Article (refereed)

Acs, Zoltán; Challis, Richard J.; Bihari, Péter; Blaxter, Mark; Hayward, Alexander; Melika, George; Csoka, Gyorgy; Penzes, Zsolt; Pujade-Villar, Juli; Nieves-Aldrey, Jose-Luis; **Schonrogge, Karsten**; Stone, Graham N.. 2010 Phylogeny and DNA barcoding of inquiline oak gallwasps (Hymenoptera: Cynipidae) of the Western Palaearctic. *Molecular Phylogenetics and Evolution*, 55 (1). 210-225. 10.1016/j.ympev.2009.12.004

Running head: PHYLOGENY AND DNA BARCODING OF CYNIPID INQUILINES

Phylogeny and DNA Barcoding of Inquiline Oak Gallwasps (Hymenoptera: Cynipidae) of the Western Palaearctic.

Zoltán Ács[1*], Richard Challis[2,3*], Péter Bihari[4], Mark Blaxter[2], Alexander Hayward[2,5], George Melika[1], Zsolt Pénzes[4,6], Juli Pujade-Villar[7], José-Luis Nieves-Aldrey[7], Karsten Schönrogge[8] and Graham N. Stone[2*]

*These authors contributed equally to this manuscript.

1. Systematic Parasitoid Laboratory, Plant Protection and Soil Conservation Service of County Vas, Ambrozy setany 2, Tanakajd 9762, HUNGARY; E-mail: acs.zoltan@gmail.com

2. Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, The King's Buildings, West Mains Road, Edinburgh EH9 3JT, UNITED KINGDOM.
E-mails: graham.stone@ed.ac.uk, mark.blaxter@ed.ac.uk

3. Department of Biology (Area 11), University of York, Heslington, York YO10 5YW. E-mail: rjc509@york.ac.uk

4. Biological Research Centre of Hungarian Academy of Sciences, Institute of Genetics, Temesvári krt 62, Szeged, 6723

5. Oxford University Department of Zoology, South Parks Road, Oxford, OX1 3PS.
E-mail: alexander.hayward@zoo.ox.ac.uk

6. University of Szeged, Department of Ecology, Szeged HUNGARY.
E-mail: penzes@bio.u-szeged.hu

7. Universitat de Barcelona, Facultat de Biologia, Departament de Biologia Animal, Avda. Diagonal 645, E- 08028 Barcelona, SPAIN.
E-mail: pujade@porthos.bio.ub.es

8. Dept. de Biodiversidad, Museo Nacional de Ciencias Naturales (CSIC), José Gutiérrez Abascal 2, 28006 Madrid, SPAIN.
E-mail: aldrey@mncn.csic.es

9. Centre for Ecology & Hydrology, CEH Wallingford, Maclean Building, Benson Lane, Wallingford, OX10 8BB, United Kingdom.
E-mail: ksc@ceh.ac.uk

Corresponding author: Dr Graham Stone, Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, The King's Buildings, West Mains Road, Edinburgh EH9 3JT, UNITED KINGDOM. Tel 44 131 6507194, Fax 0131 6506564, email graham.stone@ed.ac.uk

Short title: Phylogeny and DNA barcoding of cynipid inquilines.

Abstract

We examine phylogenetic relationships within the *Synergus* complex of herbivorous inquiline gallwasps associated with (Hymenoptera; Cynipidae; Synergini) associated with cynipid host galls on oak, a biologically diverse group whose genus-level morphological taxonomy has long been considered stable but whose species level taxonomy is problematic. We incorporate data for over 80% of recognised Western Palaearctic species in 5 morphology-based genera (*Ceroptres, Saphonecrus, Synergus, Synophrus, Ufo*), comprising sequence for two mitochondrial two mitochondrial loci (*coxI, cytb*) and one nuclear locus (28S D2). In particular, we assess the evidence for monophyly of two long-established, morphology-defined sections within the genus *Synergus* that differ in a range of biological traits between-generation polymorphism and impact on the host gall inducer (lethal *versus* non-lethal). To aid analyses of ecological interactions within oak cynipid communities, we also consider the utility of cytochrome oxidase I *(coxI)* DNA barcodes in the oak inquilines. In this assessment, we do not assume that species are delineated at a single threshold value of sequence divergence for a single gene, but examine concordance in the composition of molecular operational Taxonomic units (MOTUs) across a range of sequence divergences in each gene and across genes. We also assess the impact of sampling effort on MOTU stability.

Phylogenetic reconstructions for all three loci support monophyly for *Synergus* and *Synophrus,* but reject monophyly for *Saphonecrus* and for the two sections within *Synergus*. The suites of traits associated with the two sections of the genus *Synergus* are thus homoplasious. All three loci also reject monophyly for three *Synergus* species (*S. hayneanus, S. pallipes, S. umbraculus*). Sequences for each locus identify robust MOTUs that are largely concordant across loci for a range of cut-off values. Though many MOTU's correspond to recognised Linnean species, there is significant, multigene disagreement between groupings supported by morphology and sequence data, with both allocation of different morphospecies to the same MOTU and allocation of the same morphospecies to multiple MOTUs, regardless

of cutoff value. Our results imply that while DNA barcoding has considerable utility within this group, morphology-based identification needs major revision at both genus and species levels. Further, lifehistory traits currently attributed to single morphospecies probably confound attributes of multiple lineages. Revealing patterns of character state evolution in *Synergus* requires collection of new host association and life history data explicitly linked to DNA barcode data for the specimens concerned.

Key Words: Synergini; Cynipidae; inquiline; *Quercus*; oak; 28S rDNA; Cytochrome c oxidase I; Cytochrome b; DNA barcode;

1. Introduction

Accurate taxonomy and sample identification are crucial to analyses the ecology and evolution of species, higher-level taxonomic groups and communities. Morphological taxonomy has long been used to define species using consistent differences in external characters, but the scale of the challenge of identifying and classifying all species in this way is prohibitive (Tautz *et al*. 2003). While approximately 1.5 million, predominantly insect, species have been described to date (de Meeus & Renaud, 2002), these represent only a small proportion of estimated global diversity and molecular studies continue to reveal cryptic taxa inseparable on the basis of morphological taxonomy (e.g. Papakostas *et al.,* 2005; Smith *et al*., 2006; Bergmann & Russell, 2007; Smith *et al*., 2007; Starrett & Hedin, 2007), increasing the magnitude of the challenge.

Approaches based on genetic markers, particularly DNA sequence data, are increasingly used to augment or replace morphological taxonomic analyses (Tautz *et al*. 2003). As discussed by Vogler and Monaghan (2006), at least 3 conceptually different but related approaches have been used. DNA taxonomy uses patterns of variation in DNA sequence data to define taxa *a priori*, without reference to morphological data, though it can enable the identification of diagnostic morphological characters. This approach uses no pre-defined level of difference (e.g. % sequence divergence) to define taxa, but attempts to identify idependently evolving lineages. Because the topology of an individual gene tree can differ significantly from population and species trees, identification of such lineages is best approached using data for multiple loci (Meyer & Paulay 2005; Hickerson *et al*. 2006; Vogler & Monaghan 2006). A second approach captures the sequence diversity present in a group of samples by identifying molecular operational taxonomic units (or MOTUs; Floyd *et al.,* 2002), defined as a group of sequences differing from one another by a specified maximum number of base pairs (Blaxter 2004). MOTU richness is a useful summary measure of sequence diversity, particularly in describing samples of morphologically cryptic taxa.

However, MOTU approaches have the drawback in comparison with DNA taxonomy that the link between MOTU membership and biological species status remains unclear (Vogler & Monaghan 2006).

DNA barcoding is the third approach, and rather than defining taxa *a priori*, it uses sequence similarity at a single 'barcode' locus (in Metazoa, usually the mitochondrial locus cytochrome *c* oxidase subunit I, *coxI*) to allocate unknown specimens to morphologically determined voucher taxa (Floyd *et al*. 2002; Blaxter 2004; Hebert *et al*., 2004; Powers, 2004; Blaxter *et al*., 2005; Hajibabaei *et al*., 2005; Lambert *et al*., 2005; Ward *et al*., 2005). The underlying rationale for DNA barcoding is that sequence variation among species is greater than (and discrete from) variation within species: in other words, that sequence variation at the selected locus shows a 'barcoding gap' (Fig.1). Given this assumption, unidentified specimens that differ by less than a threshold sequence divergence (operationally taken as 2% for *coxI* in Metazoa; Hebert *et al*. 2003 and see http://www.barcodinglife.com) from voucher sequence for a reference species can be allocated to that species, and sequences more divergent than the selected threshold from any reference taxon will be allocated to a new species (Hebert *et al.,* 2003). This approach can be effective in revealing morphologically cryptic taxa, and so inform revision of morphological taxonomy (e.g. e.g. Smith et al., 2006; Challis et al., 2007; Smith et al., 2007; Stone et al., 2008).

However, there are major potential drawbacks of using a single mitochondrial barcode marker in this way. Coalescent theory predicts that both intraspecific sequence diversity and the probability of shared barcode sequences among discrete biological species will be sensitive to population demography, particularly species age and past effective population size (Moritz & Cicero 2004; Meyer & Paulay 2005; Hickerson *et al*. 2006; Nielsen & Matz 2006; Knowles & Carstens 2007). Young species with large effective population sizes are predicted not to be monophyletic for their barcode sequence, and where past population sizes vary substantially within a group of taxa, a single threshold divergence is unlikely to separate

intraspecific variation from varation among species – in other words, there may be no barcoding gap (Funk & Omland 2003; DeSalle *et al*., 2005; Meyer & Paulay, 2005; Cognato 2006; Vogler & Monaghan, 2006; Rubinoff, 2006). Heteroplasmy, introgression (with or without selective sweeps imposed by *Wolbachia* and other symbionts), potential selection on nucleotides, and the existence of nuclear pseudogenes can all lead to conflicts between species membership and grouping by DNA barcode (Hebert *et al*., 2004; Hurst & Jiggins, 2005). DNA barcoding is thus best used where (i) the presence of a barcoding gap for the barcoding locus can be demonstrated empirically, and (ii) concordance in gene tree topology can be demonstrated between the barcoding locus and at least one nuclear locus (Blaxter, 2004; Ahrens *et al*. 2007; Smith *et al*., 2007; Cardoso *et al*. 2009). More broadly, the utility of DNA barcoding is arguably greatest where existing morphological taxonomic expertise allows molecular and morphological approaches to be compared, revealing the shortfalls of traditional approaches but providing biological context to sequence-based taxa (e.g. Cardoso *et al*. 2009).

Here we examine the molecular taxonomy and potential for single locus DNA barcoding of

a group of insects whose morphological taxonomy is problematic – the inquiline gallwasps of the tribe Synergini (Hymenoptera: Cynipidae).

1.1 The inquiline gallwasps

The tribe Synergini contains *ca*. 170 species that inhabit plant galls induced by other insects – primarily other gallwasps (Cynipidae; reviewed in Csóka *et al*. 2005). The inquilines are highly specialised herbivores, able to modify the host plant tissues on which they feed, but dependent on true gall inducers to initiate gall formation (agastoparasites *sensu* Ronquist 1994; Brooks & Shorthouse, 1998; Sanver & Hawkins, 2000; Stone *et al.,* 2002b). They

attack the galls of hosts in the gallwasp tribes Cynipini (oak gallwasps, hosts to the inquiline genera *Ceroptres, Saphonecrus, Synergus, Synophrus* and *Ufo*), Diplolepidini (rose gallwasps, hosts to the inquiline genus *Periclistus*) and Aylacini (herb gallwasps, hosts to the inquiline genus *Synophromorpha*) (Ronquist, 1994, 1999; Csóka *et al*., 2005). The major exception to attack of cynipid hosts is the inquiline genus *Rhoophilus*, whose single species inhabits galls induced by Cecidosid moths on *Rhus* species (sumacs; van Noort *et al*., 2006).

Gall communities have been the subjects of numerous studies of community structure and evolution (Stone *et al.,* 2002; Stone & Schönrogge, 2003; Stone *et al*., 2009; Bailey *et al*., 2009), and cynipid inquilines have a major impact on food web structure and community species richness (Schönrogge *et al.,* 1995, 1996a, b; Schönrogge & Crawley, 2000; Stone *et al.,* 2002). The morphological taxonomy of the inquilines has been studied in depth (Mayr 1872; Wiebes-Rijks, 1979; Nieves-Aldrey & Pujade-Villar, 1985, 1986; Pujade-Villar & Nieves-Aldrey, 1990, 1993; Liljeblad & Ronquist, 1998; Nieves-Aldrey 2001; Pujade-Villar *et al*., 2003). On the basis of morphology, the most species-rich genus, *Synergus*, is divided into two sections (Sections I and II; Mayr 1872) that also differ in in a range of biological traits. Section I species are predominantly univoltine, and their development in a host gall is rarely lethal to the gall-inducer. In contrast, Section II species are predominantly bivoltine, and their development in a host gall often kills the gall inducer (Csóka *et al*., 2005). Section II species are also characterised by high between-generation variation in some adult morphological attributes, particularly size and colour (Nieves-Aldrey & Pujade-Villar, 1986; Pujade-Villar, 1992; Wiebes-Rijks, 1979), which can make morphology-based identification difficult. As a result, morphology-based identification of adults is often only possible to complexes of morphologically similar species, and identification of most immature stages, desirable in food web analyses, is currently impossible. Demonstration of effective DNA barcoding in this group would thus greatly facilitate the generation of more comprehensive food web data.

Here we assess the utility of two mitochondrial loci (*coxI* and cytochrome b, *cytb*) and one nuclear locus (the D2 region of the 28S ribosomal RNA gene) in resolving the relationships among inquiline oak gallwasp lineages, incorporating data for over 80% of recognised Western Palaearctic species and all recognised genera. We examine the concordance between mitochondrial and nuclear gene trees, particularly important here because two processes known to influence mitochondrial sequence diversity - introgression and infection with *Wolbachia* symbionts – occur in inquiline and gall inducing Cynipini (Rokas *et al*., 2001, 2002a, 2003a, b). We choose 28S D2 as the nuclear locus because it has proven valuable in resolving species level relationships in cynipids in the past, and in contrast to alternatives such as long wavelength opsin and elongation factor 1, it can be amplified with highly conserved primers (see Methods: problematic for long wavelength opsin in some cynipids: Stone *et al*., 2009) and does not require cloning (sometimes required for opsin and elongation factor 1; Rokas *et al*., 2002b; Stone *et al*., 2009). The need for further phylogenetic analysis of oak inquiline gallwasps is supported by the fact that although previous analyses support monophyly for the complex as a whole (Ronquist, 1994; Ronquist & Liljeblad, 2001; Nylander *et al*., 2004a, b; Nieves-Aldrey *et al*., 2005), the genera and species within the complex are difficult to distinguish morphologically (Pujade-Villar *et al*., 2003). Though some species in this complex have been included in previous molecular phylogenetic analyses (e.g. Rokas *et al*., 2002b; Nylander *et al*., 2004a, b; Pénzes *et al*., 2009), sampling of *Synergus* has been very limited. A recent molecular analysis of the genera *Synophrus* and *Saphonecrus* (Pénzes *et al.*, 2009) revealed the possible diphyly of the genus *Saphonecrus*, though placement of these genera relative to *Synergus* was not investigated. We address the following questions: (a) Are the genera *Synophrus, Saphonecrus* and *Synergus* monophyletic groups, and what are the phylogenetic relationships between them? (b) are Sections I and II of *Synergus* monophyletic groups? If yes, then diagnostic biological traits have been conserved within two divergent lineages. Rejection of section monophyly would

imply that biological trait evolution in *Synergus* is more labile than currently accepted. (c) Is there evidence of morphologically cryptic sequence diversity, suggesting the need for revision of current estimates of taxon diversity?

We then explore the potential for DNA barcoding in the oak-associated Synergini. First, we assess the empirical support for a barcoding gap in *coxI* and *cytb*. Rather than assuming a single cut-off threshold for definition of MOTUs (e.g. 2% sequence divergence for *coxI*), for the two mitochondrial genes we examine the impact of variation in the cutoff threshold from 0-12% on MOTU membership. Because MOTU richness and composition can be sensitive to taxon sampling (Meyer & Paulay, 2005), we also explore the impact of sampled sequence diversity on MOTU richness. We address the following additional questions: (d) Do *coxI* MOTUs capture Linnean species? If not, are separate Linnean species represented by sequences in the same MOTU (Type I error *sensu* Quicke, 2004)? (e) Do *coxI* barcodes reveal cryptic taxa unrecognised by existing morphology-based taxonomy (Type II error *sensu* Quicke, 2004)? (f) To what extent are MOTUs identified using *coxI* sequence data concordant with those supported by *cytb* and 28S D2? (g) To what extent is MOTU membership sensitive to the number of sequences included in the analysis?

2. Materials and methods

2.1 Taxon sampling

Because tests of the efficacy of DNA barcoding are highly dependent on thorough taxon sampling (Meyer & Paulay 2005), we sampled as many of the described species of Western Palaearctic Synergini associated with oak cynipid hosts as possible. We obtained data for 184 specimens (see supplementary on-line Appendix S1) comprising 33 of the 41 described species (>80%: 2/2 *Ceroptres,* 6/6 *Saphonecrus*, 23/30 *Synergus*, 2/3 *Synophrus*; Csóka *et al*.,

2005; Sadeghi *et al*., 2006). Because the ability of DNA barcodes to assign individuals to species requires adequate sampling of intraspecific as well as interspecific variation (Moritz & Cicero, 2004; Morando *et al*., 2003), where possible we incorporated samples from multiple refugial centres of intraspecific diversity (Rokas *et al*., 2003b) across the Western Palaearctic from Spain to Iran (Appendix S1). Sampling was strongest for *Synergus*, in which 11 species were represented by four or more *coxI* sequences, and four species (*S. hayneanus, S. pallicornis, S. pallipes* and *S. umbraculus*) were represented by between 10 and 15 sequences (Appendix S1). Examination of conflicts between Linnean species designation and *coxI* MOTU membership is thus restricted to *Synergus.*

Unsampled species are either extremely local in distribution, such that we were unable to obtain samples (*Synophrus olivieri, Synergus ibericus, S. ilicinus, S. ruficornis, S. subterraneus* Pujade-Villar *et al*. 2003) or are morphologically very similar to (and probable synonyms of) sampled species (*S. dacianus* of the sampled species *S. crassicornis*; *S. synophri* of the sampled species *S. hayneanus*; and *S. radiatus* of the sampled *S. pallipes*). All sampled inquilines were reared from oak galls, except for *Rhoophilus loewi*, which was reared from *Scyrotis* sp. (Lepidoptera, Cecidosidae) galls induced on *Rhus* in South Africa. Specimens were identified by Melika and Pujade-Villar, recognised experts in the field of morphological cynipid taxonomy and the authors of the existing keys (Pujade-Villar *et al*., 2003). Due to difficulties in morphological identification, some individual specimens were assigned >1 morphospecies name (see Appendix S1). The Eastern Palaearctic (Abe *et al*., 2007) and Nearctic (Burks, 1979) inquiline faunas are far less known. To begin the process of assessing relationships between these regional faunas, and in particular to explore whether Western Palaearctic MOTUs span the palaearctic into Asia, our analysis incorporates two *Synergus* species from China (*Synergus chinensis* and *S. xiaolongmeni*; det. G. Melika) (Melika *et al.* 2004), and one from Japan (*Synergus japonicus*; det. Y. Abe).

2.2 DNA Extraction and Sequencing

DNA was extracted from a single hind leg of most specimens using a simple Chelex-based protocol (Lopez-Vaamonde et al. 2001). Based on prior experience (though not a hard and fast rule), for insects less than 2 mm long we extracted DNA from the whole wasp using the DNeasy Tissue Kit (QIAGEN cat. 69504).

For cytochrome b (*cytb*), we first used the following primers: forward primer, CB1 TATGTACTACCATGAGGACAAATATC, reverse primer, CB2 ATTACACCTCCTAATTTATTAGGAAT (Jermiin & Crozier, 1994). Where amplification using the CB1/CB2 primer pair failed, an overlapping fragment was amplified using the CP1/CP2 primer pair: CP1 GATGATGAAATTGGATC, CP2 CTAATGCAATAACTCCTCC (Harry *et al.,* 1998). For cytochrome c oxidase subunit I (*coxI*), we used forward primer, LCO1490 GGTCAACAAATCATAAAGATATTGG, and reverse primer HCO2198 TAAACTTCAGGGTGACCAAAAAAT (Folmer *et al.,* 1994). For the D2 region of the nuclear 28S ribosomal RNA gene, we used forward primer CGTGTTGCTTGATAGTGCAGC, and reverse primer TCAAGACGGGTCCTGAAAGT (Heraty *et al*. 2004). We also amplified the D3-5 region of the 28S ribosomal RNA gene, using forward primer ACACACTCCTTAGCGGA, and reverse primer GACCCGTCTTGAAACACGGA (Friedrich & Tautz, 1995).

For all loci, 25 µl polymerase chain reactions (PCRs) were carried out in a PTC-200 DNA Engine (MJ Research) using 1 U Taq polymerase (Invitrogen or Promega), 2.5 µl 10x Taq buffer, 1.5 µl $MgCl_2$ (25 mM), 0.5 µl dNTPs (10 mM), 0.35 µl primers (20 pmol), 1.0 µl template DNA and 18.85 µl dH_2O. PCR products were purified using shrimp alkaline phosphatase and *E. coli* exonuclease I (USB Corporation, USA) and sequenced directly on an automated ABI Prism 3730 Genetic Analyzer machine using ABI BigDye v3.1 Terminator Sequencing chemistry. All PCR products were sequenced in both directions to minimise PCR artefacts, ambiguities and base-calling errors. Chromatogram output was checked by eye

using Sequencher 4.1 (Gene Codes) or ProSeq (Filatov, 2002). Direct sequencing of a small proportion of the *cytb* PCR products revealed mixtures of multiple *cytb*-like fragments, or sequences possessing reading frames containing stop codons or indels, suggesting the possible presence of nuclear pseudogenes (Bensasson *et al*. 2001, Rokas *et al*., 2003a). In these cases, individual PCR products were amplified by cloning (TA cloning, Invitrogen) and only specimens for which a single, correct open reading frame (ORF) bearing sequence was identified have been included in the following analyses. Though heterozygotes have been detected for the 28S D2 region in cynipids (e.g. Stone *et al*., 2007) and other taxa (e.g. Smith *et al.* 2008), none were detected in our surveys.

Our analysis incorporates 404 new sequences (*coxI,* 106; *cytb,* 143; 28S D2, 108; 28S D3-5, 47), with Genbank accession numbers in Appendix S1. Of the 184 Synergini specimens in the study, 70 had full sequences for all three genes (*cytb*, *coxI* and 28S D2). The 67 discrete haplotype sets in these specimens comprise the **maxtaxa** dataset.

2.3 Phylogenetic Analyses

2.3.1 Alignment and phylogeny reconstruction

All *coxI* (660 bp) and *cytb* (433 bp) sequences were the same length, and each gene set could be aligned unambiguously by eye. The 28S fragments were of variable length (D2: 520-572 bp, D3-5: 511-513 bp) and were aligned using MUSCLE 3.6 (Edgar 2004) using default settings. There was very little sequence variation in the D3-5 region, and we do not consider it further. Bayesian phylogenetic inference was performed in MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003). Phylogenies for individual genes and for combined gene datasets were constructed under the GTR+I+G model of sequence evolution, partitioned by codon position for *cox1* and *cytb*. For each data matrix, two independent Markov chain Monte Carlo (MCMC) runs of four Metropolis-coupled chains were performed with the gamma shape

parameter, the proportion of invariant sites, base frequencies and substitution rates unlinked across all partitions, and with default priors. MCMC runs comprised either 4 million (*cox1,* 28S D2) or 8 million (*cytb*) generations, sampled every 1000 generations with a burn-in time of 3 or 7 million generations, respectively. Models were considered to have converged when the average standard deviation of split frequencies between the two independent runs fell below 0.01, and chain parameters examined in Tracer v1.4 (Rambaut & Drummond 2007) showed stable distributions. All trees were rooted using *Ceroptres clavicornis*, since previous studies spanning multiple tribes of Cynipinae (Nylander et al. 2004a) have shown this genus to represent a lineage distinct from the *Synergus* complex.

Because molecular clock assumptions can significantly influence clade support, we compared support for clock and non-clock models for each gene in MrBayes using ln Bayes factors (ln BF). These were estimated as twice the difference in the natural log of the harmonic mean of model likelihoods of each model (2ΔlnHML), interpreted following Table 2 of Kass & Raftery (1995). By their criteria, ln BF of 2-6, 6-10 and >10 represent respectively positive, strong and very strong support for the model with higher likelihood. All three molecular markers showed sequence variation consistent with strict clock assumptions (ln BF in favour of a clock model for *cox1*=32; *cytb* =227, 28S D2=221).

2.3.2 Tests of taxon monophyly

Support for the monophyly of specific taxa was tested by using MrBayes and Bayes factors as above to compare the harmonic mean likelihoods of models in which taxon monophyly was constrained with models in which there was no such constraint.

2.3.3 Likelihood mapping

Likelihood mapping (Strimmer & von Haeseler, 1997) provides an estimate of the phylogenetic utility of a set of sequences, and was performed in TreePuzzle 5.0 (Schmidt et

al., 2002) using the HKY model of nucleotide substitution. Likelihood maps were constructed for each gene with parameters estimated from the data set and using all possible quartets of taxa.

2.4 CoxI MOTU analyses

MOTUs were defined for the complete set of unique *coxI* haplotypes at cut-off values ranging from 1-100 base pairs (*ca*. 0-12% sequence divergence) using MOTU_define 2.04 (Floyd & Blaxter, 2006). MOTU_define clusters input sequences into MOTUs by adding each sequence in turn to a local BLAST database and then taking the next sequence and performing a BLAST similarity search against the entries in the local database. If the sequence has less than the user-defined cut-off number of differences to an existing MOTU then it is added to that MOTU, otherwise it is assigned to a new MOTU. The grouping of sequences in this way is sensitive to the order in which they are added (Blaxter *et al.,* 2005) so 100 replicates using different random resampling orders were performed for each MOTU cut-off.

To examine the impact of sampling depth on MOTU richness and membership, we compared the results obtained from analysis of the **maxdata** supermatrix (matrix containing data for all 3 loci) with those obtained for a second supermatrix intended to maximise sequence diversity present in all three loci, while minimising the number of specimens with missing data. Starting from the **maxdata** supermatrix, this **maxtaxa** supermatrix was constructed by adding all specimens that either individually contributed a unique new haplotype for at least one gene, or represented a novel combination of existing haplotypes. Where two individuals shared a haplotype not present in the **maxdata** supermatrix, we added data for specimens sequenced for two genes in preference to those sequenced for one. This approach resulted in addition of data for 58 specimens, resulting in a supermatrix containing sequence for 125 specimens, almost doubling the specimen sample size. The **maxtaxa** matrix lacked data for 100 of the 375 gene sequences making up the alignment, representing 26.7%

missing data.

3. Results

3.1 Phylogenetic Relationships within the Synergus complex

3.1.1 Phylogenetic utility of coxI, cytb and 28S D2

All three molecular markers used in phylogeny reconstruction had high phylogenetic utility, all showing fewer than 10% of quartets in the unresolved central portion of the likelihood map (Figure 2). *CoxI* had the highest phylogenetic utility, with 95.8% of quartets in the well-resolved regions towards the corners, followed by *cytb* (92.9%) and 28S D2 (86.4%).

3.1.2 Relationships among major lineages of the Synergus complex

All three genes supported broadly concordant relationships among major lineages (Figs. 3-5 for *coxI, cytb* and 28S D2 respectively). All three genes supported monophyly of the large genus *Synergus*. Bayes factor comparisons strongly supported monophyly in the *coxI* data (Fig.3), while the *cytb* and 28S D2 data were equivocal (Table 1). *Synergus* was nevertheless supported as monophyletic with a posterior probability of >0.95 in the 28S D2 reconstruction (Fig.5). All three genes supported monophyly for the small genus *Synophrus* (the posterior probability of monophyly is 1.0 in all analyses; Figs. 3-5), but while *Saphonecrus* monophyly was supported (albeit weakly) by the *cytb* data (Fig.4), monophyly was rejected rather more strongly by *coxI* and 28S D2 (Table 1). Both *coxI* and 28S D2 divide *Saphonecrus* into two (Figs. 3,5): three species (*Saphonecrus barbotini, S. connatus* and *S. lusitanicus*) form part of a monophyletic clade including all sampled *Synophrus* species, while the remaining Western Palaearctic species (*Saphonecrus haimi, S. undulatus* and a recently described species *S. irani*) represent a separate lineage (the ‘haimi clade’ Figs. 3, 5). Which

of these two lineages is the sister group to *Synergus* remains poorly resolved in each of the single gene datasets, while the haimi clade of *Saphonecrus* is strongly supported (posterior probability >0.95) as the sister group in an analysis incorporating data for all three genes (Fig.S1).

395

3.1.3 Synergus *shows extensive non-monophyly of morphology-based taxa*

All three genes very strongly rejected monophyly of Hartig's sections I and II within *Synergus* (Table 1, Figs. 3-5). This implies that the biological traits characteristic of species in these sections are evolutionarily labile within this genus. There are also conflicts in *Synergus* 400 between morphological and molecular taxa at the level of Linnean species. Bayes factor comparisons for all three genes rejected monophyly of hapotypes attributed to *S. hayneanus, S. pallipes* and *S. umbraculus* (highlighted in Figs. 3-5, Table 1). Phylogenetically divergent haplotypes in all 3 genes were also attributed to *Synergus flavipes* (Figs. 3-5).

405 *3.1.4 Placement of Eastern Palaearctic taxa*

All single gene datasets show Chinese *Synergus chinensis* and *S. xiaolongmeni* and Japanese *Synergus japonicus* to represent distinct lineages nested among the Western Palaearctic lineages. There is no evidence to suggest that Eastern and Western Palaearctic *Synergus* represent discrete monophyletic radiations.

410 *3.2 MOTU Analysis*

3.2.1 CoxI and cytb both show evidence of a barcoding gap

The relationship between numbers of MOTUs and the percentage sequence divergence used to define them is shown for all three genes in Figure 6. Both *coxI* and *cytb* show a plateau of MOTU richness over a range of cut-off values compatible with the presence of a

'barcoding gap'. Because there is no formal way to define the margins of a barcoding gap, we examined MOTU composition across a range of cutoff values spanning the barcoding gap for each gene (indicated on Fig.6). For *coxI*, the selected sequence divergence values were 1.1% (most divisive, 'splitter'), 3.8% (midpoint) and 6.4% (most inclusive, 'lumper'), equivalent to divergences of 7, 25 and 42 base pairs (bp) and labelled A-C on Fig.6. The equivalent values for *cytb* were 2.1, 4.4 and 6.2% (equivalent to 9, 19 and 27 bp, points D-F, Fig.6). There is little evidence of a barcoding gap in the 28S D2 sequences, and to allow comparison of MOTU membership across the 3 genes, we selected a cut-off value of 2 bp (equivalent to 0.04 % sequence divergence, point G, Fig.6).

3.2.2 cox1 MOTUs and discordance with morphological taxonomy

The 98 *coxI* sequences assigned 31 *Synergus* complex species to 27 MOTUs at the inclusive 6.4% cut-off (Appendix 1, Fig.S1), comprising 13 clusters and 14 singleton haplotypes (six of which were found in more than one individual). Comparison of morphology-based species identification with MOTU designations at this cut-off revealed that though many MOTU's corresponded to recognised Linnean species, there was also substantial discordance: six MOTUs (Table 2) contained samples representing more than one morphospecies (Type I error), while six species are represented by samples in more than one MOTU (Type II error) (Table 3).

The divisive 1.1% cut-off assigned the same haplotypes to 40 MOTUs, comprising 20 clusters and 20 singleton haplotypes (eight of which were found in more than one sample). At this cut-off, only 4 MOTUs contained more than 1 species (Table 2), while 8 species were present in more than one MOTU (Table 3). Significantly for our inference of tupe 1 and type 2 errors, the specimens contributing to conflicts between morphological taxonomy and MOTU allocations were the same regardless of cut-off level, and included the non-monophyletic species described in section 3.1.3. The morphospecies most commonly grouped

together in a single MOTU were (i) *Synergus gallaepomiformis, S. pallicornis* and *S. pallipes* (all in Mayr Section II), grouped with *S. pallidipennis* (Section I) in MOTU 20, and (ii) *Synergus flavipes, S. hayneanus* and *S. umbraculus* (all in Mayr Section I) in MOTU 19 (Table 2, text Appendix 1). The species most frequently allocated to multiple MOTUs were *Synergus hayneanus* (Section I; 4 MOTUs at 42bp, 5 MOTUs at 7bp), *S. pallipes* (Section II; 3 and 4 MOTUs, respectively) and *S. umbraculus* (Section I; 2 and 5 MOTUs, respectively) (Table 3). Although Section II of *Synergus* is charactersised by greater intraspecific morphological diversity (see 1.1, above), there is no evidence that either Type I or Type II errors are more common in this section.

3.2.3 MOTU concordance across sampled genes

The three genes in our analysis identify highly concordant sets of MOTUs in the **maxdata** alignment (Appendix 2). At the inclusive cut-off (6.4% for *coxI*, 6.2% for *cytb*), 8/14 clusters and all singletons inferred with the 67 distinct *coxI* sequences were also inferred with *cytb* (Appendix 2). MOTUs defined by *coxI* were more stable over a range of cut-off values than those defined by *cytb*, apparent in Fig.6 as the flatter plateau between points A-C for *cox1* than between points D-F for *cytb*.

Comparison between *coxI* (6.4%) and 28S D2 (2 bp, 0.04 %) gave the strongest agreement of all between-gene comparisons. The 28S D2 data support 10/14 of the *coxI* clusters and all of the *coxI* singletons (Appendix 2). This agreement in MOTU definition between genes supports the conclusion that the disagreement revealed between morphology- and sequence-based taxa is genuine, and unlikely to be an artefact associated with the use of a particular sequence to define molecular taxa.

3.2.4 MOTU designation is robust to variation in taxon sampling and missing data

All MOTUs inferred with greater than 50% support for each of the genes using the **maxdata** set (67 specimens) were also inferred using the **maxtaxa** set (125 specimens; see Appendices 1 and 2). Addition of the 58 extra **maxtaxa** sequences had little impact on levels of Type I (the number of MOTUs containing >1 species) and Type II (the number of species in >1 MOTU) errors; each increased by 1 relative to values at corresponding cut-offs for the **maxdata** matrix, and involved almost identical sets of Linnean species (one additional MOTU containing two Linnean species is inferred in the **maxtaxa** dataset; Tables 2,3).

4. Discussion

4.1 Phylogenetic relationships within the Synergus complex

Our analyses provide well-resolved phylogenetic hypotheses for relationships between major lineages within the *Synergus* complex, but reveal widespread conflict with existing groupings based on morphology. While the genera *Synergus* and *Synophrus* are probably monophyletic, *Saphonecrus* is probably diphyletic. Our results thus confirm earlier doubts over the monophyly of this genus (Pujade-Villar & Nieves-Aldrey, 1990; Pénzes *et al*., 2009). The three *Saphonecrus* species allied to *Synophrus* (*S. barbotini, S. connatus* and *S. lusitanicus*) could reasonably be transferred to the genus *Synophrus*, while the distinct lineage comprising *Saphonecrus haimi* and *S. undulatus* could retain the genus name *Saphonecrus*. Within *Synergus*, we find no separation between Western Palaearctic taxa and our very limited sampling of three Eastern Palaearctic species from China and Japan. We find that Mayr's long-accepted morphology-based sections within the genus *Synergus* do not represent natural groups, and should be abandoned. The character used by Mayr to separate the two sections of *Synergus* - the dorso-ventral distribution of sculpturing on tergites of the metasoma (the abdomen behind the petiole) – is clearly relatively labile evolutionarily, and of no taxonomic use.

4.2 The potential utility of coxI barcodes for the Synergus complex

The widely used Folmer barcode region of the mitochondrial *coxI* gene appears to show a clear barcoding gap in the Synergus complex. MOTUs defined with this *coxI* region are relatively stable across a range of cut-off values representing 1.1-6.4% sequence divergence, and in many cases MOTU membership reflects current membership of Linnean species. The resampling inherent in MOTU_define and comparison of results for the **maxdata** and **maxtaxa** analyses also shows that the compositions of the *coxI* MOTUs are also relatively insensitive to sampling order and variation in haplotype sampling effort. There has rightly been widespread criticism of the assumption of a single sequence divergence threshold in a single locus in discriminating between biological species (see 4.3). We suggest that the stability of *coxI* MOTU membership and the congruence in MOTU membership across mitochondrial and nuclear genes supports the use of *coxI* barcodes in the Synergus complex.

4.3 Mismatches between morphological taxonomy and MOTU membership in Synergus

While some recognised *Synergus* species correspond to *coxI* MOTUs over the full range of cut-offs investigated here, others clearly do not. MOTU-based groupings suggest that both Type I errors (separation into discrete taxa where none exists) and Type II errors (cryptic lineages within single morphological species) (Quicke, 2004) exist in the *Synergus* complex. There are two general hypotheses for such mismatches between morphological taxonomy and MOTU membership: (i) that MOTU-based identification is correct, while identification based on morphological traits is flawed, and (ii) that MOTU-based identification is flawed, while identification based on morphological traits is correct. The latter hypothesis predicts a mismatch between specimen groupings based on mitochondrial sequence data and those based on nuclear sequence data. There are many reasons to expect such a mismatch (Hudson

& Turelli, 2003; Machado and Hey, 2003; Moritz & Cicero, 2004; Hurst & Jiggins, 2005). In particular, coalescent theory predicts that species can share *coxI* barcodes through incomplete sorting of ancestral polymorphism (Hickerson *et al*., 2006; Knowles & Carstens, 2007), and many species are polyphyletic rather than monophyletic for mitochondrial genes (Johnson & Cicero 2002; Funk & Omland, 2003; Meyer & Paulay, 2005). Where sorting of ancestral polymorphism is complete, species can still share barcodes through introgression (Machado and Hey, 2003; Hurst & Jiggins 2005). Exchange of mitochondrial genes by introgression has been demonstrated in gall inducing cynipids (Rokas *et al*. 2003b), and might thus be expected in inquiline gallwasps. These pitfalls of relying on a single mitochondrial locus in molecular barcoding are the same as those associated with its use in phylogeography, and can be avoided by using multiple unlinked nuclear loci to determine affinities between specimens (e.g. Knowles & Carstens, 2007; Starrett & Hedin, 2007). Although we have only used one nuclear locus here, the agreement in MOTU memberships between *coxI* and 28S D2 datasets despite the much lower levels of sequence variation in the nuclear locus (see also Rokas *et al*., 2002b; Blaxter, 2004; Ahrens *et al*. 2007) gives greater confidence that the MOTUs capture biologically meaningful entities.

If the molecular taxa identified here do represent discrete biological entities, then the morphological taxonomy of the Synergini must be flawed – either because the characters used do not adequately define taxa, or because the taxonomists identifying the specimens made mistakes in character recognition. The authors have considerable accumulated experience of working with inquiline cynipids (Nieves-Aldrey & Pujade-Villar, 1985, 1986; Pujade-Villar & Nieves-Aldrey 1990; Pujade-Villar 1992; Pujade-Villar *et al*., 2003; Sadeghi *et al*., 2006), so mistaken interpretation of specimen morphology should be rare. While it is possible that occasional placements of specimens in phylogenetically disparate MOTUs could be the result of identification error (e.g. the placement of *Synergus flavipes* Figs. 3-5), we regard taxonomist error as an unlikely explanation for two major persistent mismatches between

morphological and sequence-based groupings: (i) The inability of sequence data to discriminate between *Synergus gallaepomiformis, S. pallicornis, S. pallidipennis* and *S. pallipes*; and (ii) the widespread phylogenetic placement and the multiple MOTU allocation of specimens in the morphospecies *Synergus hayneanus* and *S. umbraculus* (Figs. 3-5, S1).

4.3.1 Failure of barcodes to discriminate among recognised morphological species

Several sets of *Synergus* species shared identical or very similar *coxI, cytb* and 28S D2 sequences (Figs. 3-5, S1). Examples include *Synergus gallaepomiformis, S. pallicornis, S. pallidipennis* and *S. pallipes* combined in *coxI* MOTU 20, *Synergus acsi, S. flavipes* and *S. variabilis* combined in *coxI* MOTU 24, and *Synergus hayneanus* and *S. umbraculus* (Fig. S1). Failure to discriminate these species using sequence barcodes implies either (i) that these morphospecies genuinely grade into each other (such that neither existing morphological characters nor barcode sequence can meaningfully discriminate among them) or (ii) that the morphological traits discriminating the species are real, but that the species have diverged so recently that sorting of both mitochondrial (*coxI*) and nuclear (28S D2) lineages between them is far from complete (Johnson & Cicero 2002; Funk & Omland, 2003; Meyer & Paulay, 2005; Hickerson *et al*., 2006). We suspect that the first explanation applies. *Synergus* species are known to show substantial phenotypic variation within and between generations each year (Wiebes-Rijks, 1979; Pujade-Villar, 1992; Nieves-Aldrey, 1986), and the characters currently used to discriminate species may simply represent redescriptions of phenotypic plasticity in a single taxon. We recommend that the morphological basis of the groups of species above should be thoroughly reviewed. If reanalysis supports their maintenance as separate taxa, then species relationships should be reinvestigated using multiple nuclear locus approaches that provide greater statistical power when lineage sorting is likely to be incomplete (Jenning and Edwards 2005; Knowles & Carstens, 2007; Starrett & Hedin, 2007).

4.3.2 Placement of morphological species in phylogenetically diverse MOTUs.

Synergus hayneanus and *S. umbraculus* are both placed in several phylogenetically divergent lineages for both mitochondrial and nuclear genes (Figs. 3-5, S1). This pattern was reconfirmed when the morphology of a subset of specimens of both species was rechecked without knowledge of their phylogenetic placement. Our results suggest the need for careful revision of the morphological traits associated with separation of *S. hayneanus* and *S. umbraculus* (and *S. reinhardi*, which was often hard to separate morphologically from *S. hayneanus*). Further, the characters used to identify *S. hayneanus* and *S. umbraculus* are either homoplasious or conserved ancestral traits, and conceal phylogenetically divergent but morphologically cryptic lineages. While such cryptic lineages have been observed in other arthropod groups (e.g. Hebert *et al*., 2004; Smith *et al.*, 2006, 2007; Starrett & Hedin, 2007), the *Synergus* complex is unusual in that its member species have a long history of taxonomic and ecological study (Ross, 1951; Eady, 1952; Wiebes-Rijks, 1979; Nieves-Aldrey & Pujade-Villar, 1985, 1986; Pujade-Villar & Nieves-Aldrey 1990; Pujade-Villar 1992; Ronquist 1994; Schönrogge *et al*., 1995, 1996a, 1996b; Liljeblad & Ronquist, 1998; Schönrogge & Crawley 2000; Pujade-Villar *et al*., 2003).

4.4 Consequences of the need to revise oak inquiline gallwasp taxonomy

A striking feature of the original *Synergus* sections established by Mayr (1872) is the correlation between section membership and biological differences. Most section I *Synergus* species are univoltine, and where there are two generations per year (in *S. crassicornis* and *S. umbraculus* in the Iberian Peninsula), adult morphology does not differ markedly between generations (Nieves-Aldrey & Pujade-Villar, 1985; Pujade-Villar, 1992). In contrast, most section II *Synergus* species are bivoltine and show generational adult dimorphism (Ross, 1951; Eady, 1952; Wiebes-Rijks, 1979; Nieves-Aldrey & Pujade-Villar, 1986; Pujade-Villar, 1992). The distribution of section membership through the *Synergus* tree (Figs. 3-5, S1)

implies that these character state combinations have evolved repeatedly in the genus. It is then of interest to examine how strictly correlated evolution of these traits has been through diversification of Synergus. However, such an analysis may be impossible using existing published data, because for several Linnean species specimens attributed to a single morphological species are placed in multiple distinct phylogenetic lineages (discussed in 4.3 below). Unless DNA sequence data can be generated for the specimens examined in past work, it will be impossible to know which of alternative possible lineages should receive the character states attributed to current Linnean species (see also Knowlton and Jackson, 1994). Revealing patterns of character state evolution in *Synergus* requires collection of new host association and life history data explicitly linked to accessions from which DNA sequence data can be generated and lineage membership determined.

Acknowledgements

This work was funded by grants from the U.K. NERC to GNS and KS (NE/E014453/1, NE/D007178/1, NER/B/504406/1, NER/B/S2003/00856, GR/12847, GST032035), by a Hungarian OTKA grant T049183 and a Bolyai Scholarship ZP, and by a British–Hungarian Intergovernmental collaborative research grant (2002–04) to GNS and GM.

References

Abe, Y., Melika, G., Stone, G.N., 2007. The diversity and phylogeography of cynipid gallwasps (Hymenoptera, Cynipidae) of the Eastern Palaearctic and their associated communities. Oriental Insects 41, 169-212.

Ahrens, D., Monaghan, M.T., Vogler, A.P., 2007. DNA-based taxonomy for associating adults

and larvae in multi-species assemblages of chafers (Coleoptera : Scarabaeidae). Mol. Phyl. Evol. 44 436-449.

Bailey, R., Schönrogge, K., Cook, J.M., Melika, G., Csóka, G., Thúroczy, C., & Stone, G.N., 2009. Host niches and defensive extended phenotypes structure parasitoid wasp communities. PLoS Biology, 7(8): e1000179.

Bensasson, D., Zhang, D.-X., Hartl D.L., Hewitt, G.M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16, 314-321.

Bergmann, P.J., Russell, A.P., 2007. Systematics and biogeography of the widespread Neotropical gekkonid genus *Thecadactylus* (Squamata), with the description of a new cryptic species. Zool. J. Linn. Soc. 149, 339-370.

Blaxter, M. 2004. The promise of a DNA taxonomy. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 359, 669–679.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., Abebe, E., 2005. Defining operational taxonomic units using DNA barcode data. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 360, 1935-1943.

Brooks, S.E., Shorthouse, J.D., 1998. Developmental morphology of stem galls of *Diplolepis nodulosa* (Hymenoptera: Cynipidae) and those modified by the inquiline *Periclistus pirata* (Hymenoptera: Cynipidae) on *Rosa blanda* (Rosaceae). Can. J. Bot. 76, 365-381.

Burks, B.D., 1979. Cynipidae. In: Krombein, K.V., Hurd, P.D., Jr., Smith, D.R., Burks, B.D. (Eds.), Catalog of Hymenoptera in America North of Mexico. Vol. I. Smithsonian Institution Press, Washington, D.C., USA, pp. 1060-1107.

Cardoso, A., Serrano, A., Vogler, A.P., 2009. Morphological and molecular variation in tiger beetles of the Cicindela hybrida complex: is an 'integrative taxonomy' possible? Mol. Ecol.

18, 648-664.

Challis, R.J., Stone, G.N., Mutun, S., J.-L., Nieves-Aldrey, Preuss, S., Rokas, A., Aebi, A., Sadeghi, E., Tavakoli, M., 2007. Longitudinal range expansion and cryptic eastern species in the Western Palaearctic oak gallwasp *Andricus coriarius*. Mol. Ecol. 16, 2103-2114.

Cognato, A. 2006. Standard percent DNA sequence difference for insects does not predict species boundaries. J. Econ. Entomol. 99, 1037-1045.

Csóka, Gy., Stone, G.N., Melika, G., 2005. Biology, ecology, and evolution of gall-inducing Cynipidae. In: Raman, A., Schaefer, C.W., Withers, T.M. (Eds.), Biology, Ecology, and Evolution of Gall-inducing Arthropods. Science Publishers, Inc., New Hampshire, USA, pp. 573-642.

de Meeus, T., Renaud, F. 2002. Parasites within the new phylogeny of eukaryotes. Trends Parasitol. 18, 247-251.

De Salle, R., Egan, M.G., Siddall, M., 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 360, 1905-1916.

Eady, R.D. 1952. A revision of section I (Mayr, 1872) of the genus *Synergus* (Hym., Cynipidae) in Britain, with a species new to science. Trans. Soc. Brit. Entomol. 11, 141-152.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797.

Filatov, D.A. 2002. ProSeq: a software for preparation and evolutionary analysis of DNA datasets. Mol. Ecol. Notes 2, 621-624.

Floyd, R. Blaxter, M.L., 2006. MOTU_define.pl, available from

http//www.nematodes.org/bioinformatics/MOTU/index.shtml

Floyd, R., Eyualem, A., Papert, A., Blaxter, M.L. 2002. Molecular barcodes for soil nematode identification. Mol. Ecol. 11, 839–850.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol. Marine Biol. and Biotechnol. 3, 294-299.

Friedrich, M. Tautz, D., 1995. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution myriapods. Nature 637, 165-167.

Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Syst. 34, 397–423.

Hajibabaei, M., De Waard, J.R., Ivanova, N.V., Ratnasingham, S., Dooh, R.T., Kirk, S.L., Mackie, P.M., Hebert, P.D.N., 2005. Critical factors for assembling a high volume of DNA barcodes. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 360, 1959-1967.

Heraty, J., Hawks, D., Kostecki, J.S., Carmichael, A., 2004. Phylogeny and behaviour of the Gollumiellinae, a new subfamily of the ant-parasitic Eucharitidae (Hymenoptera: Chalcidoidea). Syst. Entomol. 29, 544-559.

Harry, M., Solignac, M., Lachaise, D., 1998. Molecular evidence for parallel evolution of adaptive syndromes in fig-breeding *Lisocephala* (Drosophilidae). Mol. Phylogen. Evol. 9, 542-541.

Hebert, P.D.N., Cywinska, A., Ball, S.L., de Waard, J.R., 2003. Biological identifications through DNA bar-codes. Proc. Roy. Soc. Lond. Ser. B 270, 313-322.

690 Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Nat. Acad. Sci. U.S.A. 101, 14812–14817.

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of birds through DNA barcodes. PLoS Biology 2, 1657-1663.

695 Hickerson, M.J., Meyer, C., Moritz, C. 2006. DNA-barcoding will fail to discover new animal species over broad parameter space. Syst. Biol. 55,729-739.

Hudson, R.R., Turelli, M., 2003. Stochasticity overrules the "threetimes rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. Evolution 57,182–190.

700 Hurst, G.D.D., Jiggins, F.M., 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc. Roy. Soc. Lond. Ser. B 272, 1525-1534.

Jermiin, L.S., Crozier, R.H., 1994. The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger* sequence divergence in Hymenoptera may be associated with
705 nucleotide content. J. Mol. Evol. 38, 282-294.

Johnson, N.K., Cicero, C., 2002. The role of ecological diversification in sibling speciation of *Empidonax* flycatchers (Tyrannidae): Multigene evidence from mtDNA. Mol. Ecol. 11, 2065-2081.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Amer. Statist. Assoc. 90, 773–795.

710 Knowles, L.L., Carstens, B.C., 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56, 887-895.

Knowlton, N., Jackson, J.B.C., 1994. New taxonomy and niche partitioning on coral reefs: Jack of all trades or master of none? Trends Evol. Ecol. 9, 7-9.

Lambert, D.M., Baker, A., Huynen, L., Haddrath, O., Hebert, P.D.N., Millar, C. D., 2005. Is a large-scale DNA-based inventory of ancient life possible? J. Hered. 96, 279-284.

Liljeblad, J., Ronquist, F., 1998. A phylogenetic analysis of higher-level gall wasp relationships (Hymenoptera: Cynipidae). Syst. Entomol. 23, 229-252.

Lopez-Vaamonde, C., Rasplus, J.Y., Weiblen, G., Cook, J.M., 2001. Molecular phylogenies of fig wasps: partial co-cladogenesis of pollinators and parasites. Mol. Phylogenet. Evol. 21, 55-71.

Machado, C. A., Hey, J., 2003. The causes of phylogenetic conflict in a classic *Drosophila* species group. Proc. R. Soc. Lond. B Biol. 270, 1193-1202.

Mayr, G. 1872. Die Einmiethler der mitteleuropäischen Eichengallen. Verhandlungen der Zoologisch-Botanischen Geselschaft in Wien 22, 669-726.

Melika, G., Acs, Z., Bechtold, M., 2004. New species of cynipid inquilines from China (Hymenoptera: Cynipidae: Synergini). Acta Zool. Acad. Sci. Hung. 50, 319-336.

Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. PloS Biology 3(12), e422.

Morando, M., Avila, L.J., Sites, J.W., 2003. Sampling strategies for delimiting species: Genes, individuals, and populations in the *Liolaemus elongatus-kriegi* complex (Squamata : Liolaemidae) in Andean-Patagonian South America. Syst. Biol. 52, 159-185.

Moritz, C., Cicero, C., 2004. DNA bar-coding: Promise and pitfalls. PLoS Biology 2, 1529-1531.

Nielsen, R., Matz, M., 2006. Statistical approaches for DNA barcoding. Syst. Biol. 55, 162-169.

Nieves-Aldrey, J.L., 2001. Hymenoptera, Cynipidae. Fauna Ibérica, vol. 16. (ed. M. A. Ramos et al.). pp. 636. Museo Nacional de Ciencias Naturales (CSIC). Madrid.

Nieves-Aldrey, J.-L., Pujade-Villar, J., 1985. Sobre las especies ibéricas de la Sección 1 (Mayr, 1872) del género *Synergus* Htg. Eos 61, 219–237.

Nieves-Aldrey, J.-L., Pujade-Villar, J., 1986. Sobre las especies ibéricas de la Sección II (Mayr, 1872) del género *Synergus* Htg. con descripción de una especie nueva (Hym., Cynipidae, Cynipinae). Eos 62, 137-165.

Nieves-Aldrey, J.-L., Vårdal, H., Ronquist, F., 2005. Comparative morphology of terminal-instar larvae of Cynipoidea: phylogenetic implications. Zool. Scripta 34, 15-36.

Nylander, J.A.A., Buffington, M.L., Liu, Z., Nieves-Aldrey, J.-L., Liljeblad, J., Ronquist, F., 2004a. Molecular Phylogeny and Evolution of Gall Wasps. In: Nylander JAA *Bayesian Phylogenetics and the Evolution of Gall Wasps*, PhD Thesis, University of Uppsala, Uppsala, Sweden.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves Aldrey, J.-L., 2004b. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47-67.

Papakostas, S., Triantafyllidis, A., Kappas, I., Abatzopoulos, T.J., 2005. The utility of the 16S gene in investigating cryptic speciation within the *Brachionus plicatilis* species complex. Mar. Biol. 147, 1129-1139.

Pénzes, Z., Melika, G., Bozsóki, G., Bihari, P., Mikó, I., Tavakoli, M., Pujade-Villar, P., Fehér, B., Fülöp, D., Szabó, K., Bozsó, M., Sipos, B., Somogyi, K., Stone, G.N. 2009. Systematic re-appraisal of the gall-usurping genus *Synophrus* Hartig, 1843 (Hymenoptera: Cynipidae: Synergini). Systematic Entomology, doi: 10.1111/j.1365-3113.2009.00482.x.

Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W.H., Casane, D., 2004.

Phylogenomics of Eukaryotes: Impact of Missing Data on Large Alignments. Mol. Biol. Evol. 21, 1740-1752.

Powers, T. 2004. Nematode molecular diagnostics: from bands to barcodes. Ann. Rev. Phytopathol. 42, 367-383.

Pujade-Villar, J. 1992. Sobre *Synergus ilicinus* (Barbotin, 1972) n. comb. y otras especies de *Synergus* Htg. en el Nordeste de la Peninsula Ibérica (Hym., Cynipidae, Synergini). Boletín de la Asociación española de Entomología 16, 129-148.

Pujade-Villar, J., Melika, G., Ros-Farré, P., Ács, Z., Csóka, G., 2003. Cynipid inquiline wasps of Hungary, with taxonomic notes on the Western Palaearctic fauna (Hymenoptera: Cynipidae, Cynipinae, Synergini). Folia Entomologica Hungarica 64, 121-170.

Pujade-Villar, J., Nieves-Aldrey, J.-L., 1990. Revisión de las especies europeas del género *Saphonecrus* Dalla Torre and Kieffer, 1910 (Hymenoptera: Cynipidae: Cynipinae). Bulletín de la Institució Catalana d'Historia Natural 58, 45–55.

Pujade-Villar, J. and J. L. Nieves-Aldrey. 1993. Revisión de las especies europeas del género *Ceroptres* Htg. 1840 (Hymenoptera: Cynipidae). *Boletin de la Asociación Española de Entomologia* 17, 49–63.

Quicke, D. 2004. The world of DNA bar-coding and morphology – collision or synergism and what of the future? The Systematist 23, 8-12.

Rannala, B., Huelsenbeck, J.P., Yang, Z., Nielsen, R., 1998. Taxon sampling and the accuracy of large phylogenies. Syst. Biol. 47, 702–710.

Robinson, D.R., Foulds, L.R., 1981. Comparison of phylogenetic trees. Mathematical Biosciences 53, 131-147.

Rokas, A., Atkinson, R.J., Brown, G., West, S.A., Stone, G.N., 2001. Understanding patterns of genetic diversity in the oak gall wasp *Biorhiza pallida*: demographic history or a *Wolbachia* selective sweep? Heredity 87, 294–305.

Rokas, A., Atkinson, R.J., Nieves-Aldrey, J.-L., West, S.A., Stone, G.N., 2002a. The incidence and diversity of *Wolbachia* in gallwasps (Hymenoptera; Cynipidae) on oak. Mol. Ecol. 11, 1815-1829.

Rokas, A., Atkinson, R.J., Webster, L., Csóka, Gy., Stone., G.N., 2003a. Out of Anatolia: longitudinal gradients in genetic diversity support an eastern origin for a circum-Mediterranean oak gallwasp *Andricus quercustozae*. Mol. Ecol. 12, 2153-2174.

Rokas, A., Melika, G., Abe, Y., Nieves-Aldrey, J.-L., Cook J.M., Stone, G.N., 2003b. Lifecycle closure, lineage sorting, and hybridisation revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. Mol. Phylogen. Evol. 26, 36-45.

Rokas, A., Nylander, J.A.A., Ronquist, F., Stone, G.N., 2002b. A maximum-likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae): implications for insect phylogenetic studies. Mol. Phylogen. Evol. 22, 206-219.

Ronquist, F. 1994. Evolution of parasitism among closely related species: phylogenetic relationships and the origin of inquilinism in gall wasps (Hymenoptera, Cynipidae). Evolution 48, 241-266.

Ronquist, F. 1999. Phylogeny, classification and evolution of the Cynipoidea. Zoologica Scripta 28, 139-164

Ronquist, F., Huelsenbeck., J. P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Ronquist, F., Liljeblad, J., 2001. Evolution of the gall wasp-host-plant association. Evolution 55, 2503–2522.

Ross, J. 1951. A study of some British species of *Synergus* (Hymenoptera, Cynipoidea). *Trans. Soc. Brit. Entomol.* 11, 81-95.

Rubinoff, D. 2006. Utility of mitochondrial DNA barcodes in species conservation. Conserv. Biol. 20, 1026-1033.

Sadeghi, S.E., Melika, G., Pujade-Villar, J., Pénzes, Z., Ács, Z., Bechtold, M., Assareh, M.H. Tavakoli, M., Yarmand, H., Askary, H., Stone, G.N., Azizkhani, E., Zargaran, M.R., Aligolizade, D., Barimani, H., Dordaei, A.A., 2006. Oak cynipid gall inquilines of Iran (Hym.: Cynipidae: Synergini), with descriptions of new species. J. Entomol. Soc. Iran 25, 15-50.

Sanver, D., Hawkins, B.A., 2000. Galls as habitats: the inquiline communities of insect galls. Basic Appl. Ecol. 1, 3-11.

Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18, 502-504.

Schönrogge, K., Crawley, M.J., 2000. Quantitative webs as a means of assessing the impact of alien insects. J. Anim. Ecol. 69, 841-868.

Schönrogge, K., Stone, G.N., Crawley, M.J., 1995. Spatial and temporal variation in guild structure: Parasitoids and Inquilines of *Andricus quercuscalicis* Burgsd. (Hymenoptera: Cynipidae) in its native and alien ranges. Oikos 72, 51-60.

Schönrogge, K., Stone, G.N., Crawley, M.J., 1996a. Alien herbivores and native parasitoids: rapid development of guild structure in an invading gall wasp, *Andricus quercuscalicis* (Hymenoptera: Cynipidae). Ecol. Entomol. 21, 71–80.

Schönrogge, K., Stone, G.N., Crawley, M.J., 1996b. Abundance patterns and species richness of the parasitoids and inquilines of the alien gall former *Andricus quercuscalicis* Burgsdorf (Hymenoptera: Cynipidae). Oikos 77, 507–518.

Smith, M.A., Rodriguez, J.J., Whitfield, J.B., Deans, A.R., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc. Nat. Acad. Sci. USA 105, 12359-12364.

Smith, M.A., Woodley, N.E., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). Proc. Nat. Acad. Sci. U.S.A. 103, 3657-3662.

Smith, M.A., Wood, D.M., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proc. Nat. Acad. Sci. U.S.A. 104, 4967-4972.

Starrett, J., Hedin, M., 2007. Multilocus genealogies reveal multiple cryptic species and biogeographical complexity in the California turret spider *Antrodiaetus riversi* (Mygalomorphae, Antrodiaetidae). Mol. Ecol. 16, 583-604.

Stone, G.N., Atkinson, R.J., Rokas, A., Nieves-Aldrey, J.-L., Melika, G., Ács, Z., Csóka, Gy., Hayward, A., Bailey, R., Buckee, C., McVean, G.A.T., 2008. Evidence for widespread cryptic sexual generations in apparently asexual *Andricus* gallwasps. Mol. Ecol. 17, 652-665.

Stone, G.N., Challis, R.J., Atkinson, R.J., Csóka, G., Hayward, A., Mutun, S., Preuss, S., Rokas, A., Sadeghi, E., Schönrogge, K., 2007. The phylogeographic clade trade: tracing the impact of human-mediated dispersal on the colonisation of northern Europe by the oak

gallwasp *Andricus kollari*. Mol. Ecol. 16, 2768-2781.

Stone, G.N., Hernandez-Lopez, A., Nicholls, J.A., di Pierro, E., Pujade-Villar, J., Melika, G., Cook, J.M., 2009. Extreme host plant conservatism during at least 20 million years of host plant pursuit by oak gallwasps. Evolution, 63, 854-869.

Stone, G.N., Schönrogge, K., 2003. The adaptive significance of insect gall morphology. Trends Ecol. Evol. 18, 512-522.

Stone, G.N., Schönrogge, K., Atkinson, R.J., Bellido, D., Pujade-Villar, J., 2002. The population biology of oak gall wasps (Hymenoptera: Cynipidae). Ann. Rev. Entomol. 47, 633-668.

Strimmer, K., von Haeseler, A., 1997. Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. Proc. Nat. Acad. Sci. USA 94, 6815-6819.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., Vogler, A.P.2003. A plea for DNA taxonomy. Trends Ecol. Evol. 18, 70-74.

van Noort, S., Stone, G.N., Whitehead, V.B., Nieves-Aldrey, J.-L., 2006. Biology and redescription of *Rhoophilus loewi* (Cynipidae: Cynipoidea: Hymenoptera), with evolutionary implications on the inquilinism in gall wasps. Biol. J. Linn. Soc. 90, 153-172.

Vogler, A.P., Monaghan, M.T., 2006. Recent advances in DNA taxonomy. J. Zool. Syst. Evol. Res. 45, 1–10.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N., 2005. DNA bar-coding Australia's fish species. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 360, 1847-1857.

Wiebes-Rijks, A.A. 1979. A character analysis of the species of *Synergus* Hartig, Section II (Mayr, 1872) (Hymenoptera, Cynipidae). Zool. Mededelingen 53, 297-321.Heraty J,

Hawks D, Kostecki JS, Carmichael A (2004) Phylogeny and behaviour of the
875 Gollumiellinae, a new subfamily of the ant-parasitic Eucharitidae (Hymenoptera:
Chalcidoidae). Syst. Entomol. 29, 544-559.

Table 1. Summary of ln Bayes Factor (BF) analyses of alternative models for the *cox1* (n=83 sequences), *cytb* (n=136) and 28S D2 (n=81) datasets. Ln Bayes Factors are calculated as twice the difference in lnHML (2ΔlnHML) between alternative models, calculated from 2 runs for each model in MrBayes. For each single gene dataset, BF tests strongly supported a strict clock model, which is used as the reference for comparison with models in which taxon monophyly was enforced. Our inference follows Kass and Raftery (1995), with ln BF of 2-6, 6-10 and >10 taken to represent respectively positive, strong and very strong support for the model with higher likelihood. For all three loci we used a GTR+I+G model, partitioned by codon position in *cox1* and *cytb*.

Model	*cox1*	*cox1* inference	*cytb*	*cytb* inference	28S D2	28S D2 inference
HML for strict clock GTR+I+G	-7039.8		-5875.5		-2242.2	
Synergus monophyly	**20.9**	Very strongly supported	**-2.8**	Very weakly rejected	**-5.4**	Weakly rejected, but monophyletic with a posterior probability >0.95 in the MrBayes consensus tree.
Synergus section 1 monophyly	-959.3	Rejected	-1457.5	Rejected	-273.9	Rejected
Synergus section 2 monophyly	-961.5	Rejected	-1454.7	Rejected	-260.2	Rejected
Saphonecrus monophyly	-19.3	Rejected	**6.5**	Strongly supported	-63.5	Rejected
Synergus hayneanus monophyly	-267.1	Rejected	-753.8	Rejected	-245.3	Rejected
Synergus pallipes monophyly	-229.0	Rejected	-438.6	Rejected	-153.4	Rejected
Synergus umbraculus monophyly	-617.1	Rejected	-1343.6	Rejected	-242.0	Rejected

Table 2. *coxI* MOTUs identified from the **maxdata** and **maxtaxa** alignments containing specimens of more than one morphological species. MOTU numbers refer to text Appendices 1 and 2, and Figure S1. The first two columns refer to results incorporating the inclusive 6.4% (42 bp) cut-off, while the righthand column summarises changes when the divisive 1.1% (7 bp) cut-off is used. Specimens with combined species names showed morphological characters indicative of each species in current morphology-based taxonomy. Numbers in brackets after species names refer to membership of Mayr's *Synergus* sections I and II.

MOTU number	Species combined at the 6.4% cut-off	Differences observed at the 1.1% cut-off
3	*Saphonecrus barbotini* + *Saphonecrus lusitanicus*	-› 2 single-species MOTUs
6	*Synergus hayneanus* (I) + *S. hayneanus/reinhardi* (I) + *S. hayneanus/umbraculus* (I)	No change
19	*Synergus flavipes* (I) + *S. hayneanus* (I) + *S. umbraculus* (I)	-› 4 MOTUs: 2 multispecies (*S.flavipes* + *S. umbraculus*; *S. hayneanus* + *S. umbraculus*), and 2 single-species (each *S. umbraculus*).
20	*Synergus gallaepomiformis* (II) + *S. pallicornis* (II) + *S. pallipes* (II) + *S. pallidipennis* (I)	-› 3 MOTUs: 1 multipecies containing all 4 species, and 2 single-species, single-sequence (*S. pallicornis*; *S. pallipes*).
21	*Synergus clandestinus* (I) + *S. crassicornis* (I)	-› 2 single-species MOTUs
24	*Synergus acsi* (II) + *S. flavipes* (I) + *S. palmirae* (II) + *S. variabilis* (II)	-› 4 MOTUs: 3 single-species (*S. acsi*; *S. flavipes*; *S. palmirae*) and 1 multispecies (*S. flavipes* + *S. variabilis*).
Added only in the **maxtaxa** alignment		
26	*Synergus consobrinus* (I) + *S. pallipes* (II)	-› 2 single-species MOTUs

Table 3. Morphological species allocated to multiple *coxI* MOTUs in the **maxdata** and **maxtaxa** alignments. MOTU numbers refer to text Appendices 1 and 2, and Figure S1. Specimens with combined species names showed morphological characters indicative of each species in current morphology-based taxonomy. Numbers in each cell refer to the **maxdata** alignment, with numbers of MOTUs in the **maxtaxa** alignment in brackets if different. Numbers after species names refer to membership of Mayr's *Synergus* sections I and II.

Species	Number of MOTUs at the 6.4% cut-off	Number of MOTUs at the 1.1% cut-off
Saphonecrus undulatus	1	2
Synergus apicalis/*tibialis* (II)	2	2
Synergus flavipes (I)	2	2
Synergus hayneanus (I)	4	5
Synergus pallicornis (II)	1	2
Synergus pallipes (II)	2(3)	2(4)
Synergus reinhardi (I)	2	2
Synergus umbraculus (I)	2	5
Total number of species in >1 MOTU	6	8

Figure Legends

Figure 1. Diagrammatic representation of the barcoding gap. (a) Phylogenetic representation of sampled sequences, showing the separation of intraspecific and interspecific sequence variation assumed in single locus barcoding. (b) The relationship between the threshold used to divide MOTUs (here, % sequence divergence) and the number of MOTUs defined in a given sample (MOTU richness). In this idealised example, the barcoding gap is revealed as a plateau in MOTU richness over a threshold range.

Figure 2. Likelihood maps produced using TreePuzzle 5.0 (Schmidt et al. 2002) showing the phylogenetic utility of the three molecular markers used in this study. The upper row shows the distribution of likelihoods for each of the possible quartets of taxa for each gene. The lower row indicates the proportion of quartets that were poorly resolved (central portion) and well resolved (corners) for each gene. The number within the inner triangle indicates the percentage of quartets whose vectors place them less than halfway from wholly unresolved (the centre of the larger triangle) to fully resolved (any vertex of the larger triangle).

Figure 3. Bayesian majority rule consensus phylogeny for *coxI*, assuming a GTR+I+G strict clock model of sequence evolution. Vertical bars at right indicate main clades within the Synergus complex of oak inquiline gallwasps. All unlabelled nodes have a posterior probability of ≥95%, and values for other nodes with support >50% are shown. Membership of Mayr's *Synergus* sections is indicated by a filled circle after the sample name for Section I and by an open circle for Section II. Taxon labels of the form a/b refer to specimens showing morphological characters of species a and b within *Synergus*. Full morphology-based identifications are given for each specimen in Appendix S1. Coloured taxon labels illustrate the separation of three morphospecies (*Synergus hayneanus* in green, *S. pallipes* in blue, *S. umbraculus* in red, *S. flavipes* in purple) among multiple MOTUs. Scale bar indicates 0.1 substitutions per site.

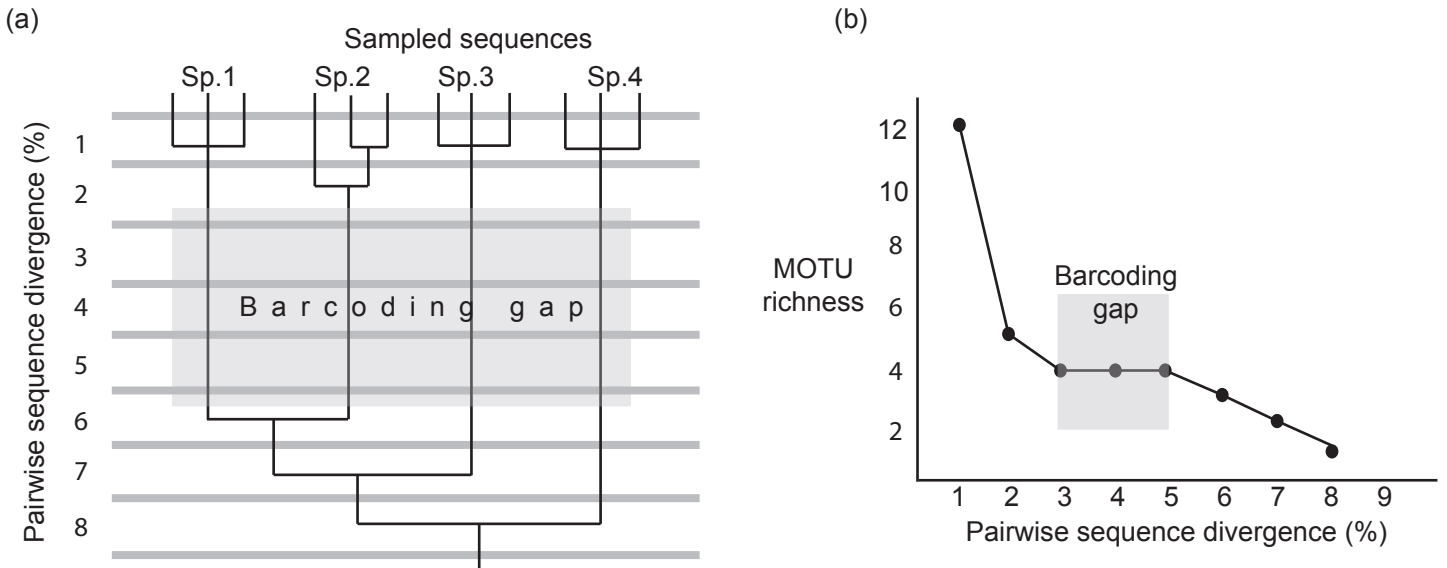
Figure 4. Bayesian majority rule consensus phylogeny for *cytb*, assuming a GTR+I+G strict clock model of sequence evolution. Vertical bars at right indicate main clades within the *Synergus* complex of oak inquiline gallwasps. All unlabelled nodes have a posterior probability of ≥95%, and values for other nodes with support >50% are shown. Taxon labels of the form a/b refer to specimens showing morphological characters of species a and b within Synergus. Coloured taxon labels illustrate the separation of three morphospecies (*Synergus hayneanus* in green, *S. pallipes* in blue, *S. umbraculus* in red, *S. flavipes* in purple) among multiple MOTUs. Full morphology-based

identifications are given for each specimen in Appendix S1. Scale bar indicates 0.1 substitutions per site.

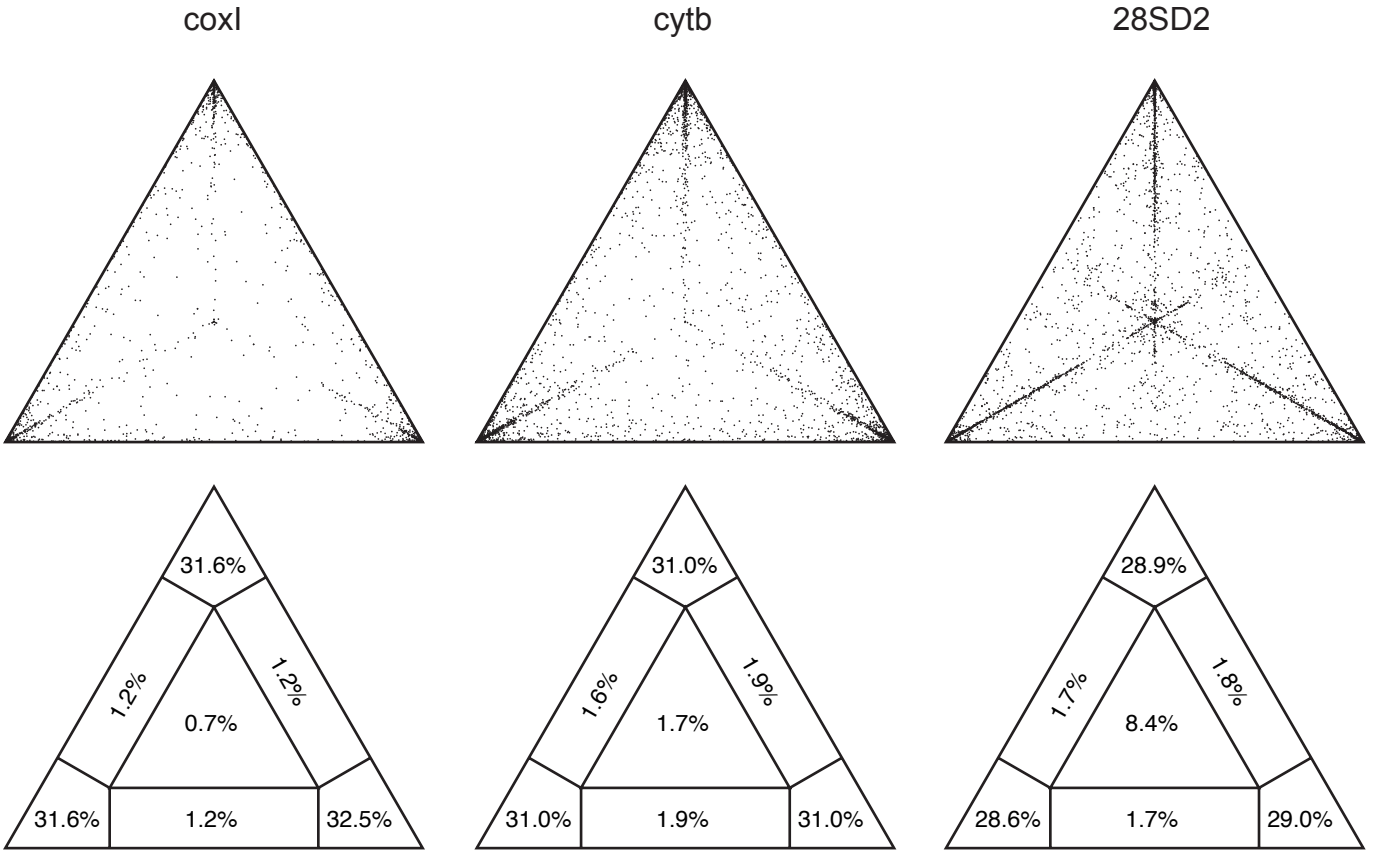
Figure 5. Bayesian majority rule consensus phylogeny for 28S D2, assuming a GTR+I+G strict clock model of sequence evolution. Vertical bars at right indicate main clades within the Synergus complex of oak inquiline gallwasps. All unlabelled nodes have a posterior probability of ≥95%, and values for other nodes with support >50% are shown. Membership of Mayr's *Synergus* sections is indicated by a filled circle after the sample name for Section I and by an open circle for Section II. Taxon labels of the form a/b refer to specimens showing morphological characters of species a and b within Synergus. Taxon labels of the form a/b refer to specimens showing morphological characters of species a and b within Synergus. Coloured taxon labels illustrate the separation of three morphospecies (*Synergus hayneanus* in green, *S. pallipes* in blue, *S. umbraculus* in red, *S. flavipes* in purple) among multiple MOTUs. Full morphology-based identifications are given for each specimen in Appendix S1. Scale bar indicates 0.01 substitutions per site.

Figure 6. Variation in the numbers of MOTUs defined at cutoffs between 0 and 12% for *coxI*, *cytb*, and 28S D2. Arrows indicate the selected divisive, intermediate and inclusive cutoffs for *coxI* (A-C), *cytb* (D-F) and the single cutoff for 28S D2 (G). Values are means ± 1 standard error for 100 sampling replicates.

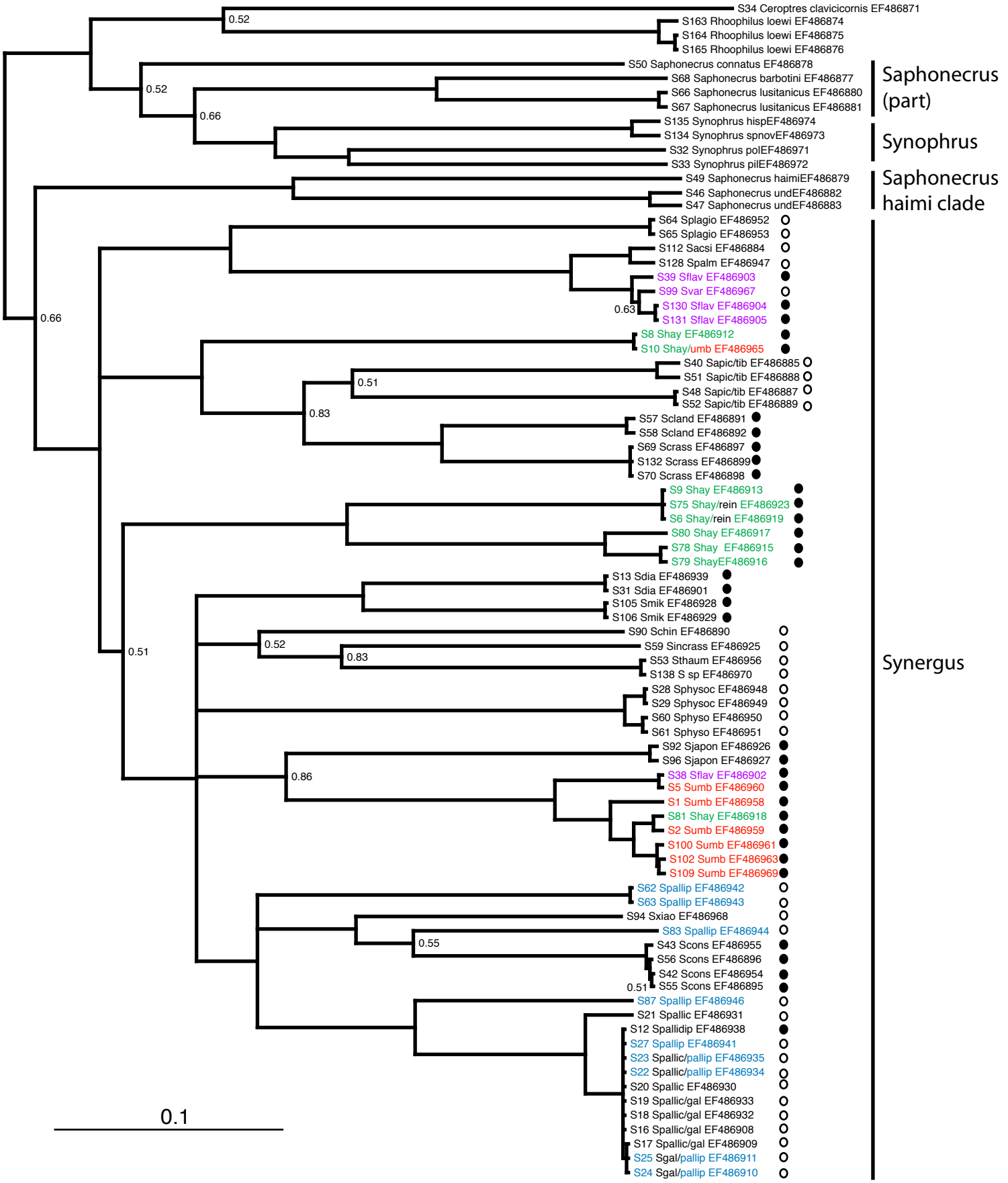
Acs et al. Fig.1



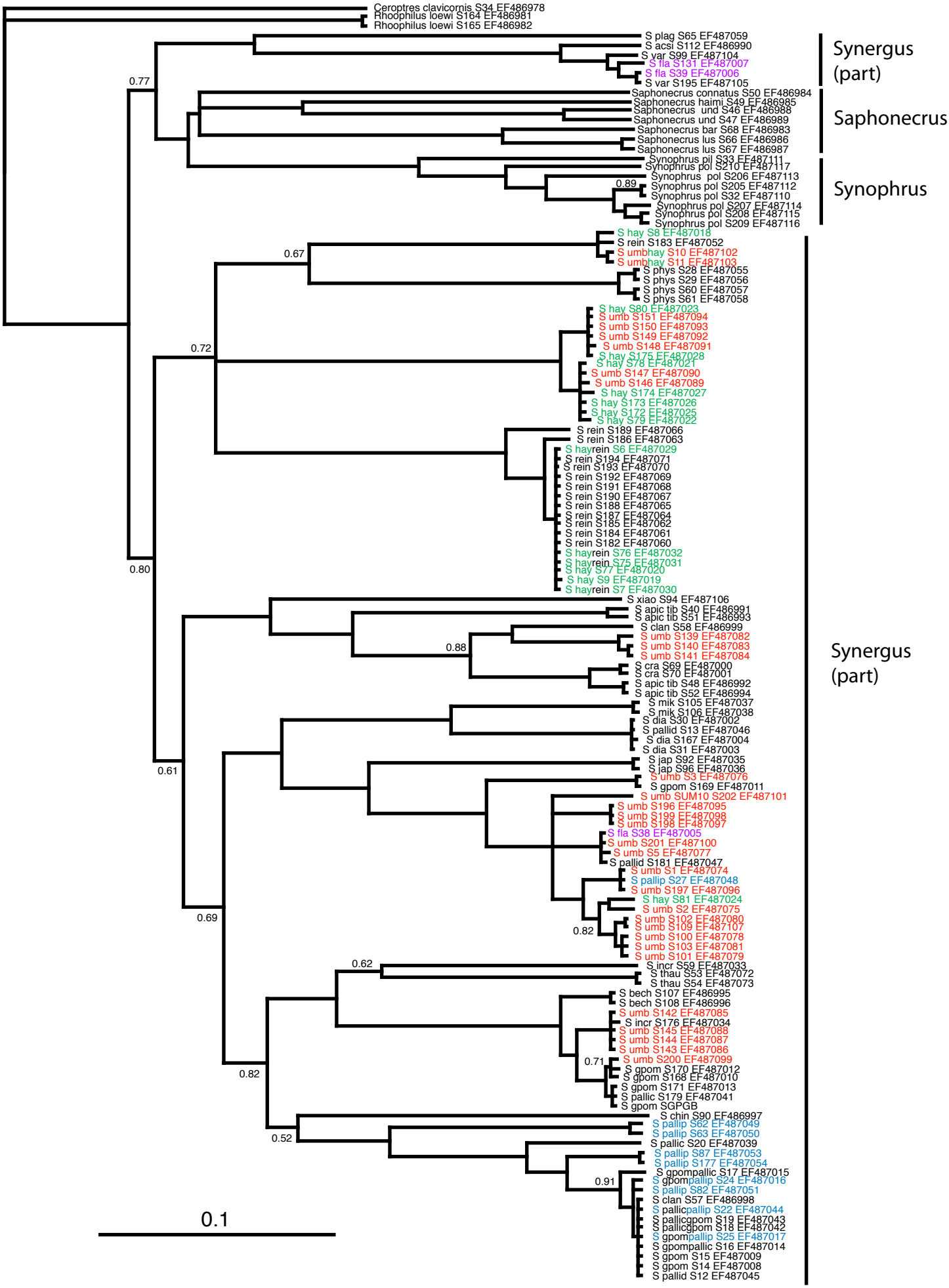
Acs et al. Fig.2



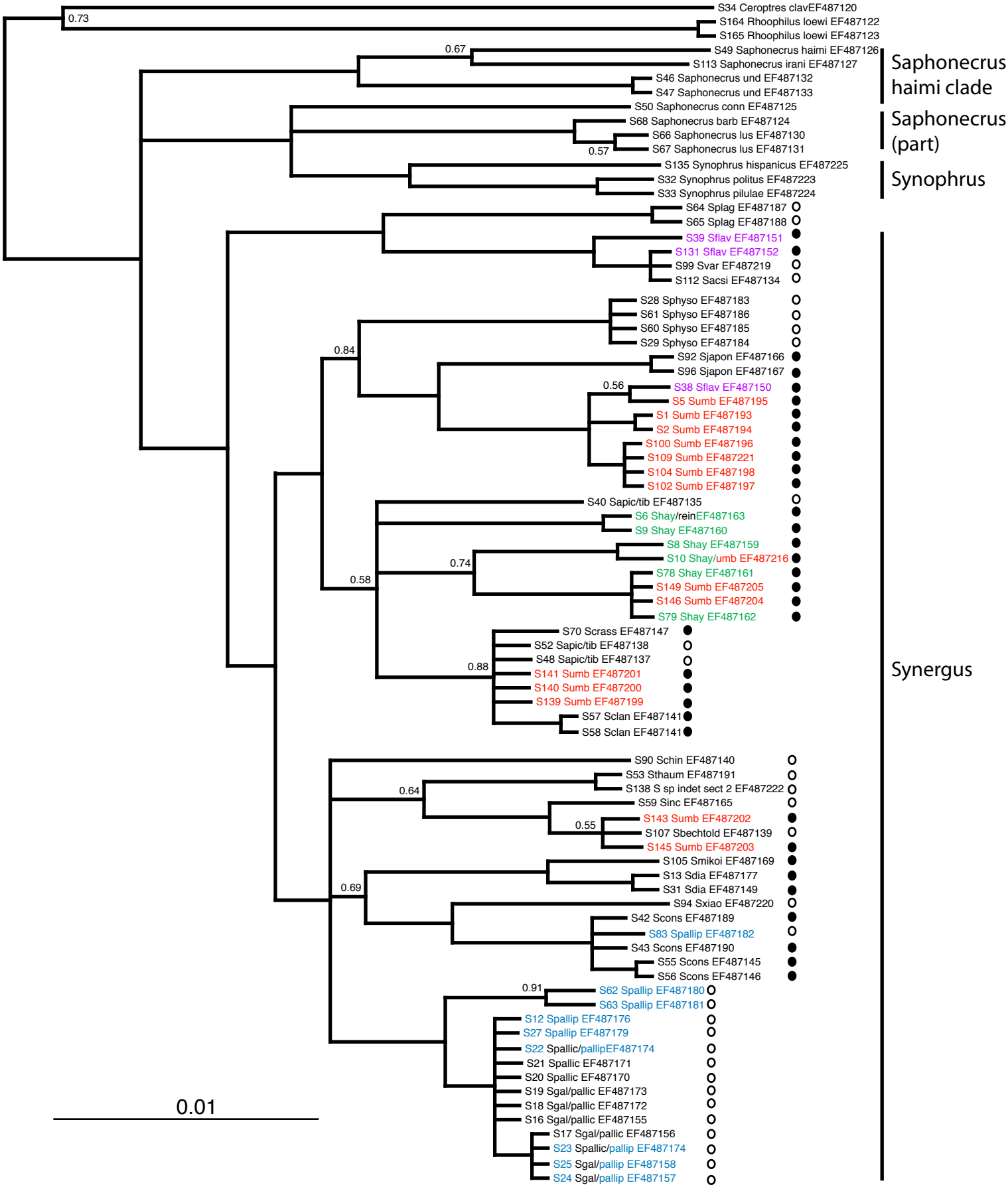
Acs et al. Fig.3



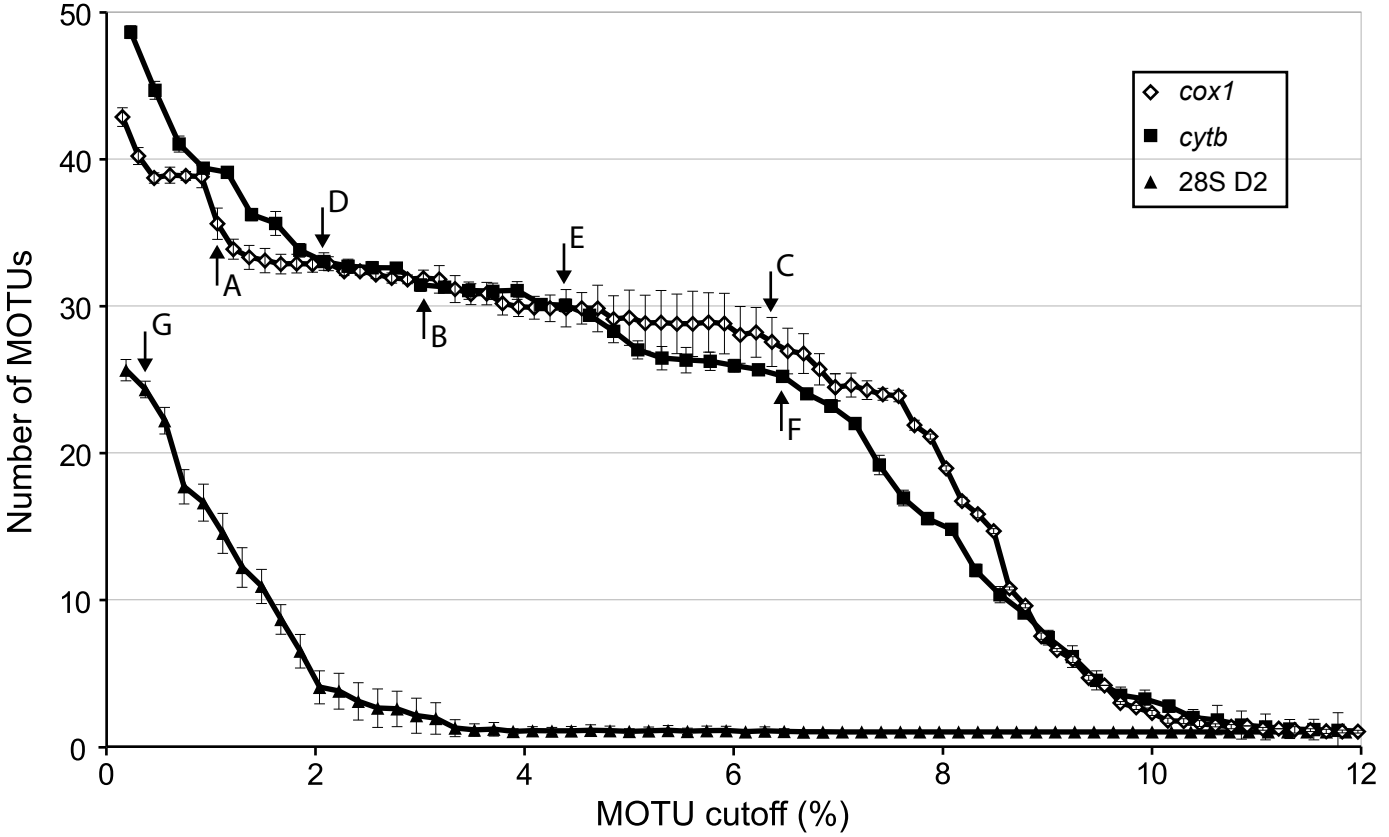
Acs et al. Fig.4



Acs et al. Fig.5



Acs et al. Fig.6



Supplementary Figure 1. Bayesian majority rule consensus phylogeny for the three-gene (*coxI*, *cytb* and 28 SD2) **maxdata** supermatrix, which includes at least one sequence for each *coxI* MOTU defined in Appendices 1 and 2. Full morphology-based identifications are given for each specimen in Appendix S1. Vertical bars at right indicate main clades within the Synergus complex of oak inquiline gallwasps. All unlabelled nodes have a posterior probability of ≥95%, and values for other nodes with support >50% are shown. MOTU numbers (Mx) are for *coxI* at the 6.4% (41 bp, inclusive) cut-off (Appendix 1). Membership of Mayr's *Synergus* sections is indicated by a filled circle after the sample name for Section I and by an open circle for Section II. Full morphology-based identifications are given for each specimen in Appendix S1. Scale bar indicates 0.1 substitutions per site.

Acs et al. Fig. S1