analysis; writing – review and editing. **Graham N. Stone**: Conceptualization; funding acquisition; writing – original draft; writing – review and editing; project administration; supervision; resources.

ACKNOWLEDGEMENTS

We would like to thank Jean-Yves Rasplus and Astrid Cruaud for their willingness to share barcoding and UCE data, providing critical clues to the life history of *Cecinothofagus* prior to publication. Jean-Yves Rasplus, Matt Buffington, Bonnie Blaimer and three anonymous reviewers provided many valuable comments on the manuscript.

FUNDING INFORMATION

Support for genome sequencing was obtained from the National Genomics Infrastructure and Science for Life Laboratory (Project ID P14912). For computation, we used resources from the Swedish National Infrastructure for Computing at UPPMAX (projects 2017/7-283 and 2017097) and at NSC (projects 2021/5-118 and 2021/23-157), partially funded by the Swedish Research Council (2018-05973). Additional support was obtained from the European Union's Horizon 2020 research and innovation program, Marie Skłodowska Curie Actions (642241 to Erik Gobbo and Fredrik Ronquist); the Swedish Research Council (2018-04620 to Fredrik Ronquist); the UK Natural Environment Research Council (NE/J010499 and NBAF375 to Graham N. Stone); the French National Research Agency (ANR-15-CE12-0010-01/DASIRE to Nicolas Lartillot and ANR-19-CE02-0008 to Antoine Branca). José Luis Nieves-Aldrey was supported by the research project MINECO/FEDER, UE CGL2015-66571-P.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Raw sequencing data is available under EBI Bioprojects PRJEB13424, PRJEB45812, PRJEB51101 and PRJEB37996. Scripts, datasets, and result files are available from https://github.com/ronquistlab/cynipoid_phylogenomics.

ORCID

Jack Hearn  <https://orcid.org/0000-0003-3358-4949>

Erik Gobbo  <https://orcid.org/0000-0001-9897-8610>

José Luis Nieves-Aldrey  <https://orcid.org/0000-0002-4711-7455>

Antoine Branca  <https://orcid.org/0000-0001-8475-1386>

James A. Nicholls  <https://orcid.org/0000-0002-9325-563X>

Georgios Koutsovoulos  <https://orcid.org/0000-0003-3406-3715>

Nicolas Lartillot  <https://orcid.org/0000-0002-9973-7760>

Graham N. Stone  <https://orcid.org/0000-0002-2737-696X>

Fredrik Ronquist  <https://orcid.org/0000-0002-3929-251X>

REFERENCES

- Abe, Y., Ide, T. & Wachi, N. (2011) Discovery of a new gall-inducing species in the inquiline tribe Synergini (Hymenoptera: Cynipidae): inconsistent implications from biology and morphology. *Annals of the Entomological Society of America*, 104, 115–120. Available from: <https://doi.org/10.1603/AN10149>
- Abe, Y., Melika, G. & Stone, G.N. (2007) The diversity and phylogeography of cynipid gallwasps (Hymenoptera: Cynipidae) of the Oriental and Eastern Palaearctic regions, and their associated communities. *Oriental Insects*, 41, 169–212. Available from: <https://doi.org/10.1080/00305316.2007.10417504>
- Andrews, S. (2010) FastQC: A quality control application for high throughput sequence data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Aranda-Rickett, A., Rothen, C., Diez, P., Gonzalez, A. & Marazzi, B. (2017) Sugary secretions of wasp galls: a want-to-be extrafloral nectar? *Annals of Botany*, 120, 765–774. Available from: <https://doi.org/10.1093/aob/mcx075>
- Ashmead, W.H. (1903) Classification of the gall-wasps and parasitic cynipoids, or the superfamily Cynipoidea. *Psyche (Cambridge)*, 10, 210–215.
- Askew, R.R. (1999) Confirmation of an association of *Synergus* Hartig and *Saphonecrus* Dalla Torre and Kieffer (Hym., Cynipidae) with oak galls of Cecidomyiidae (Dipt.). *Entomologist's Monthly Magazine*, 135, 89–90.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19, 455–477. Available from: <https://doi.org/10.1089/cmb.2012.0021>
- Blaimer, B.B., Gotzek, D., Brady, S.G. & Buffington, M.L. (2020) Comprehensive phylogenomic analyses re-write the evolution of parasitism within cynipoid wasps. *BMC Evolutionary Biology*, 20, 155. Available from: <https://doi.org/10.1186/s12862-020-01716-2>
- Blaimer, B.B., Santos, B.F., Cruaud, A., Gates, M.W., Kula, R.R., Mikó, I. et al. (2023) Key innovations and the diversification of hymenoptera. *Nature Communications*, 14, 1212. Available from: <https://doi.org/10.1038/s41467-023-36868-4>
- Blair, K.G. (1949) *Ceroptres arator* Hartig (Hym., Cynipidae): inquiline or parasite? *Entomologist's Monthly Magazine*, 85, 254–255.
- Brandão-Dias, P.F.P., Zhang, Y.M., Pirro, S., Vinson, C.C., Weinersmith, K.L., Ward, A.K.G. et al. (2022) Describing biodiversity in the genomics era: a new species of Nearctic Cynipidae gall wasp and its genome. *Systematic Entomology*, 47, 94–112.
- Buffington, M.L. (2008) A revision of Australian Thrasorinae (Hymenoptera: Figitidae) with a description of a new genus and six new species. *Australian Journal of Entomology*, 47, 203–212.
- Buffington, M.L., Brady, S.G., Morita, S.I. & van Noort, S. (2012) Divergence estimates and early evolutionary history of Figitidae (Hymenoptera: Cynipoidea). *Systematic Entomology*, 37, 287–304. Available from: <https://doi.org/10.1111/j.1365-3113.2012.00617.x>
- Buffington, M.L. & Liljeblad, J. (2008) The description of Euceroptinae, a new subfamily of Figitidae (Hymenoptera), including a revision of Euceroptres Ashmead, 1896 and the description of a new species. *Journal of Hymenoptera Research*, 17(1), 44–56.
- Buffington, M.L. & Nieves-Aldrey, J.L. (2011) Revision of Plectocynipinae (Hymenoptera: Figitidae) with descriptions of a new genus and three

- new species from Chile. *Proceedings of the Entomological Society of Washington*, 113(2), 91–108.
- Buffington, M.L., Nylander, J.A.A. & Heraty, J.M. (2007) The phylogeny and evolution of Figitidae (Hymenoptera: Cynipoidea). *Cladistics*, 23, 403–431. Available from: <https://doi.org/10.1111/j.1096-0031.2007.00153.x>
- Cambier, S., Ginis, O., Moreau, S.J.M., Gayral, P., Hearn, J., Stone, G.N. et al. (2019) Gall wasp transcriptomes unravel potential effectors involved in molecular dialogues with oak and rose. *Frontiers in Physiology*, 10, 926. Available from: <https://doi.org/10.3389/fphys.2019.00926>
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34, i884–i890. Available from: <https://doi.org/10.1093/bioinformatics/bty560>
- Cornell, H.V. (1983) The secondary chemistry and complex morphology of galls formed by the Cynipinae (Hymenoptera): why and how? *American Midland Naturalist*, 110, 225–234. Available from: <https://doi.org/10.2307/2425263>
- Di Franco, A., Poujol, R., Baurain, D. & Philippe, H. (2019) Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evolutionary Biology*, 19, 21. Available from: <https://doi.org/10.1186/s12862-019-1350-2>
- Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T. & Glenn, T.C. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary time-scales. *Systematic Biology*, 61, 717–726.
- Geoffroy, E.L. (1762) *Histoire abrégée des insectes qui se trouvent aux environs de Paris, dans laquelle ces animaux sont rangés suivant un ordre méthodique, Tome second*. Paris: Durand, 690 pp., pls. 11–22.
- Gobbo, E., Lartillot, N., Hearn, J., Stone, G.N., Abe, Y., Wheat, C.W. et al. (2020) From inquillines to gall inducers: genomic signature of a life-style transition in *Synergus* gall wasps. *Genome Biology and Evolution*, 12, 2060–2073. Available from: <https://doi.org/10.1093/gbe/evaa204>
- Harper, L.J., Schönrogge, K., Lim, K.Y., Francis, P. & Lichtenstein, C.P. (2004) Cynipid galls: insect-induced modifications of plant development create novel plant organs. *Plant, Cell & Environment*, 27, 327–335. Available from: <https://doi.org/10.1046/j.1365-3040.2004.01145.x>
- Harris, M.O. & Pitzschke, A. (2020) Plants make galls to accommodate for-eigners: some are friends, most are foes. *New Phytologist*, 225, 1852–1872. Available from: <https://doi.org/10.1111/nph.16340>
- Hartig, T. (1840) Über die Familie der Gallwespen. *Zeitschrift für Entomologie(Germar)*, 2, 176–209.
- Hearn, J., Blaxter, M., Schönrogge, K., Nieves-Aldrey, J.L., Pujade-Villar, J., Huguet, E. et al. (2019) Genomic dissection of an extended phenotype: oak galling by a cynipid gall wasp. *PLoS Genetics*, 15, e1008398. Available from: <https://doi.org/10.1371/journal.pgen.1008398>
- Heraty, J., Ronquist, F., Carpenter, J.M., Hawks, D., Schulmeister, S., Dowling, A.P. et al. (2011) Evolution of the hymenopteran megara-diation. *Molecular Phylogenetics and Evolution*, 60, 73–88. Available from: <https://doi.org/10.1016/j.ympev.2011.04.003>
- Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522. Available from: <https://doi.org/10.1093/molbev/msx281>
- Ide, T., Kusumi, J., Miura, K. & Abe, Y. (2018) Gall inducers arose from inquillines: phylogenetic position of a gall-inducing species and its relatives in the inquiline tribe Synergini (Hymenoptera: Cynipidae). *Annals of the Entomological Society of America*, 111, 6–12. Available from: <https://doi.org/10.1093/aesa/sax065>
- Jehl, P., Sievers, F. & Higgins, D.G. (2015) OD-Seq: outlier detection in multiple sequence alignments. *BMC Bioinformatics*, 16, 269. Available from: <https://doi.org/10.1186/s12859-015-0702-1>
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermini, S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589. Available from: <https://doi.org/10.1038/nmeth.4285>
- Kapli, P. & Telford, M.J. (2020) Topology-dependent asymmetry in systematic errors affects phylogenetic placement of Ctenophora and Xenacoelomorpha. *Science Advances*, 6, eabc5162. Available from: <https://doi.org/10.1126/sciadv.abc5162>
- Kerzhner, I.C. (1991) Case 2292 histoire abregée des insectes qui se trouvent aux environs de Paris (Geoffroy, 1762): proposed conservation of some generic names (Crustacea and Insecta). *Bulletin of Zoological Nomenclature*, 48, 2.
- Kieffer, J.J. (1904) Description de quelques Cynipides exotiques dont l'un forme un genre nouveau. *Bulletin de la Société d'Histoire Naturelle de Metz*, 23, 59–66.
- Kinsey, A.C. (1920) Phylogeny of cynipid genera and biological characteristics. *Bulletin of the American Museum of Natural History*, 42, 357–402.
- Klopfstein, S., Villhelmsen, L., Heraty, J.M., Sharkey, M. & Ronquist, F. (2013) The hymenopteran tree of life: evidence from protein-coding genes and objectively aligned ribosomal data. *PLoS One*, 8, e69344. Available from: <https://doi.org/10.1371/journal.pone.0069344>
- Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., Wägele, J.W. et al. (2010) Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Frontiers in Zoology*, 7, 10. Available from: <https://doi.org/10.1186/1742-9994-7-10>
- Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2017) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34, 772–773.
- Lartillot, N. & Philippe, H. (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, 21, 1095–1109. Available from: <https://doi.org/10.1093/molbev/msh112>
- László, Z. & Tóthmérész, B. (2006) Inquiline effects on a multilocular gall community. *Acta Zoologica Academiae Scientiarum Hungaricae*, 52, 373–383.
- Latreille, P.A. (1802) *Histoire Naturelle, Générale et Particulière, des Crustacés et des Insectes*, Vol. 13. Paris: F. Dufart, 432 pp.
- Liljeblad, J., Nieves-Aldrey, J.L., Nesser, S. & Melika, G. (2011) Adding another piece to the puzzle: the description of a South African gall wasp and a new tribe (Hymenoptera: Cynipidae). *Zootaxa*, 2806, 35–52.
- Liljeblad, J. & Ronquist, F. (1998) A phylogenetic analysis of higher-level gall wasp relationships (Hymenoptera: Cynipidae). *Systematic Entomology*, 23, 229–252. Available from: <https://doi.org/10.1046/j.1365-3113.1998.00053.x>
- Liljeblad, J., Ronquist, F., Nieves-Aldrey, J.L., Fontal-Cazalla, F., Ros-Farré, P., Gaitros, D. et al. (2008) A fully web-illustrated morphological phylogenetic study of relationships among oak gall wasps and their closest relatives (Hymenoptera: Cynipidae). *Zootaxa*, 1796, 1–73.
- Linnaeus, C. (1758) *Systema naturae* (10th ed.). Stockholm: Laurentius Salvius, 824 pp.
- Lobato-Vila, I., Bae, J., Roca-Cusachs, M., Kang, M., Jung, S., Melika, G. et al. (2022) Global phylogeny of the inquiline gall wasp tribe Synergini (Hymenoptera: Cynipoidea: Cynipidae): first insights and establishment of a new cynipid tribe. *Zoological Journal of the Linnean Society*, 195, 1338–1354. Available from: <https://doi.org/10.1093/zoolinnean/zlab085>
- Martinson, E.O., Werren, J.H. & Egan, S.P. (2022) Tissue-specific gene expression shows a cynipid wasp repurposes oak host gene networks to create a complex and novel parasite-specific organ. *Molecular Ecology*, 31, 3228–3240. Available from: <https://doi.org/10.1111/mec.16159>

- Maslin, B.R., Miller, J.T. & Seigler, D.S. (2003) Overview of the generic status of *Acacia* (Leguminosae: Mimosoideae). *Australian Systematic Botany*, 16, 1–18. Available from: <https://doi.org/10.1071/SBO2008>
- Melika, G. (2006) Gall wasps of Ukraine. *Cynipidae*. *Vestnik Zoologii*, 21, 1–300.
- Melika, G. & Abrahamson, W.G. (2002) Review of the world genera of oak cynipid wasps (Hymenoptera: Cynipidae: Cynipini). In: Melika, G. & Thuróczy, C. (Eds.) *Parasitic wasps: evolution, systematics and biological control*. Budapest: Agroinform, pp. 150–190.
- Minh, B.Q., Hahn, M.W. & Lanfear, R. (2020) New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution*, 37, 2727–2733. Available from: <https://doi.org/10.1093/molbev/msaa106>
- Minh, B.Q., Nguyen, M.A.T. & von Haeseler, A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30, 1188–1195. Available from: <https://doi.org/10.1093/molbev/mst024>
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37, 1530–1534. Available from: <https://doi.org/10.1093/molbev/msaa015>
- Mo, Y.K., Lanfear, R., Hahn, M.W. & Minh, B.Q. (2022) Updated site concordance factors minimize effects of homoplasy and taxon sampling. *Bioinformatics*, 39, btac741. Available from: <https://doi.org/10.1093/bioinformatics/btac741>
- Mortimer, N.T., Goecks, J., Kacsoh, B.Z., Mobley, J.A., Bowersock, G.J., Taylor, J. et al. (2013) Parasitoid wasp venom SERCA regulates *Drosophila* calcium levels and inhibits cellular immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 9427–9432. Available from: <https://doi.org/10.1073/pnas.1222351110>
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32, 268–274. Available from: <https://doi.org/10.1093/molbev/msu300>
- Nicholls, J.A., Stone, G.N. & Melika, G. (2018) A new genus of oak gall-wasp, *Protobalandricus* Melika, Nicholls & Stone (Hymenoptera: Cynipidae: Cynipini) from California. *Zootaxa*, 4472, 141–152.
- Nieves-Aldrey, J.L. (1994) Revision of West-European genera of the tribe Aylacini Ashmead (Hymenoptera, Cynipidae). *Journal of Hymenoptera Research*, 3, 175–206.
- Nieves-Aldrey, J.L. (2001) Hymenoptera, Cynipidae. In: Ramos, M., Alba-Tercedor, J., Bellés-i-Ros, X., Gosálbez-i-Noguera, J., Guerra-Sierra, A., Macpherson-Mayol, E. et al. (Eds.) *Fauna Ibérica*. Madrid: Museo Nacional de Ciencias Naturales, CSIC, pp. 1–636.
- Nieves-Aldrey, J.L. (2022) Description of *Fumariphilus* Nieves-Aldrey, gen. nov., a new genus of herb gall wasps, with a key to genera of the tribe Aulacideini (Hymenoptera: Cynipidae). *Zootaxa*, 5155, 393–413. Available from: <https://doi.org/10.11646/zootaxa.5155.3.5>
- Nieves-Aldrey, J.L., Gómez, J.F. & Hernández, N.M. (2004) Nuevos datos sobre *Aulacidea freesei* y *Phanacis zwoelferi* (Hymenoptera, Cynipidae, Aylacini), inductores de agallas en *Silybum marianum* (Asteraceae), en la Península ibérica, incluyendo la descripción y comparación de sus larvas terminales y sus agallas. *Boletín Sociedad Entomológica Aragonesa*, 34, 85–93.
- Nieves-Aldrey, J.L., Liljeblad, J., Hernández Nieves, M., Grez, A. & Nylander, J.A.A. (2009) Revision and phylogenetics of the genus *Paraulax* Kieffer (Hymenoptera, Cynipidae) with biological notes and description of a new tribe, a new genus, and five new species. *Zootaxa*, 2200, 1–40. Available from: <https://doi.org/10.11646/zootaxa.2200.1.1>
- Nieves-Aldrey, J.L. & San, B.G. (2015) Revision of the Neotropical genus *Eschatocerus* Mayr (Hymenoptera, Cynipidae, Eschatocerini) with biological notes and the first description of the terminal larva. *Zootaxa*, 4012, 135–155. Available from: <https://doi.org/10.11646/zootaxa.4012.1.7>
- Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P. & Nieves-Aldrey, J.L. (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology*, 53, 47–67. Available from: <https://doi.org/10.1080/10635150490264699>
- Paretas-Martínez, J., Restrepo-Ortiz, C., Buffington, M. & Pujade-Villar, J. (2011) Systematics of Australian Thrasorinae (Hymenoptera, Cynipoidea, Figitidae) with descriptions of Mikeiinae, new subfamily, two new genera, and three new species. *ZooKeys*, 108, 21–48.
- Pérez, Z., Melika, G., Bozsóki, G., Bihari, P., Mikó, I., Tavakoli, M. et al. (2009) Systematic re-appraisal of the gall-urprising wasp genus *Synophrus* Hartig, 1843 (Hymenoptera: Cynipidae: Synergini). *Systematic Entomology*, 34, 688–711.
- Peters, R.S., Krogmann, L., Cx, M., Donath, A., Gunkel, S., Meusemann, K. et al. (2017) Evolutionary history of the Hymenoptera. *Current Biology*, 27, 1013–1018.
- Polidori, C. & Nieves-Aldrey, J.L. (2014) Diverse filters to sense: great variability of antennal morphology and sensillar equipment in gall wasps (Hymenoptera: Cynipidae). *PLoS One*, 9, e101843. Available from: <https://doi.org/10.1371/journal.pone.0101843>
- Pujade-Villar, J. (1984) Estudi del comportament individual de *Diastrophus rubi* (Bouché) (Hymenoptera, Cynipinae). *Sessió Conjunta d'Entomologia ICHN-SCL*, 3, 125–132.
- Pujade-Villar, J., Wang, Y., Zhang, W., Mata-Casanova, N., Lobato-Vila, I., Dénes, A.-L. et al. (2020) A new *Diplolepis* (Hymenoptera, Cynipidae, Diplolepidini) species from China: a rare example of a rose gall-inducer of economic significance. *ZooKeys*, 904, 131–146. Available from: <https://doi.org/10.3897/zookeys.904.46547>
- Rasplus, J.Y., Nieves-Aldrey, J.L. & Cruaud, A. (2022) *Cecinothofagus* Nieves-Aldrey & Liljeblad (Hymenoptera, Cynipidae) is likely an endoparasitoid of the gall-maker genus *Aditrochus* Rübtsaamen (Hymenoptera, Pteromalidae). *Journal of Hymenoptera Research*, 93, 33–42. Available from: <https://doi.org/10.3897/jhr.93.89507>
- Ronquist, F. (1994) Evolution of parasitism among closely related species: phylogenetic relationships and the origin of Inquilinism in gall wasps (Hymenoptera: Cynipidae). *Evolution*, 48(2), 241–266.
- Ronquist, F. (1995) Phylogeny and early evolution of the Cynipoidea (Hymenoptera). *Systematic Entomology*, 20, 309–335. Available from: <https://doi.org/10.1111/j.1365-3113.1995.tb00099.x>
- Ronquist, F. (1999) Phylogeny, classification and evolution of the Cynipoidea. *Zoologica Scripta*, 28, 139–164. Available from: <https://doi.org/10.1046/j.1463-6409.1999.00022.x>
- Ronquist, F. & Liljeblad, J. (2001) Evolution of the gall wasp-host plant association. *Evolution*, 55, 2503–2522. Available from: <https://doi.org/10.1111/j.0014-3820.2001.tb00765.x>
- Ronquist, F. & Nieves-Aldrey, J.L. (2001) A new subfamily of Figitidae (Hymenoptera, Cynipoidea). *Zoological Journal of the Linnean Society*, 133, 483–494.
- Ronquist, F., Nieves-Aldrey, J.L., Buffington, M.L., Liu, Z., Liljeblad, J. & Nylander, J.A.A. (2015) Phylogeny, evolution and classification of gall wasps: the plot thickens. *PLoS One*, 10, e0123301. Available from: <https://doi.org/10.1371/journal.pone.0123301>
- Ronquist, F., Nylander, J.A.A., Vårdal, H. & Nieves-Aldrey, J.L. (2018) Life history of *Parnips* and the evolutionary origin of gall wasps. *Journal of Hymenoptera Research*, 65, 91–110. Available from: <https://doi.org/10.3897/jhr.65.24115>
- Ros Farre, P.R. & Pujade-Villar, J. (2007) Plectocynipinae, a new subfamily of Figitidae and description of a new Neotropical genus of Thrasorinae (Hymenoptera: Cynipoidea). *Zootaxa*, 1583, 1–13.
- Salichos, L. & Rokas, A. (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, 497, 327–331.
- Samacá-Sáenz, E., Egan, S.P. & Zaldívar-Riverón, A. (2020) Species diversity in the braconid wasp genus *Allorhogas* (Doryctinae) associated with cynipid galls on live oaks (*Quercus*: Fagaceae) using

- natural history, phylogenetics, and morphology. *Insect Systematics and Diversity*, 4: 3, 1–20. Available from: <https://doi.org/10.1093/isd/ixaa011>
- Sanver, D. & Hawkins, B.A. (2000) Galls as habitats: the inquiline communities of insect galls. *Basic and Applied Ecology*, 1, 3–11. Available from: <https://doi.org/10.1078/1439-1791-00001>
- Sharkey, M.J., Carpenter, J.M., Vilhelmsen, L., Heraty, J., Liljeblad, J., Dowling, A.P.G. et al. (2011) Phylogenetic relationships among superfamilies of Hymenoptera. *Cladistics*, 27, 1–33. Available from: <https://doi.org/10.1111/j.1096-0031.2011.00366.x>
- Shorthouse, J.D. (1973) The insect community associated with rose galls of *Diplolepis polita* (Cynipidae, Hymenoptera). *Quaestiones Entomologicae*, 9, 55–98.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539. Available from: <https://doi.org/10.1038/msb.2011.75>
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210–3212.
- Stone, G.N. & French, V. (2003) Evolution: have wings come, gone, and come again? *Current Biology*, 13, 436–438.
- Stone, G.N. & Schönrogge, K. (2003) The adaptive significance of insect gall morphology. *Trends in Ecology & Evolution*, 18, 512–522.
- Stone, G.N., Schönrogge, K., Atkinson, R.J., Bellido, D. & Pujade-Villar, J. (2002) The population biology of gall wasps (Hymenoptera: Cynipidae). *Annual Review of Entomology*, 47, 633–668.
- Talavera, G. & Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, 56, 564–577. Available from: <https://doi.org/10.1080/10635150701472164>
- Tooker, J.F. & Helms, A.M. (2014) Phytohormone dynamics associated with gall insects, and their potential role in the evolution of the gall-inducing habit. *Journal of Chemical Ecology*, 40, 742–753. Available from: <https://doi.org/10.1007/s10886-014-0457-6>
- van Noort, S., Stone, G.N., Whitehead, V.B. & Nieves-Aldrey, J.L. (2007) Biology of *Rhoophilus loewi* (Hymenoptera: Cynipoidea: Cynipidae), with implications for the evolution of inquilineism in gall wasps. *Biological Journal of the Linnean Society*, 90, 153–172.
- Vyrzhikovskaja, A.V. (1963) New gall wasps (Hymenoptera: Cynipidae) from dog-rose in Central Asia and Kazakhstan. *Entomologicheskoye Obozrenije*, 42, 651–659.
- Wang, H.-C., Minh, B.Q., Susko, E. & Roger, A.J. (2018) Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Systematic Biology*, 67, 216–235. Available from: <https://doi.org/10.1093/sysbio/syx068>
- Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G. et al. (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, 35, 543–548.
- Yoshimoto, C. (1970) A new subfamily of Cynipoidea (Hymenoptera) from Nepal. *Canadian Entomologist*, 86, 145–154.
- Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T.-Y. (2017) GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8, 28–36. Available from: <https://doi.org/10.1111/2041-210X.12628>
- Zhang, Y.M., Buffington, M.L., Looney, C., László, Z., Shorthouse, J.D., Ide, T. et al. (2020) UCE data reveal multiple origins of rose galls in North America: global phylogeny of *Diplolepis* Geoffroy (Hymenoptera: Cynipidae). *Molecular Phylogenetics and Evolution*, 153, 106949. Available from: <https://doi.org/10.1016/j.ympev.2020.106949>
- Zhang, Y.M., Egan, S.P., Driscoe, A.L. & Ott, J.R. (2021) One hundred and sixty years of taxonomic confusion resolved: *Belonocnema* (Hymenoptera: Cynipidae: Cynipini) gall wasps associated with live oaks in the USA. *Zoological Journal of the Linnean Society*, 193, 1234–1255. Available from: <https://doi.org/10.1093/zoolinnea/zlab001>
- Zheng, G.X., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M. et al. (2016) Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature Biotechnology*, 34, 303–311. Available from: <https://doi.org/10.1038/nbt.3432>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Phylogenetic results for six equally-sized, quality-ranked subsets of the T34-G542 dataset, analysed using IQ-Tree under the C60 + I + G5 model. The raw alignments were subjected to filtering and quality ranking by HmmCleaner only. Support values (ultrafast bootstrap) are shown on branches only if they are less than 100%. (A) Less than 1.8% of sites filtered out (best quality). (B) From 1.8% to 3.5% filtered out. (C) From 3.5% to 4.9% filtered out. (D) From 4.9% to 7.4% filtered out. (E) From 7.4% to 11% filtered out. (F) More than 11% filtered out (worst quality). The two best subsets (A, B) yield congruent results except for the position of *Eschatocerus*, which varies slightly but without strong conflict in support values. The results for these subsets are also consistent with those from the best alignments using other filtering and quality ranking criteria (Figures 3, S2 and S3).

Figure S2. Phylogenetic results for six equally-sized, quality-ranked subsets of the T34-G542 dataset, analysed using IQ-Tree under the C60 + I + G5 model. The raw alignments were subjected to filtering and quality ranking by Gblocks and HmmCleaner. Support values (ultrafast bootstrap) are shown on branches only if they are less than 100%. (A) Less than 16.5% of sites filtered out (best quality). (B) From 16.5% to 28.8% filtered out. (C) From 28.8% to 39.6% filtered out. (D) From 39.6% to 52.2% filtered out. (E) From 52.2% to 64.1% filtered out. (F) More than 64.1% filtered out (worst quality). The two best subsets (A, B) yield congruent results except for the position of *Eschatocerus*, which varies slightly but without strong conflict in support values. The results for these subsets are also consistent with those from the best alignments using other filtering and quality ranking criteria (Figures 3, S1 and S3).

Figure S3. Phylogenetic results for six quality-ranked subsets of the T34-G542 dataset, analysed using IQ-Tree under the C60 + I + G5 model. The raw alignments were subjected to filtering and quality ranking by OD-Seq only. Support values (ultrafast bootstrap) are shown on branches only if they are less than 100%. (A) No outlier sequence removed (best quality). (B) One to three sequences removed. (C) Four or five sequences removed. (D) Six or seven sequences removed. (E) Eight or nine sequences removed. (F) Ten or more sequences removed (worst quality). Because of the discrete nature of the OD-Seq filtering (number of sequences removed), the best quality subset (A) is significantly smaller than the others,

explaining the lower support values. Except for some variations with weak support, the two best subsets yield phylogenetic results that are congruent with each other and with those from the best alignments using other filtering and quality ranking criteria (Figures 3, S1 and S2).

Figure S4. Testing the potential effect of long-branch taxa on phylogenetic results using IQ-Tree under the C60 + I + G5 model rather than PhyloBayes. Otherwise, this is an identical replicate of the analyses shown in the Main Text (Figure 4). Branch support values (ultrafast bootstrap) are only shown if they are less than 100%. (A) Analysis of the full taxon set. (B) *Cecinothofagus* excluded. (C) *Eschatocerus* excluded. (D) Outgroups excluded. (E) *Cecinothofagus* and *Eschatocerus* excluded. (F) *Eschatocerus* and outgroups excluded. Results are identical to those obtained with PhyloBayes and the CAT-F81 model (Figure 4), except a minor and weakly supported difference in the positions of *Ceroptres* and *Iraella* when *Eschatocerus* and outgroups are excluded (F).

Figure S5. Results from gene tree concordance analysis. (A) Gene tree consistency indices. (B) Gene tree concordance factors. (C) Site concordance factors. The shading of nodes corresponds to the certainty or concordance factor values, with darker shades corresponding to higher values.

Figure S6. Comparison of IQ-Tree C60 analyses using different criteria for taxon completeness and alignment quality. (A) Analysis of the T34-G532 dataset, including the alignments with less than 26% of site columns removed by Gblocks. (B) Analysis of the T30-G1898 dataset, including the alignments with less than 10% of site columns removed by Gblocks. The latter analysis provides notably less robust resolution of the Qwaqwaiini (66% vs. 98% bootstrap support) and the Aulacideini (90% vs. 96% for a different clade). The positions of *Ceroptresini* and *Eschatocerini* remain uncertain in both analyses.

Figure S7. Results from the IQ-Tree analysis of the combined UCE data from this study and that of Blaimer et al. (2020). Branch support values (ultrafast bootstrap) are only shown if they are less than 100%. Note that the combined analysis strongly supports a sister-group relationship between Qwaqwaiini and Rhoophilini, and a sister-group relationship between *Eschatocerini* and remaining Cynipidae (s. str.), and that Aylacini (s. str.) is not closely related to Phanacidini + Aulacideini.

Table S1. Species metadata and assembly statistics for the genomes and transcriptomes used in this study. Columns contain information on collection location where available, whether the data was assembled here or previously, and links to raw read data. Genome/transcriptome assembly quality metrics include complete and partial BUSCO (version 4.0.6) scores for both Hymenoptera and Eukaryota datasets respectively, assembly size, N50 and number of contigs, repeated for contigs >200 bp.

Table S2. Completeness of different BUSCO gene sets per taxon. The table gives the number of genes of each gene set present in each taxon, as well as the proportion of the total set being present in the taxon. Taxon names correspond to those used in the datasets. Note that, because of taxonomic changes or incorrect annotations in public databases, *Beloc_tr* is used for *Belonocnema kinseyi*, *Calli_sp* for *Neuroterus valhalla* and *Andr_qln* for *Druon quercuslanigerum*.

How to cite this article: Hearn, J., Gobbo, E., Nieves-Aldrey, J.L., Branca, A., Nicholls, J.A., Koutsovoulos, G. et al. (2023) Phylogenomic analysis of protein-coding genes resolves complex gall wasp relationships. *Systematic Entomology*, 1–28. Available from: <https://doi.org/10.1111/syen.12611>