Hybrid capture data unravels a rapid radiation of pimpliform parasitoid wasps (Hymenoptera: Ichneumonidae: Pimpliformes)

Seraina Klopfstein[1,2], Barbara Langille[3], Tamara Spasojevic[1,2], Gavin R. Broad[4], Steven J. B. Cooper[3,5], Andrew D. Austin[3], Oliver Niehuis[6,7]

[1] University of Bern, Institute of Ecology and Evolution, Baltzerstrasse 6, 3012 Bern, Switzerland

[2] Naturhistorisches Museum Bern, Bernastrasse 15, 3005 Bern, Switzerland

[3] Australian Centre for Evolutionary Biology and Biodiversity, School of Biological Sciences, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia

[4] Natural History Museum, Department of Life Sciences, Cromwell Road, London SW7 5BD, UK

[5] South Australian Museum, Evolutionary Biology Unit, North Terrace, Adelaide, SA 5000, Australia

[6] Center for Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

[7] University of Freiburg, Department of Evolutionary Biology and Ecology, Institute for Biology I (Zoology), Hauptstraße 1, 79104 Freiburg, Germany (current affiliation)

Running head: Rapid radiation of pimpliform parasitoid wasps

Corresponding author: Seraina Klopfstein, Naturhistorisches Museum Bern, Bernastrasse 15, 3005 Bern, Switzerland, Tel. +41 31 350 72 45, E-mail: klopfstein@nmbe.ch

Abstract

The parasitoid wasp family Ichneumonidae is among the most diverse groups of organisms, with conservative estimates suggesting that it contains more species than all vertebrates together. However, ichneumonids are also among the most severely understudied groups, and our understanding of their evolution is hampered by the lack of a robust higher-level phylogeny of this group. Based on newly generated transcriptome sequence data, which were filtered according to several criteria of phylogenetic informativeness, we developed target DNA enrichment baits to capture 93 genes across species of Ichneumonidae. The baits were applied to DNA of 55 ichneumonids, with a focus on Pimpliformes, an informal group containing nine subfamilies. The target DNA capture efficiency was consistently high across the investigated species, including three distantly related outgroups. Phylogenetic trees were inferred under maximum likelihood and Bayesian approaches, both at the nucleotide and amino acid levels. We found maximum support for the monophyly of Pimpliformes but low resolution and very short branches close to its base, strongly suggesting a rapid radiation. Different analytical approaches recovered different trees, which precludes robust inferences of the evolution of different parasitoid strategies, such as idiobiosis and koinobiosis. Two genera and one genus-group were consistently recovered in unexpected parts of the tree, prompting changes in their higher-level classification: *Pseudorhyssa* Merrill, currently classified in the subfamily Poemeniinae, is transferred to the tribe Delomeristini within Pimplinae, and *Hemiphanes* Förster is moved from Orthocentrinae to Cryptinae. Likewise, the tribe Theroniini is resurrected for the *Theronia*-group of genera (stat. rev.). Phylogenetic analyses, in which we gradually increased the numbers of genes, revealed that the initially steep increase in mean clade support slows down at around 40 genes. Consideration of up to 93 genes still left various nodes in the inferred phylogenetic tree poorly resolved. It remains to be shown whether more extensive gene or taxon sampling can resolve the early evolution of the pimpliform subfamilies. However, the fact that the single-gene analyses did not reveal strong incongruences, but instead all showed low resolution at the base of Pimpliformes, suggests the possibility of a hard polytomy.

Keywords: next-generation sequencing, hybrid capture, phylogenomics, phylogenetic informativeness, classification

Introduction

The Ichneumonidae constitute the largest family of parasitoid wasps, with to date more than 25,000 species described and conservative estimates suggesting that several fold more are yet to be discovered . Ichneumonid wasps in the vast majority of cases attack immature stages of holometabolous insects and, to a lesser extent, immature and mature spiders . They can also be predators of arachnid egg sacs and even more rarely, facultatively herbivorous . They play a crucial role in most terrestrial ecosystems by regulating the abundance of other species, including pests in agriculture and forestry . Despite the fact that there are probably many more ichneumonid wasps than vertebrates (about 68,600 species; the group has received only limited attention from systematists and evolutionary biologists and the phylogenetic relationships among and circumscription of higher-level taxa within the family are still very poorly understood . These limitations hinder inferences pertaining to evolutionary pathways of parasitoid life styles, major changes in host ranges, and co-evolution with their hosts. The current classification of the family is based mostly on morphological data , and the only available molecular phylogeny of the family is based on a single gene (28S rRNA) that is highly sensitive to alignment parameters . The current consensus in ichneumonid taxonomy and phylogeny was summarized by Quicke and Broad *et al.* . The authors pointed out that the current classification recognising up to 42 extant subfamilies will likely be subject to future changes. Seven informal subfamily groups are currently recognised, mostly based on larval morphology and on results from a phylogeny obtained from the combined analysis of the 28S rRNA gene and a morphological character matrix . Three of these seven groups contain most of the ichneumonid species: Ichneumoniformes, Ophioniformes, and Pimpliformes .

We here focus on the clade Pimpliformes, which includes nine subfamilies, the Acaenitinae, Collyriinae, Cylloceriinae, Diacritinae, Diplazontinae, Orthocentrinae, Pimplinae, Poemeniinae, and Rhyssinae. They show a great variety in their exploited host groups and life history strategies. Specifically, they include parasitoids of, in descending order of importance, Lepidoptera, Hymenoptera, Coleoptera, Diptera, and Araneae; they act as ectoparasitoids and endoparasitoids, and as idiobionts and koinobionts . Idiobionts stop the development of their hosts at the time of parasitisation, with the wasp female usually paralysing the host larva or pupa permanently before laying her egg on top of or, more rarely, inside it . Koinobionts, in contrast, allow their hosts to continue their development, often over several moulting cycles; they are typically endoparasitoids, although the subfamily Tryphoninae and the *Polysphincta* genus-group in Pimplinae are conspicuous exceptions in being koinobiont ectoparasitoids . All four strategies are represented by several subfamilies in

the Pimpliformes, and several transitions have probably taken place, both between idiobiosis and koinobiosis and between ecto- and endoparasitism within this group. Also, a single transition to Diptera as a host group has been postulated , but the phylogenetic relationships among the pimpliform subfamilies are too uncertain to draw any robust conclusions .

In addition to studying subfamily relationships within Pimpliformes, we also use this study to examine operational questions regarding molecular phylogenetic analyses. Namely, we ask: how many and which genes are needed to resolve the pimpliform tree? A much-cited study from the early days of phylogenomics examined the relationships among eight species of yeast as inferred from 106 genes (comprising 127 kilobases (kb)) and concluded that "very large concatenated data sets of ~20 genes were required to provide 95% bootstrap support for all nodes" . This gene number might seem low now, given that hundreds to thousands of genes are regularly sequenced to infer large phylogenies . However, the discussion triggered by this study focused on two questions that are now more timely than ever: what are the causes of systematic biases in phylogenomics, and how many genes are really needed if they are chosen carefully?

The importance of the above questions is underlined by the increasing number of phylogenomic studies that arrive at contradictory conclusions, despite including hundreds to thousands of genes . The identification of systematic biases often requires re-analysis of the entire dataset, which can be prohibitive for very large datasets. The same applies to analyses that aim to study the influence of different taxon sampling schemes to assess the robustness of phylogenetic inference . Thus, increasing dataset size often comes at the cost of reduced rigour in the analysis methodology. The computational limitations imposed by many of the more advanced analytical methods, such as certain Bayesian inference methods, and the use of complex evolutionary models, represent another important argument for carefully choosing markers instead of simply increasing dataset size.

Several criteria have been suggested to aid in identifying the phylogenetically most informative genes, such as i) previous performance in a different, typically smaller taxon set ; ii) evolutionary rate and its distribution among sites ; iii) clock-like evolution ; and iv) stationarity of nucleotide or amino acid composition . However, different studies of empirical datasets have arrived at conflicting conclusions about the performances of these criteria , with the consequence that the pathway to efficient experimental design in phylogenomics remains obscure. It is even likely that identification of the best-performing approach will remain highly context-specific, with small differences in taxon sampling and analytical approach

making a large impact on the preferred strategy. Thus, only few criteria might emerge that have general applicability.

In any case, increasing the number of genes sampled can be expected to have a positive effect on the robustness of phylogenetic inference, at least up to a point, for instance by allowing identification of regions of uncertainty in a phylogenetic hypothesis that are due to incongruences among the individual gene trees . Next-generation sequencing facilitates the acquisition of unprecedented amounts of data, and techniques such as hybrid capture (also called anchored/hybrid enrichment or target DNA enrichment) enable cost-effective application even to non-model taxa . Hybrid capture approaches typically use RNA baits of about 100–150 bp in length, which show a close match to the target DNA sequences, in order to enrich the target DNA in a genomic extract . The enrichment step facilitates the pooling of large numbers of samples on a single sequencing lane, thus decreasing sequencing costs considerably. The technique has been applied to taxonomic groups with various levels of divergence, with capture success often still high in distantly related groups . Given that only a handful of markers have been established for ichneumonid phylogenetics in the past , and divergences within the Ichneumonidae likely date back to the upper Cretaceous , a hybrid capture approach is very promising for robustly inferring the phylogeny of this family.

We here aimed to resolve the higher-level relationships within Pimpliformes to gain insights into the evolution of this group, especially with respect to parasitoid strategy and host range. To this end, we used previously and newly generated transcriptome data to identify candidate genes, filtered them by several criteria of phylogenetic informativeness, and created a set of baits to enrich 152 exons of a total of 93 genes. We used this bait set to enrich the target DNA in 55 species of ichneumonids and two distantly-related outgroup taxa. Furthermore, we designed and tested oligonucleotide PCR primers for amplifying and subsequently Sanger-sequencing eight relatively long exons, identified in the captured gene set, which can be used in future ichneumonid studies. We applied several methods for phylogenetic reconstruction and examined them for congruence and conflict. Furthermore, based both on random and non-random gene sampling strategies, we studied subsets of our data to assess strategies to improve phylogenetic studies of the group in the future.

Materials and Methods

This study included five major steps: i) transcriptome sequencing and analysis, ii) estimation of phylogenetic informativeness of genes, iii) hybrid capture laboratory work and

data analysis, iv) development and testing of oligonucleotide primers for PCR and Sanger sequencing, and v) phylogenetic analyses. These individual steps are detailed below.

Most of the bioinformatics for this study were conducted using newly developed scripts, which are provided here as supplementary material (Supplementary File S1). We mostly used shell scripting (bash) to create a pipeline of programs for specific tasks and the APE package in the program R to manipulate alignments and trees. All calculations were performed on UBELIX (http://www.id.unibe.ch/hpc), the HPC cluster at the University of Bern, Switzerland.

Transcriptome data, orthology prediction and transcriptome phylogenies

Transcriptomes of two pimpliform wasp species were sequenced and published by Peters *et al.* . We additionally sequenced whole body transcriptomes of adult specimens for eight additional species from five pimpliform subfamilies and three other ichneumonids, including *Xorides praecatorius* Fabricius as a representative of the Xoridinae, that has been hypothesized to be sister to all remaining ichneumonids . For a full list of specimens used in this study, see Table 1 (and the more detailed Supplementary File S2). As we sampled a single species per genus in most cases, we henceforth refer to genera only. The specimens for the transcriptome analysis were freshly collected in a mountainous area in the canton of Bern, Switzerland, and either identified alive after a cooling period in the fridge, or collected alongside other, identical-looking individuals at the same location, with the latter kept in ethanol as 'pseudo-vouchers' or 'paragenophores' for later confirmation of the preliminary identifications (Supplementary File S2). Specimens chosen for the transcriptome analysis were either ground directly in RNAlater solution or shock-frozen at -80 °C. Total RNA was extracted using the Promega ReliaPrep™ Tissue Miniprep system. Library preparation was done using the Illumina TruSeq stranded RNA method, which includes poly-A selection, followed by size selection. The resulting 11 libraries were indexed, multiplexed and sequenced on one Illumina HiSeq3000 lane (150 bp paired-end, 250 million reads) at the Institute of Genetics of the University of Bern (VetSuisse).

The raw reads from transcriptome sequencing were de-multiplexed and trimmed using Trimmomatic 0.33 , which removed adapters, leading and trailing bases below a phred score of 10, and sections in a sliding window of size 4 with an average score below 15. Read quality was assessed with FastQC 0.11.2 . Transcriptomes were assembled with Trinity 2.1.1 in paired-end mode on six CPUs and with a maximum memory setting of 24 Gb of RAM under

188default settings. Combined with transcriptomes from the 1KITE project , this resulted in 19
189ichneumonid and six braconid transcriptomes (Table 1). The raw reads of the newly
190sequenced taxa are archived at the National Center for Biotechnology Information (Genbank),
191NIH, under BioProject PRJNA450386.

192 To identify orthologous transcripts of putative single-copy protein-coding genes
193among these 25 transcript libraries, we searched them for 3,260 genes that the OrthoDB
194database suggested to be single-copy in Hymenoptera . For this purpose, we used the
195software Orthograph , which employs the best reciprocal hit criterion, and the ortholog set
196designed by Peters *et al.* . Orthograph was run with the same settings and the identified
197transcripts were post-processed in the same manner (*e.g.*, stop codons were masked with
198NNN) as outlined by Peters *et al*. , except that we removed the DNA sequences of the
199following species, which were part of the ortholog set: *Acromymrex echinatior* (Forel), *Apis*
200*mellifera* Linnaeus, *Camponotus floridanus* (Buckley), *Harpegnathos saltator* Jerdon, and
201*Tribolium castaneum* (Herbst). We retained the DNA sequences of the jewel wasp *Nasonia*
202*vitripennis* (Walker), whose DNA and amino acid sequences were also part of the ortholog
203set. We exploited the gene modes of this species to cut the aligned DNA sequences of
204orthologous transcripts at putative exon-exon boundaries during target DNA enrichment bait
205design (see below). All orthologous DNA sequences were aligned at the amino acid level
206(provided by Orthograph) with MAFFT version 7.123 applying the L-INS-i alignment
207algorithm. The alignments obtained were used as blueprints to align the corresponding
208nucleotide sequences with a modified version of Pal2Nal version 14.1 .

209 We used a custom R script to trim leading and trailing amino acid positions in the
210resulting multiple DNA sequence alignments of each gene to ensure that the flanking
211positions had at least 90% coverage across the 26 included taxa. To obtain candidate loci for
212which to assess phylogenetic informativeness, we used a coverage filter, retaining only genes
213with at least 600 bp in total, at least 300 bp present in every taxon, and at most 5% of the
214alignment consisting of gaps or missing data. Only 723 genes fulfilled the above criteria and
215their alignments are provided as supplementary material (Supplementary File S3). In order to
216ensure that there was no cross-contamination between taxa resulting from the pooled Illumina
217sequencing procedure, we checked the alignments of the 723 genes for raw pairwise distances
218below 1 % and found only a single case which can be explained by a true close relationship
219(0.87 % between *Dolichomitus imperator* and *D. quercicolus* in gene Nasvi2EG0030079). To
220examine the phylogenetic signal in the transcriptome data, we conducted phylogenetic

analyses on this gene set with RAxML 8.1.2 , both at the amino-acid and nucleotide levels. Partitioning schemes and amino acid substitution models for each partition were obtained using PartitionFinder2 with the rcluster algorithm using equal weights on the clustering criteria rate, state frequencies, model parameters, and among-site rate variation. Since RAxML only allows for a single model across partitions when analysing nucleotide sequence data, even though the model parameters can be estimated separately for each partition, we applied the GTRCAT model when analysing the nucleotide alignments. When analysing the amino acid sequences, we applied the identified fixed-rate model from PartitionFinder in combination with the CAT model of among-site rate variation. To assess strength in signal, we ran 100 non-parametric bootstrap replicates. As all nodes obtained maximum bootstrap support, we also conducted gene jackknifing by randomly sampling sets of 10 genes 100 times. Each of these datasets was analysed in RAxML as amino acid alignments (under the PROTCATJTT substitution model, the most commonly selected model by PartitionFinder on the full dataset) and as first plus second codon positions (using the GTRCAT nucleotide substitution model), partitioning by gene and/or codon position.

Estimating phylogenetic informativeness

To obtain predictors of phylogenetic informativeness for the 723 candidate genes, we first trimmed each of them down to 600 bp that showed the best taxon coverage. We then obtained five predictors: achieved mean clade support per node, ML quartet mapping score, sites with optimal evolutionary rates, clocklikeness, and the stationarity of nucleotide frequencies. First, to get a measure of clade support, we ran single-gene ML analyses under the GTR + Γ + I model , partitioning by codon position and with 50 bootstrap replicates. We then recorded the sum of bootstrap support values for 12 nodes which were unequivocally retrieved in all phylogenetic analyses of the concatenated transcriptome data and that are in agreement with taxonomy and with previous phylogenetic studies : Braconidae, Ichneumonidae, Ichneumonidae without *Xorides* Latreille, Ophioniformes, Ichneumoniformes, Pimpliformes, (Pimplinae + Rhyssinae + Poemeniinae), Pimplinae, (Rhyssinae + Poemeniinae), Pimplini, and the genera *Pimpla* Fabricius and *Dolichomitus* Smith. We refer to this measure as the ‘taxonomic bootscore’. We also separately recorded if the rooting of the Ichneumonidae was in accordance with a previous molecular and morphological analyses , i.e., whether *Xorides* was recovered as the sister group to all other ichneumonids.

Maximum likelihood quartet mapping has been shown to perform well as a predictor of phylogenetic informativeness in datasets simulated at varying evolutionary rates . We obtained treelikeness scores under the HKY model for 1,000 randomly chosen taxon quartets, using a custom R script (Supplementary File S1).

To estimate evolutionary rates at each site of the alignment, we first transformed the tree obtained from the ML analysis of the complete amino acid dataset into an ultrametric tree using the penalized likelihood function 'chronopl' from the APE package in R . An ML estimate of the transition/transversion ratio was obtained for each gene using this tree and APE's 'optim.pml' function. Transition/transversion ratios and an ultrametric tree were both used as input for the program DNArates , which calculates site-specific evolutionary rates under ML. The number of sites which evolve at a rate between 0.1 and 0.4 expected substitutions per root-tip distance was recorded for each gene, a range found to result in near-optimal performance in a simulation study .

Clocklikeness of the genes was measured by running single-gene relaxed-clock analyses under the independent gamma rates (IGR) model in MrBayes 3.2 . The data were partitioned into first plus second versus third codon positions and analysed under the GTR + Γ + I model, with the topology constrained to include the taxonomic groups that comprised the taxonomic bootscore (see above) to avoid convergence issues and obtain comparable results. The median estimate of the variance of the clock model ('igrvar' parameter) was used as a measure of clocklikeness.

To test for deviations from the stationarity hypothesis of nucleotide composition, we used the Chi-square test of homogeneity as implemented in PAUP* 4.0 . This test does not account for phylogenetic signal in the data, but we assume that it is conservative and thus sufficient for our purpose of identifying and excluding genes that show a large variation in nucleotide composition across taxa.

Aiming for a good balance between the number of enriched genes and pooled taxa, we developed baits of 120 bp length each for 100 candidate genes using the program BaitFisher . In order to span the sequence variation that can be expected in such a taxonomically broad sample, we included separate baits for each cluster of sequences with at least 0.12 nucleotide sequence divergence for a specific bait region. We used the intron-exon boundaries from the *Nasonia* genome to inform bait positions and, in a second step, filtered the obtained baits against the genome of *Microplitis demolitor* Wilkinson , a member of the Braconidae and, at

least at the time of our bait design analyses, the closest relative to our ingroup with an annotated genome. Baits were only retained if they had a single hit and 95% overlap in the *Microplitis* genome. We applied different tiling designs: we tiled five baits with an offset of 20 bp across a target coding exon in 51 genes (68 exons), three baits with an offset of 20 bp across a target coding exon in 36 genes (77 exons), and a single bait to capture a given target coding exon in six additional genes (8 exons). In total, we designed 5,419 baits to capture 153 exons for a total of 93 genes. Bait nucleotide sequences are provided in FASTA format in the supplementary material (Supplementary File S4) and were ordered as RNA baits from the company MYcroarray (now Arbor Biosciences, Ann Arbor, USA).

Hybrid enrichment laboratory work

The wet laboratory procedure was as follows: genomic DNA was extracted using the Gentra Pure-gene extraction kit (Gentra systems, Minnepolis, MN USA). Extractions were from either single legs or whole bodies, with the extracted specimens kept as vouchers in a number of museum collections (Supplementary File S2). DNA concentration was measured using fluorometric quantification (Invitrogen Qubit) for all samples and for about a quarter of the samples also with the Agilent 2200 TapeStation System. DNA was sheared to an average fragment distribution of about 400 bp using a Diagenode Bioruptor sonicator, with the shearing success subsequently verified on the TapeStation. Libraries were constructed following Meyer and Kircher , using reagents from ThermoScientific and unique combinations of dual i7 and i5 index sets to allow for subsequent pooling. BeckmanCoulter AMPure XP beads were used for all purifications following the bead ratios specified by Meyer and Kircher . We assessed the success of library preparation by measuring DNA concentration with fluorometric quantification and for about a third of the samples also with a standard qPCR run using the LightCyler 96 Real-Time PCR System. In some of the libraries, we detected adapter dimers after running them on a standard electrophoresis gel. We cut these from an agarose gel, which had been stained with GelGreen, and purified the gel slices excluding the adapter dimers using Qiagen MinElute Gel Extraction columns and the manufacturer's protocol.

Prior to capture, we divided the target baits into ½ capture reactions. We enriched one to four samples per ½ capture depending on importance of the taxa and similarity of the DNA concentration in the libraries. The target capture was performed with a 24 h hybridization at 65 ºC following the MYbaits protocol v3 from Microarray. The post-capture amplification was completed using KAPA HiFi DNA Polymerase (Kapa Biosystems, USA) with the

following PCR protocol: 98 ºC for 30 s, followed by 12 cycles of 98 ºC for 20 s, 60 ºC for 30 s, and 72 ºC for 1 min, with a final extension of 5 min at 72 ºC. Illumina IS5 and IS6 primers were used to attach to the i7 and i5 ends of the library.

We verified enrichment success with a final run on the Agilent 2200 TapeStation System. Size selection was performed using AMPure XP beads in order to remove the small nucleotide fragments that may have been introduced during the capture. Samples were multiplexed into three pools of 14, 24, and 23 samples at equal concentrations and were run on one lane each on the Illumina MiSeq (pool 1) and the Illumina Next-Seq platforms (pools 2 and 3), obtaining 150 bp paired-end reads (Australian Genome Resource Facility (AGRF) in Adelaide, Australia).

From raw reads to alignments

We closely followed the strategy described by Bragg *et al*. , which is tailored for hybrid capture data from highly divergent taxa, but developed our own scripts. All scripts used in our bioinformatics pipeline are publicly available, including detailed comments (Supplementary file S1).

The paired-end reads were de-multiplexed and partial adapter sequences removed using the program Cutadapt . We used Trimmomatic to remove leading and trailing bases below a phred score of 30 and sections in a sliding window of size 4 with an average score below 25. Read quality was ascertained with FastQC 0.11.2 . BlastX was used to search for reads in each sample that showed a significant match with each of the enriched genes, a strategy making use of the more conserved amino-acid regions in comparison to nucleotide sequences across taxa separated by long divergence times . Only after this step were reads assembled into contigs separately per gene and sample, first, by using the program Velvet to generate contigs under a range of Kmer values (25 to 75 bp), and second, by merging (some of) the resulting contigs with the software Cap3 . To ascertain that the identified contigs indeed represent the targeted single-copy genes, we used BlastX to align the contigs to the predicted mRNA of the *Nasonia* official gene set , only retaining best matches. For most genes, this resulted in more than a single contig, as introns were often too long to be fully captured. To identify intron-exon boundaries, we used the 'protein2genome' method in the aligner Exonerate with all the transcriptome sequences as queries, keeping the longest match. Exonerate outputs the sequence identities of its matches, and we removed those that were below 60%, a threshold obtained by comparing the transcriptome sequences among each

other. After translation into amino acids, the remaining matching contigs were aligned to all transcriptomes and to each other in the multiple-sequence aligner MAFFT using the E-INS-i algorithm . The contigs were then translated back into nucleotide sequences using a custom R script. As Exonerate only finds complete introns, and contigs often started and/or ended in intronic sections, we developed an R script that identifies leading and trailing intron portions based on a drop in the pairwise distance when intron-exon boundaries are crossed. This script identified many but not all remaining intronic sequences, leaving out mostly those overhangs which were too short to calculate a meaningful distance. Thus, we had to correct some alignments manually (a log of these changes is given as Supplementary File S5). The remaining contigs were used as queries in a BlastN search against the original reads with a restrictive e-value of 10^{-10} to obtain coverage values and perform variant calling. Sequences were retained from a minimum coverage of eight identical bases, with variants called as ambiguity codes at a minimum coverage of four or at least 10% of the total coverage, with most positions obtaining much higher coverage (Supplementary File S6). The alignments of all 93 genes are provided as supplementary material (Supplementary File S7).

Primer development for Sanger sequencing

One of the main hindrances for developing PCR oligonucleotide primers for amplification and direct Sanger sequencing of nuclear protein-coding genes is the short average length of exons in most arthropod genes; this is problematic because both the length variability and high AT content of introns can pose problems for successful PCR and sequencing. We initially attempted to develop primer pairs for the eight genes with the highest phylogenetic informativeness, but in almost all cases found several introns ranging up to more than 1,000 base pairs in length within each of the target segments. To overcome this issue, we used the intron-exon boundaries identified in the *Nasonia* genome, together with the data obtained from the hybrid enrichment, to search for long exons (> 300 bp) or for exons separated by an intron that was relatively short (less than about 200 bp in most of the taxa whose genes we captured via hybrid enrichment; Table 2).

Primer design was done manually in Geneious 7.1.3 based on the *Nasonia* genome and the obtained ichneumonid transcriptome and hybrid capture sequences, while primer synthesis was by Microsynth (Balgach, Switzerland). The primer pairs (Table 2) were tested in a three-step procedure: i) touchdown PCR with a provisional annealing temperature on eight taxa sampled across ichneumonids; ii) PCR without touchdown with a temperature gradient on four new taxa; iii) touchdown PCR at the optimal temperature from step 2 on five new and

three previously failed taxa. The PCR mixtures (25 µl) contained 10 pmol of each forward and reverse primer, 12.5 µl GoTaq® Hot Start Green Master Mix (Promega, USA) and 1.5 µl DNA extract. PCR protocols consisted of an initial denaturation for 3 min at 94 °C, followed by 36–37 cycles of 30 s at 94 °C, 45 s at optimal annealing temperature (Table 2), 1–1.5 min at 72 °C, followed by a final extension step for 10 min at 72 °C. For the touchdown protocol, two cycles with annealing temperature 5 °C higher and two cycles with annealing temperature 2 °C higher than optimal were added before 36–37 cycles at annealing temperature. The successfully amplified DNA fragments were sequenced by LGC Genomics (Berlin, Germany). Finally, we applied the primers to a broad set of more than 100 ichneumonid and two non-ichneumonid species to ascertain broad applicability.

Phylogenetic analyses

The final set of alignments covering the 93 enriched genes was subjected to various phylogenetic analyses. First, we analysed it under the maximum likelihood (ML) criterion in RAxML 8.1.2 after determining partitioning schemes in PartitionFinder2 , as described for the phylogenetic analyses of the transcriptomes. As RAxML only allows for a single substitution model for all nucleotide partitions, we ran the GTRCAT model for the nucleotide alignments, while the amino acid alignments were analysed under the partition-specific substitution chosen by PartitionFinder2 and the CAT model of among-site rate variation. To assess signal strength, we ran 1,000 non-parametric bootstrap replicates. The concatenated genes were analysed at the nucleotide level by including i) all codon positions and ii) first and second codon positions only. We additionally analysed the concatenated dataset at the amino acid level. The amino acid alignment was also subjected to Bayesian analyses using fixed amino-acid substitution models (as obtained from PartitionFinder2) in MrBayes 3.2 and the empirical profile mixture model implemented in PhyloBayes-MPI 1.7 . In the analyses with MrBayes, we increased the relative probability of the topology and branch-length moves to 200 in order to counter-balance the large number of substitution model parameters estimated separately for the 56 partitions. In both Bayesian analyses, we conducted four independent runs to assess convergence.

Reconstructing the evolution of parasitoid lifestyle

As an example of an important aspect of the parasitoid lifestyle, we used ancestral state reconstruction to examine transitions between idiobiosis and koinobiosis in

ichneumonids. Biological data for the included genera were extracted from several sources . For many ichneumonid genera, no sufficiently detailed records of their parasitoid strategy exist, but if such information was available, it typically indicated that closely related genera and even members of the same subfamilies shared the same strategy. Thus, we extrapolated from related taxa in several cases (see Supplementary File S8 for details). Ancestral states were reconstructed under maximum parsimony using the 'hsp_max_parsimony' function in the R package 'castor' , which allows for reconstruction of missing tip data. We assessed the impact of topological uncertainty on the number and direction of evolutionary transitions by including consensus and ML trees from the various Bayesian and ML analyses, plus 1,000 trees each from the Bayesian posterior distributions as obtained from the analyses of the amino acid alignments with PhyloBayes and MrBayes, respectively.

Increase of tree resolution with dataset size

To assess the relationship between the number of genes included and the mean clade support of the resulting tree, we randomly sampled two to 90 genes, with 20 replicates for up to 20 genes and 10 replicates from 30 genes. Each concatenated gene set was analysed in RAxML at the nucleotide level, partitioned by codon position and using the GTRCAT model with 50 bootstrap replicates. Mean clade support was measured as the sum of bootstrap support values above 50%.

In order to re-evaluate the success of the criteria 'clade support', 'clocklikeness' and 'evolutionary rates', we estimated their scores again under the full taxon sampling for each gene separately. For the clade support measure, RAxML analyses were conducted at nucleotide level, partitioning by codon position and under a GTRCAT model. Genes were ranked according to decreasing mean clade support as obtained from 100 bootstrap replicates. To obtain a measure of the clocklikeness of the genes, we ran Bayesian relaxed-clock analyses under the independent-gamma-rates model (IGR) in MrBayes 3.2 . We used the uniform branch length and topology prior and an exponential distribution with rate 1 as the prior on the clock variance. As the topology signal in some single-gene datasets was quite low, we used three constraints in order to force a reasonable rooting in the clock analyses : Ichneumonidae, Braconidae, and Ichneumonidae + Braconidae. Analyses were run for 10 Mio generations, a value that was increased to 15 Mio generations if the average standard deviation of split frequencies (ASDSF), a measure for topology convergence, had not dropped below 0.02. The genes were then sorted according to descending median clock variance. As a third criterion, we estimated site-specific rates in the program DNARates as previously for

448the transcriptome sequence data and ranked the genes according to the number of sites in the
449'near-optimal' rates bracket .

450**Results**

451*Transcriptomes and phylogenetic informativeness*

452 An average of 10^8 bp ($8.2x10^7 – 1.3x10^8$) were sequenced from each transcriptome
453library, which assembled into ca. 10^5 contigs ($7.2x10^4 – 1.5x10^5$). Both these numbers are
454about 3.6 times larger than those of the transcriptome data published by Peters *et al.*
455(Supplementary File S9), which can be explained by the fact that we pooled fewer samples
456per lane. Searching for 3,260 single-copy genes, we found comparable numbers in our
457transcriptomes and those published by Peters *et al.* (3,059 compared to 2,969). However,
458different genes were missing in different taxa, so that only 1,908 genes were present across all
45925 transcriptomes. Filtering these for sections of at least 600 bp lengths and good coverage
460across all taxa resulted in 723 genes spanning more than 568 kbp in length.

461 Analysing the above 723 genes both as concatenated amino acids, first and second
462codon positions, and full nucleotide alignments under ML, we recovered partly conflicting
463trees (Fig. 1), each of which obtained maximal bootstrap support at every node. Such a result
464might be expected from a taxon sampling as sparse as this one, even when a large number of
465genes are analysed, and thus we focus on the few well-supported relationships and on
466identifying the main regions of conflict among the transcriptome trees. As expected from
467previous studies, *Xorides* was recovered as the sister to the remaining ichneumonids and the
468higher taxonomic groupings Ophioniformes, Ichneumoniformes, and Pimpliformes were
469recovered in all three datasets. However, the relationships among the pimpliform subfamilies
470showed conflicts and the subfamily Pimplinae, which was recovered as monophyletic in the
471amino-acid analysis, was split into two groups in the nucleotide analyses, both when
472including and when excluding third codon positions. The two analyses at the nucleotide level
473showed a single conflict in regard to the placement of *Rhyssa* Gravenhorst either as sister to
474*Deuteroxorides* Viereck + Pimpinae II (Fig. 1) or as sister to (*Collyria* Schøidte +
475*Coleocentrus* Gravenhorst) + *Dolichomitus*. In order to identify regions in the tree with lower
476support, we performed gene-jackknifing based on 100 datasets of 10 randomly chosen genes
477each, analysed at the amino acid level or including first and second codon positions. These
478analyses recovered mostly, but not always, the same nodes that were in conflict between
479datasets as less well supported in the respective trees. These nodes also tended to be separated
480by comparatively short branches in all analyses (Fig. 1).

The 723 candidate genes were filtered based on their degree of phylogenetic informativeness, considering five measures (Table 3): i) no significant deviation from compositional stationarity in the nucleotides, including third-codon positions (p-value of Chi-square test above 0.05); ii) clock variance below 0.015 ; iii) root of Ichneumonidae correct (*Xorides* recovered as sister to the remaining ichneumonid taxa); and (iv) at least 80 of the 600 nucleotides evolving at a near-optimal rate. This filtering approach left 117 genes, which were ranked by the taxonomic bootscore (sum of bootstrap support for 12 chosen groups recovered in the 723-gene and previous analyses, see Materials and Methods), and the 100 best genes were chosen for the hybrid capture approach. Seven of these turned out not to include any exons that were long enough for bait design, at least not among the exons for which we had sufficient taxon coverage, which left a set of baits to capture 152 exons for a total of 93 genes.

Hybrid capture success and Sanger sequencing primers

All 93 target genes were enriched successfully, with an average length of the captured coding target sequence of 376 bp per taxon after quality and coverage filtering (across genes: median = 415 bp, min = 150 bp, max = 1,197 bp; see Supplementary File S9). Calculated across all 59 taxa in the hybrid capture approach, the minimum coverage across the target genes was 331 bp, and on average 90 genes were recovered with more than 100 bp. The exception was the outgroup sample PTE_cf_*Eupelmophotismus* from the superfamily Chalcidoidea, which had much lower enrichment success (108 bp on average across genes, and only 19 genes with more than 100 bp). It is unclear if this was due to its taxonomically more distant position from the taxa on whose DNA the bait design was based on, or due to the low quality of the DNA extract from the rather small specimen that we studied. Another even more distantly related outgroup, *Gasteruption* Latreille (Evanoidea), showed only slightly lower success when compared to the ingroup samples (337 bp on average per gene, 83 genes with more than 100 bp). Comparing the coverage of the hybrid capture with the transcriptome sequence data, we found that the latter performed better, with an average of 436 bp across taxa and genes. The full alignment of 93 genes for 84 taxa contained 41,565 bp with a gap proportion of 14%, a value that dropped to 12.4% when only the 75 ichneumonid taxa were included.

Trying to identify comparatively long exons, in the entire 93-gene dataset we found only four exons that were longer than 450 bp and eight more that were longer than 350 bp. Nevertheless, we developed oligonucleotide primers for eight genes that amplified and

sequenced with good consistency (Table 2, Supplementary File S10). In one case (gene NasViEG013087), a combination of two primer pairs were needed to obtain target amplicons from DNA of the majority of taxa. In two cases, the amplified fragments spanned one intron, which was almost consistently about 100 bp in length, even in distantly related taxa. The exonic portion of the fragments ranged between 328 bp and 536 bp (Table 2) and was amplified in species across the entire ichneumonid tree and represented most of the tested subfamilies. Success was lower in the two specimens from different families (Braconidae and Gasteruptiidae), but additional testing is needed to see if that was due to low quality of the extracts or divergent sequences.

Phylogeny of Pimpliformes

The five different phylogenetic analyses that we performed recovered different trees depending on data type and analysis method, with most incongruence located close to the base of the pimpliform radiation (Figs 2 and 3). This area of the resultant phylogenies also exhibited very short branches. The incongruent nodes were mostly the same as those conflicting between the nucleotide versus amino acid versions of the transcriptome analyses, but appeared much more complex because of the inclusion of additional subfamilies in the hybrid capture taxon sampling (Fig. 1). For several nodes at the base of the pimpliform radiation, no analysis retrieved a significant signal (clade support below 0.95 posterior probability or 85% bootstrap support: Fig. 3), and most of the gene trees were also unresolved at these nodes (Supplementary File S11). In contrast, the base of the inferred tree was well resolved. *Xorides* as the representative of Xoridinae consistently appeared at the base of a monophyletic Ichneumonidae, and the three informal subfamily groups Ophioniformes, Ichneumoniformes and Pimpliformes were recovered with maximum support in all analyses. The only exception from this consistently well-supported resolution was the placement of the only representative of the Labeninae (*Poecilocryptus* Cameron), which branched off either right after the branch leading to *Xorides* (Figs 3b, c), Ophioniformes (Figs 3a, d, e), or Ichneumoniformes (Fig. 3f). The latter would imply a sister group relationship between Labeninae and Pimpliformes.

The monophyly of most pimpliform subfamilies was well-supported by all analyses. Exceptions were the Pimplinae (because the genus *Xanthopimpla* Saussure and sometimes also the tribe Pimplini took up a position closer to the base of the pimpliform radiation), and Acaenitinae (with the tribe Acaenitini recovered separately from *Coleocentrus* in the tribe Coleocentrini). Pimplinae (excluding *Xanthopimpla*) retrieved maximum support only in the

amino acid analyses (Figs 3a, b, c, d), while Orthocentrinae were only supported by Bayesian inference analyses (Fig. 3a, b, c). The only relationships among higher groups of Pimpliformes that were recovered consistently were sister group relationships between Rhyssinae and Poemeniinae, between the pimpline tribe Delomeristini and the *Theronia* genus-group, and (in all but the ML analysis of codon positions 1 and 2) between the Orthocentrinae and Diacritinae. The pimpline tribe Ephialtini was consistently recovered as monophyletic, which was also true for Pimplini (excluding the *Theronia* genus-group) and for Delomeristini (including *Pseudorhyssa* Merrill).

Three taxa were consistently recovered in places that are in conflict with their current classification (Figs 2 and 3). The genus *Pseudorhyssa*, regarded as a member of Poemeniinae since Gauld (1991), was placed with maximum support in all analyses in the tribe Delomeristini (Pimplinae). The *Theronia*-group of genera (represented in our analyses by one species each of *Theronia* Holmgren, *Nomosphecia* Gupta and *Parema* Gupta) appeared as the sister to the Delomeristini, including *Pseudorhyssa,* while *Hemiphanes* Förster, which has been classified as a member of the informal *Helictes* genus-group within Orthocentrinae, grouped with the Ichneumoniformes, a placement that also received maximum support, even though its position within Ichneumoniformes is not resolved. In the cases of both *Hemiphanes* and *Pseudorhyssa*, the new placements were also recovered in most of the single-gene trees, while they were never placed with the subfamilies they are currently classified in (Supplementary File S11). The placement of *Pseudorhyssa* with Delomeristini was supported in 89 out of 93 gene-trees, 44 of them with bootstrap support of at least 85%, and that of *Hemiphanes* with Ichneumoniformes by 81 gene-trees, 22 of them with at least 85% bootstrap support. The *Theronia*-group showed more variation in its placement among the single-gene trees, appearing with Delomeristini in 25 and with Pimplini in 15, while it showed various placements in the remaining gene trees (Supplementary File S11).

Inferring the evolution of parasitoid life styles

Despite the low resolution at the base of the pimpliform radiation, the maximum parsimony reconstructions of parasitoid ecology retrieved very similar numbers of transitions across analyses (Fig. 3, Table 4). However, the directions of change within Pimpliformes varied strongly among analyses, with the most recent common ancestor of Pimpliformes recovered as an idiobiont with high probability in the PhyloBayes and nucleotide analyses, while it was retrieved as a koinobiont in most trees for the MrBayes analysis of the amino acid data (Table 4; the bootstrap trees from the ML analysis at the amino acid level were

equivocal). Nevertheless, it is interesting to note that the idiobiont and koinobiont Pimpliformes, tended to cluster together in most of the analyses, with the exception of *Xanthopimpla* (in most trees) and the derived *Polysphincta*-group of genera. For two included taxa, the mode of parasitoidism is unknown; the reconstruction for Diacritinae was always in favour of it being a koinobiont, while it was sometimes equivocal for *Coleocentrus*.

How many genes do we need?

Random subsampling of the full dataset of 93 genes resulted in a curve of increasing mean clade support with increasing dataset size (Fig. 4). The curve first showed a very steep increase up to about 40 genes, after which the improvement slowed down. Even at 93 genes, only a 90% mean clade support was achieved, with lower values mostly due to the basal pimpliform nodes (Figs 2 and 3). However, it is unclear whether saturation was reached, which suggests that additional gene sampling could lead to a further increase in tree resolution even under the same taxon sampling and analytical method.

When not sampling genes randomly but instead ranking them according to different criteria, we found a clear improvement for up to about 30 genes under all criteria except the ranking by clocklikeness (Fig. 4). The rankings of the different criteria were very highly correlated (Supplementary File S12), with the highest correlation coefficients between gene length and the number of sites evolving at a near-optimal rate, while clocklikeness showed the weakest correlation with the other criteria. Thus, gene length seems to be the main factor determining differences in single-gene success in our hybrid capture dataset, even though the evolutionary-rate and clade support criteria performed slightly better between 5 and 10–30 genes (Fig. 4).

When comparing the ranking obtained from the full hybrid-capture dataset with the gene rankings obtained from the transcriptome sequence data, we found no significant correlation between gene rankings, even though the bootscore and mean clade support measures were weakly correlated with each other (Spearman: Rho = 0.18, p = 0.08; Supplementary File S12). Ranking genes according to their performance in the transcriptome sequence data also did not lead to any improvement over random gene sampling (results not shown). Besides the obviously different taxon sampling, the two datasets also differed strongly in gene length, as the hybrid capture was only enriching some of the exons, while nearly complete genes were available in the transcriptome sequence dataset.

Discussion

612*New molecular resources for Ichneumonoidea*

613 The bait set for hybrid enrichment proposed here proved highly efficient in enriching
614152 exons of 93 genes, not only across the target group, the ichneumonids, but also in the
615sister family Braconidae and in the genus *Gasteruption*, a member of a distantly related group
616of apocritan wasps (Evanioidea: Gasteruptiidae). Given the small number of genes previously
617used in ichneumonid phylogenetics , this bait set represents a significant step forward. We
618also provide oligonucleotide primers for PCR and Sanger sequencing of eight markers that
619were tested successfully across Ichneumonidae and several outgroups. This set complements a
620recent study very well which suggested oligonucleotide primers for apocritan Hymenoptera,
621but did not include any representatives of Ichneumonoidea . The largest obstacle for
622developing PCR primers is the large density of introns and corresponding short exon lengths
623found in many arthropod taxa. Despite the fact that the divergence between *Nasonia* and
624Ichneumonidae probably dates back to the Triassic , most of the intron-exon boundaries were
625found to be conserved, which might be because introns typically insert at certain sequence
626motifs . This high conservation of intron positions across groups encourages the use of
627distantly related models for intron-exon boundaries for designing phylogenetic markers, be it
628for PCR amplification or in bait design for hybrid captures.

629 Of the eight genes for whose amplification we developed PCR primers, two actually
630included an intron, but one that was found to be short in almost all the taxa we sequenced.
631Preliminary analyses showed that the captured intronic sequences could not be aligned at the
632deep phylogenetic divergences found in our dataset, but they might prove useful in future
633studies as markers for species delimitation. Given previous reports of shortcomings of the
634standard mitochondrial barcode locus cytochrome oxidase subunit 1 (*COI*) in insects and
635especially ichneumonids , additional nuclear markers that are sufficiently variable at the
636species level, such as introns, are sorely needed. However, it remains to be shown how useful
637our intronic markers are at delimiting species.

638*Taxonomic implications of the pimpliform tree*

639 Our combined transcriptome and hybrid capture data represent an important step
640forward from the previous molecular datasets that targeted higher ichneumonid relationships,
641all of which included only a single gene (28S rRNA) and were highly sensitive to alignment
642strategies . Even the addition of morphological data in Quicke *et al*. left many parts of the
643ichneumonid tree inconsistent between different analytical approaches; this might be due to

644high levels of homoplasy found in these parasitoid wasps, which can partly be explained by
645convergent adaptations to the same host groups .

646 Given the sparse taxon sampling in the transcriptome data, which omitted members of
647several pimpliform subfamilies, we focus our discussion here on the main analyses of the
648combined dataset, but mention results from the transcriptome trees where appropriate. Many
649previously suggested higher relationships, most of which were based on morphology, were
650confirmed here. This is case for the placement of Xoridinae , which is unequivocally
651recovered as sister to the remaining ichneumonids here for the first time . The informal
652subfamily groupings Ophioniformes, Ichneumoniformes and Pimpliformes, and the
653monophyly of most subfamilies, were also recovered with good support.

654 With respect to relationships among the Pimpliformes subfamilies, results of our
655amino-acid analyses (Figs 2, 3a-d) are similar to the "best-guess" phylogeny shown in
656Quicke ; which is a summary of the analyses in Quicke *et al.* , in terms of the grouping of
657Poemeniinae, Rhyssinae and Pimplinae (but with the exception of *Xanthopimpla*, see below).
658However, this grouping was not found in the nucleotide analyses (Figs 3c, d), and we
659recovered different (but partly weakly supported) relationships among the remaining
660subfamilies. The positions of Acaenitinae, Collyriinae, and Diplazontinae were highly
661inconsistent among our analyses (e.g., compare Figs 3a and 3e) and thus remain unclear, but
662the low support in this part of each tree precludes rejection of the hypothesis that they all
663branch off close to the base of a group containing the koinobiont, endoparasitoid pimpliform
664subfamilies . In contrast to previous suggestions , Diplazontinae and Orthocentrinae were
665never recovered as sister groups; instead, Diplazontinae were repeatedly resolved as sister to
666the remaining pimpliform subfamilies (Figs 1, 3c, 3e, 3f), even though this placement was
667again not always recovered. Instead of grouping with Diplazontinae, the Orthocentrinae
668formed a clade with Diacritinae and Cylloceriinae in almost all analyses, usually with good to
669even maximum support. A close relationship between Cylloceriinae and Orthocentrinae has
670long been suggested based on morphological and 28S evidence and some authors (e.g.
671Humala, 2007) maintain one subfamily for the cylloceriine and orthocentrine genera. Wahl
672(1990) defined a clade of Koinobiont endoparasitoids of Diptera comprising Cylloceriinae,
673Diplazontinae and Orthocentrinae based mainly on larval morphology: the distinctive fused
674hypostomal-stipital plate, not otherwise found in Ichneumonidae. Given these results and the
675lack of convincing apomorphies in adult morphology, it is possible the the hypostomal-stipital
676plate has evolved more than once in ichneumonid larvae feeding within Diptera larvae.

Diacritinae and Orthocentrinae have also been grouped together previously, especially by Perkins and Townes , who both included *Diacritus* Förster in their Plectiscinae (roughly equivalent to Orthocentrinae), even though Townes did so only "provisionally" and later classified Diacritinae as a separate tribe within an expanded Pimplinae that also included Rhyssinae and Poemeniinae . Morphologically, a close relationship between Orthocentrinae and Diacritinae is not without merit. *Diacritus* shares a fringe of dense setae along the inner side of the apex of the hind tibia with Orthocentrinae (even though it is somewhat less pronounced), and the first metasomal segment, with its elongate shape and fusion between sternite and tergite, is reminiscent of some orthocentrine genera such as *Proclitus* Förster, *Dialipsis* Förster and *Symplecis* Förster. The fact that both Orthocentrinae and Cylloceriinae are parasitoids of Diptera (Supplementary File S6) might provide a clue as to where to look for hosts of Diacritinae, for which no host data are available to date.

Wahl and Gauld suggested that the Acaenitinae consists of a grade of genera previously classified in the tribe Coleocentrini and the monophyletic former Acaenitini, but only referred to an unpublished, preliminary cladistic analysis to support this claim. It is a distinct possibility that the former grade is not only para- but indeed polyphyletic with respect to other pimpliform subfamilies . Morphological similarities between Acaenitinae and other subfamilies have been pointed out in the past, especially with Collyriinae , with which many acaenitine genea share the median tubercle on the clypeus, the propodeum lacking any transverse carinae, and the nervellus of the hind wing intercepted high up. In our analyses, the four genera of the former Acaenitini were always recovered together but almost never sharing a unique common ancestor with the sole representative of the former Coleocentrini included here, *Coleocentrus* (Figs 2 and 3). In the analyses of the transcriptome sequence data, *Coleocentrus* always clustered with *Collyria* (Fig. 1; no other acaenitine genera were included), and this sister-group relationship also appeared in some of the analyses of the full dataset (e.g., Fig. 3d). However, *Coleocentrus* clustered inconsistently with various pimpliform groups, depending on data type and analysis, and support for a separation of the two acaenitine groups was weak in most analyses. Furthermore, the Acaenitinae including *Coleocentrus* were recovered as monophyletic in some of the single-gene trees (Supplementary File S9). Thus, we await an increased taxon sampling of the *Coleocentrus*-group of genera (especially including specimens of *Procinetus* Förster) before suggesting any taxonomic changes.

The subfamily Pimplinae (excluding *Xanthopimpla*) was only recovered as monophyletic in the analyses of the amino-acid alignment (Figs 2, 3a–d) and not those of the nucleotide sequence data (Figs 3e–f). The placement of *Xanthopimpla* might have been obscured by the long subtending branch of our exemplar species (*X. varimaculata* Cameron), an issue that could be remedied in the future by a more extensive sampling of this species-rich genus. *Xanthopimpla* shares at least two unique characters with *Lissopimpla* Kriechbaumer of the Pimplini, namely the transversely divided clypeus and (with some exceptions in *Xanthopimpla*) a transverse carina at the anterior end of the notauli (unfortunately, no representatives of *Lissopimpla* were included in our study). Contrary to various authors , *Echthromorpha* Holmgren does not share the transversely divided clypeus and its mandibles, although narrowed with the lower tooth very short, are not twisted as in *Lissopimpla* and *Xanthopimpla*. *Xanthopimpla* also seems to be among the oldest extant pimpline genera when considering the fossil record , which could support an early branching of this genus. Together with *Lissopimpla*, *Xanthopimpla* might have to be included in a new tribe or even subfamily in the future, but we await an analysis with more extensive taxon sampling before formalizing such a change. As for the relationships among the other Pimplinae genera, it is interesting that the results recovered here are strongly reminiscent of the tribal classification suggested by Townes : he combined today's Delomeristini with the *Theronia*-group of genera and *Pseudorhyssa* in the tribe Theroniini, but expressed uncertainty in both cases. Later cladistic analyses based on morphological characters moved *Pseudorhyssa* to Poemeniinae and the *Theronia*-group into Pimplini . Our analyses provide support for Townes's classification, which was largely intuitive. The grouping of *Pseudorhyssa* in Delomeristini was recovered with maximum support in all analyses of the full dataset (Fig. 2) and was consistent even across single-gene trees (Supplementary File S11); we therefore move this genus to the tribe Delomeristini (Pimplinae), with the result that the tribe Pseudorhyssini Wahl & Gauld, 1998, becomes a junior synonym of Delomeristini Hellén, 1919; **syn nov**. We note that the continuation of the epomia along the ventral edge of the pronotum in *Pseudorhyssa* is similar to that found in *Delomerista* Förster and *Atractogaster* Kriechbaumer (the epomia is much shorter in *Perithous*, the other genus of Delomeristini) and dissimilar to the more pronounced bulging of the pronotum along the carina in Poemeniinae. Additionally, *Pseudorhyssa* has a weak, ventrally placed glymma on the first metasomal tergite in common with other Delomeristini and unlike Poemeniinae. The placement of the *Theronia* genus-group with Delomeristini was not recovered in a majority of gene trees, and the current placement within Pimplini also occurred in some gene trees. However, the group was consistently recovered as

743monophyletic here and is well-defined on morphological grounds ; thus, we resurrect the
744Theroniini as a tribe for the *Theronia* genus-group (stat. rev.), as was advocated by Carlson
745(1979).

746 Our results clearly placed *Hemiphanes* in the Ichneumoniformes-group of subfamilies,
747both in the concatenated analyses and in the vast majority of the gene trees but, on the basis of
748our taxon sampling, did not firmly suggest a subfamily placement. We here transfer
749*Hemiphanes* to Cryptinae. Whilst the morphology of *Hemiphanes* does not present any
750unambiguous synapomorphies with Cryptinae, it is a better fit than Orthocentrinae, as is borne
751out by our molecular results. Characters in support of this placement include the fused first
752metasomal sternite and tergite, absence of a glymma, anteriorly narrow and posteriorly much
753wider first tergite, and strongly sclerotized remaining tergites. There is a distinct sternaulus on
754the anterior 0.3 – 0.4 of the mesopleuron, hind wing vein M+Cu is straight and the nervellus
755(distal abscissa of Cu) is strong. Furthermore, fore wing vein 2m-cu has one bulla, the
756ovipositor has a nodus, and the male aedeagus is roughly ‘drop-shaped’ distally, not dorso-
757ventrally flattened as in Pimpliformes. In several respects, including the shape of the clypeus,
758large hypopygium, short first metasomal sternite, spiracle of the first tergite anterior to the
759mid-length, wing venation and general structure, *Hemiphanes* is surprisingly similar to
760*Sphecophaga* Westwood, a cryptine parasitoid of eusocial vespine wasps; the two key out to
761the same couplet in a recent key to ichneumonid subfamilies . Features such as the twisted
762mandibles and fringe of setae on the inner aspect of the hind tibia do not strongly support the
763placement of *Hemiphanes* in the Orthocentrinae, as the mandibles are more robust and fringe
764less dense than in most orthocentrines. The sum of characters suggest that *Hemiphanes*,
765particularly the wing venation and reduced propodeal carinae, suggest that *Hemiphanes*
766belongs in Cryptinae rather than Phygadeuontinae, and the pronounced, triangular expansions
767of the metanotum are a feature of the tribe Aptesini rather than Cryptini, so we place
768*Hemiphanes* in Aptesini, despite its overall similarity with *Sphecophaga*.

769*Rapid radiation of Pimpliformes*

770 Rapid radiations have been suggested in several groups of parasitoid wasps and are
771mostly attributed to the formation of new ecological niches through host switches. A rapid
772radiation after host switching is possibly the case for the microgastroid subfamilies of the
773Braconidae, which changed to lepidopteran hosts and at the same time acquired poly-DNA
774viruses to overcome their hosts' immune systems . However, previous studies inferring rapid
775radiations were based on the analysis of small sets of genes, usually no more than three or

four, which could explain the low resolution in the phylogenetic inferences. In Chalcidoidea, a recent transcriptome study based on more than 3,200 genes suggested a rapid radiation of some of its families during the late Cretaceous, although the inferred branch lengths of their dated tree still mostly span about 5–10 million years and are, thus, not that rapid after all . Our study might therefore be viewed as the most convincing example to date of a rapid radiation in parasitoid wasps, supported by both considerable gene sampling and very short inferred branch lengths (Fig. 2). As a very rough estimate, assuming the most recent common ancestor of Pimpliformes existed about 66 million years ago (see below) and using average branch lengths as clock lengths, the shortest branches at the base of the radiation would be less than one million years long; but this result should be confirmed by a properly calibrated molecular clock analysis.

Our ancestral-state reconstructions inferred at least one switch between idiobiosis and koinobiosis during the most rapid phase of the radiation (Fig. 3). In addition, Diptera were acquired as hosts by the subfamilies Diplazontinae, Orthocentrinae, and Cylloceriinae (and possibly Diacritinae), according to our trees probably in two separate events (but see previous section where we discuss that we cannot exclude that they are indeed monophyletic, which would imply a single switch). Diptera are rather rare as hosts for Ichneumonidae, and the effect of this host switch on diversification rates and patterns deserves further study. The same is true for the switches between parasitoid strategies, including between ecto- and endoparasitoidism. We here refrain from a detailed discussion of the implications of our results, given the inconsistency between analyses (Fig. 3). A better-resolved tree is needed to address such questions.

The timing of the rapid radiation of the pimpliform subfamilies remains unclear as no dated tree for the group is yet available. However, the fossil record indicates that it might have taken place around or just after the Cretaceous-Paleogene boundary. Fossil ichneumonids from the Cretaceous are all classified in now extinct subfamilies , with the exception of a species from Canadian amber that was tentatively placed in Labeninae , a group consistently branching earlier than Pimpliformes in our analyses. On the other hand, several clear representatives of the subfamilies Acaenitinae, Orthocentrinae, Pimplinae and Rhyssinae are known from the upper and middle Eocene . If a dated phylogenetic analysis confirms the timing of the pimpliform radiation as directly after the K-Pg boundary, this group's history can be compared to other rapid radiations that have presumably taken place around the same time, such as that of Neoaves , placental mammals and legumes .

Do we need more genes, more taxa, or better methods?

Rapid radiations that happened in the deep past constitute some of the most difficult phylogenetic problems. The low probability of any substitution happening along the short basal branches, combined with the risk that such true signal is masked by subsequent changes along the typically long branches leading from there, might mean that very large amounts of data need to be generated in order to observe sufficient signal for resolving deep, rapid radiations . Biases in the data arising from model misspecification, non-random patterns of missing data, or among-lineage rate variation and associated effects such as long-branch attraction, might easily obscure any true phylogenetic signal from such radiations . Consequently, several studies targeting old radiations using genome-scale datasets have failed to completely resolve the underlying species tree , and others retrieved conflicting hypotheses depending on dataset or analytical methods .

Is it possible to distinguish between a lack of signal due to an extremely rapid radiation (or in other words, a truly hard polytomy), insufficient gene or taxon sampling, and systematic biases in a dataset? Analyses of subsets of a dataset can, to a certain extent, exclude under-sampling as a major disruptive factor. Our phylogenetic inferences based on gene sets of different sizes indicate a slow-down in the increase of mean clade support with increasing gene sampling and indications of a levelling-off at an average bootstrap per node of about 0.9, even though it is not entirely clear if a further increase might still be possible if more than 93 genes were sampled (Fig. 4).

It also remains unclear whether our approach to marker choice has in fact improved the inference of the phylogeny, given that neither of the criteria clocklikeness, number of sites at near-optimal rates, or single-gene bootstrap support clearly outperformed mere gene length in the analyses of subsets of genes (Fig. 4). However, it is possible that the initial choice of markers from the transcriptome gene-set already eliminated most of the inferior genes and, thus, left us with insufficient power to detect any remaining effect of these predictors. The efficiency of some of the criteria we applied are also not beyond doubt. For instance, clade support values for certain clades might be high despite very low resolution in other parts of the tree, especially if there is a difference in node depth . Using sums of clade support might therefore favour genes that resolve the more numerous, shallow nodes well, while not accounting sufficiently for the deeper nodes. Furthermore, we used a combined approach with the sequential application of several different criteria for choosing genes, and the relative performance and combinability of these criteria still needs to be evaluated .

Tests for the amount and extent of taxon sampling that is necessary to resolve a deep radiation are even more difficult to perform, as the addition of taxa always changes the nature of the phylogenetic problem. In our case, the trees derived from the 723-gene transcriptome dataset of 26 taxa was fully resolved in terms of maximum bootstrap support in all analyses, while the combined hybrid capture data of 93 genes of 84 taxa exposed the weakly-resolved rapid radiation at the base of Pimpliformes. The larger dataset included the additional subfamilies Cylloceriinae, Diacritinae, and Orthocentrinae, and several additional tribes. Thus, it included additional nodes which were difficult to resolve. The most striking topological difference between the trees was the placement of Diplazontinae: breaking down the branch leading to *Syrphophilus* Dasch, the only diplazontine included in the transcriptome dataset, obviously had the effect of moving this subfamily further up the pimpliform tree (compare Figs 2 and 3). Consideration of phylogenetic informativeness led to the conclusion that targeted taxon sampling should focus on species that would branch off as closely to the unresolved nodes as possible . In our case, a promising approach would be to sample multiple genera per subfamily, such as in Diacritinae and Collyriinae, but also focusing on more basally-branching members of other subfamilies such as Acaenitinae. While such an approach of course cannot increase the length of the most basal branches of the pimpliform radiation, it breaks down their long subtending branches and can therefore ameliorate long-branch attraction phenomena that could obscure the true phylogenetic signal.

Systematic biases such as nucleotide composition biases and other mismatches of the substitution model can be exposed by analytical methods, for example, the test for uneven nucleotide composition performed here , but also by more advanced methods falling into the class of model adequacy testing . In most cases, such biases may be datatype-specific and will thus translate into conflicts between different dataset types, such as amino acids versus nucleotides or all three versus the first two codon positions. Such conflicts were apparent especially in our transcriptome sequence analyses, while the complete data typically exhibited low support at those nodes that were in conflict between analyses. The absence of supported conflict between data types might indicate that substitution-model misspecification was not the major factor preventing resolution of the pimpliform radiation. Also, in contrast to previous analyses that recovered strongly supported conflict among the single-gene trees , those here were mostly unresolved at the base of Pimpliformes, which indicates that gene-tree/species-tree analyses would not yield improved resolution . Nevertheless, better evolutionary models might facilitate the extraction of what little signal is left in sequence data from the early and rapid pimpliform radiation. Another alternative is the analysis of different

data types that show comparatively low levels of homoplasy, for instance certain ecological or morphological characters or characters pertaining to genome morphology .

In conclusion, we postulate here one of the most rapid radiations uncovered to date in parasitoid wasps and speculate that the most promising approach to improve its resolution probably lies in increased taxon sampling, even though the sampling of additional markers or even other data types remains an alternative. The molecular resources and analysis pipeline we present here represent a significant step forward in phylogenetics of ichneumonids and beyond and will in the future hopefully allow a more robust inference of the relationships among these wasps and of the evolutionary history of their varied parasitoid ecologies.

Acknowledgments

We are very grateful to everyone who provided rare specimens for this study: Eric Chapman (University of Kentucky, U.S.A.), the Swedish Malaise Trap Project (www.stationlinne.se, Sweden), Masato Ito (Kobe University, Japan), Ilari Sääksjärvi (University of Turku, Finland), Stanislav Korenko (Czech University of Life Sciences Prague), Philippe Louâpre (Université de Bourgogne, France), and Marc Neumann (Walterswil SO, Switzerland). Muriel Fragnière (University of Bern, Switzerland) helped with the planning of the transcriptome part of this study, prepared the RNA libraries and conducted the transcriptome sequencing. We also thank Kathy Saint and Tessa Bradford (University of Adelaide, Australia) for helpful advice on the hybrid-enrichment lab work. Hugh Robertson (University of Illinois, U.S.A.) provided us with access to the genome of *Microplitis demolitor* before its release for assessing intron-exon boundaries for bait design.

This study was funded by the Swiss National Science Foundation (grant PZ00P3_154791 to SK).

References

Andrews S. (2014) FastQC version 0.11.2. Cambridge, Babraham Institute.

Askew R.R., Shaw M.R. (1986) Parasitoid communities: Their size, structure and development. In: Waage JK, Greathead D editors. Insect Parasitoids. London, Academic Press, p. 225–264.

Austin A.D. (1985) The function of spider egg sacs in relation to parasitoids and predators, with special reference to the Australian fauna. *Journal of Natural History*, **19**, 359–376.

Belshaw R., Fitton M.G., Herniou E., Gimeno C., Quicke D.L.J. (1998) A phylogenetic reconstruction of the Ichneumonoidea (Hymenoptera) based on the D2 variable region of 28S rivosomal RNA. *Systematic Entomology*, **23**, 109–123.

Bergsten J., Bilton D.T., Fujisawa T., Elliott M., Monaghan M.T., Balke M., Hendrich L., Geijer J., Herrmann J., Foster G.N., Ribera I., Nilsson A.N., Barraclough T.G., Vogler A.P. (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, **61**, 851–869.

Bolger A.M., Lohse M., Usadel B. (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170. *Bioinformatics*, **30**, 2114–2120.

Bragg J.G., Potter S., Bi K., Moritz C. (2016) Exon capture phylogenomics: efficacy across scales of divergence. *Molecular Ecology Resources*, **16**, 1059–1068.

Broad G.R., Shaw M.R., Fitton M.G. (2018) *The ichneumonid wasps of Britain and Ireland (Hymenoptera: Ichneumonidae)*. Royal Entomological Society, Telford.

Brown J.M. (2014) Detection of implausible phylogentic inferences using posterior predictive assessment of model fit. *Systematic Biology*, **63**, 334–348.

Carlson, R.W. (1979) Family Ichneumonidae. In: Krombein, K.V., Hurt, P.D., Smith, D.R., Burks, B.D., editors. Catalog of Hymenoptera in America north of Mexico. Washington. Pp. 315-740.

Collins T.M., Fedrigo O., Naylor G.J.P. (2005) Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. *Systematic Biology*, **54**, 493–500.

Dibb N.J., Newman A.J. (1989) Evidence that introns arose at proto-splice sites. *The EMBO Journal*, **8**, 2015–2021.

dos Reis M., Inoue J., Hasegawa M., Asher R.J., Donoghue P.C.J., Yang Z. (2012) Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **279**, 3491–3500.

Dowton M., Cameron S., Dowavic J.I., Austin A.D., Whiting M.F. (2009) Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. *Molecular Biology and Evolution*, **26**, 1607–1617.

Doyle V.P., Young R.E., Naylor G.J.P., Brown J.M. (2015) Can we identify genes with increased phylogenetic reliability? *Systematic Biology*, **64**, 824–837.

Duchêne D., Duchêne S., Ho S.Y.W. (2018) PhyloMAd: Efficient assessment of phylogenomic model adequacy. *Bioinformatics*, **bty103**.

Edwards S.V., Liu L., Pearl D.K. (2007) High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5936–5941.

Eggleton P. (1989) The phylogeny and evolutionary biology of the Pimplinae (Hymenoptera: Ichneumonidae). Department of Pure and Applied Biology. London, Imperial College, p. 295.

Ericson P.G.P., Klopfstein S., Irestedt M., Nguyen J.M.T., Nylander J.A.A. (2014) Dating the diversification of the major lineages of Passeriformes (Aves). *BMC Evolutionary Biology*, **14**, 8.

Fischer M., Steel M.A. (2009) Sequence length bounds for resolving a deep phylogenetic divergence. *Journal of Theoretical Biology*, **256**, 247–252.

Gatesy J., DeSalle R., Wahlberg N. (2007) How many genes should a systematist sample? Conflicting insights from a phylogenomic matrix characterized by replicated incongruence. *Systematic Biology*, **56**, 355–363.

Gauld I.D. (1988) Evolutionary patterns of host utilization by ichneumonid parasitoids (Hymenoptera: Ichneumonidae and Braconidae). *Biological Journal of the Linnean Society*, **35**, 351–377.

Gauld I.D. (1991) The Ichneumonidae of Costa Rica, 1. *Memoirs of the American Entomological Institute*, **47**, 1–589.

Gauld I.D., Dubois J. (2006) Phylogeny of the *Polysphincta* group of genera (Hymenoptera: Ichneumonidae; Pimplinae): a taxonomic revision of spider ectoparasitoids. *Systematic Entomology*, **31**, 529-564.

Gauld I.D., Mound L.A. (1982) Homoplasy and the delineation of holophyletic genera in some insect groups. *Systematic Entomology*, **7**, 73–86.

Gauld I.D., Sithole R., Gómez J.U., Godoy C. (2002a) The Ichneumonidae of Costa Rica, 4. *Memoirs of the American Entomological Institute*, **66**, 1–768.
Gauld I.D., Wahl D.B., Broad G.R. (2002b) The suprageneric groups of the Pimplinae (Hymenoptera: Ichneumonidae): a cladistic re-evaluation and evolutionary biological study. *Zoological Journal of the Linnean Society*, **136**, 421–485.
Gauld I.D., Ward S., Mallet V. (2000) The Ichneumonidae of Costa Rica, 3. Introduction and keys to species of the subfamilies Brachyicyrtinae, Cremastinae, Labeninae and Oxytorinae, with an appendix on the Anomaloninae. *Memoirs of the American Entomological Institute*, **63**, 1–453.
Geuten K., Massingham T., Darius P., Smets E., Goldman N. (2007) Experimental design criteria in phylogenetics: where to add taxa. *Systematic Biology*, **56**, 609–622.
Glenn T.C., Nilsen R., Kieran T.J., Finger J.W., Pierson T.W., Bentley K.E., Hoffberg S., Louha S., Garcia-De-Leon F.J., del Rio Portilla M.A., Reed K., Anderson J.L., Meece J.K., Aggery S., Rekaya R., Alabady M., Belanger M., Winker K., Faircloth B.C. (2016) Adapterama I: Universal stubs and primers for thousands of dual-indexed Illimina libraries (iTru & iNext).
Goldman N. (1998) Phylogenetic information and experimental design in molecular systematics. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **265**, 1779–1786.
Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–654.
Hartig G., Peters R.S., Borner J., Etzbauer C., Misof B., Niehuis O. (2012) Oligonucleotide primers for targeted amplification of single-copy nuclear genes in Apocritan Hymenoptera. *PLoS One*, **7**, e39826.
Hasegawa M., Kishino H., Yano T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.
Heath T.A., Hedtke S.M., Hillis D.M. (2008) Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, **46**, 239–257.
Hedtke S.M., Townsend T.M., Hillis D.M. (2006) Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Systematic Biology*, **55**, 522–529.
Henricson A., Forslund K., Sonnhammer E.L.L. (2010) Orthology confers intron position conservation. *BMC Genomics*, **11**, 412.
Huang X., Madan A. (1999) CAP3: A DNA Sequence Assembly Program. *Genome Research*, **9**, 868–877.
Hugall A., O'Hara T., Hunjan S., Nilsen R., Moussali A. (2015) An exon-capture system for the entire class Ophiuroidea. *Molecular Biology and Evolution*, **33**, 281–294.
Humala, A.E. (2007) To the system of the subfamily Orthocentrinae s. l. (Hymenoptera, Ichneumonidae). In: Rasnitsyn, A.P., Gokhman, V.E., editors. [Studies on hymenopterous insects. Collection of scientific papers.] (In Russian with English summaries). Moscow: KMK Scientific Press Ltd. Pp. 76-85.
i5K Consortium. (2013) The i5K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity*, **104**, 595–600.
IUCN. (2017) IUCN Red List of Threatened Species, version 2017.3. Gland, Switzerland, IUCN.
Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Li J.W., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldon T., Capella-Gutierrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X.J., Dixon A., Li S.B., Li N., Huang Y.H., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C.V., Lovell P.V., Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Velazquez A.M.V., Alfaro-Nunez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z.J., Zeng Y.L., Liu S.P., Li Z.Y., Liu B.H., Wu K., Xiao J., Yinqi X., Zheng Q.M., Zhang Y., Yang H.M., Wang J., Smeds L., Rheindt F.E., Braun M.,

Fjeldsa J., Orlando L., Barker F.K., Jonsson K.A., Johnson W., Koepfli K.P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alstrom P., Edwards S.V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T.P., Zhang G.J. (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.
Katoh K., Standley D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.
Klopfstein S. (2014) Revision of the Western Palaearctic Diplazontinae (Hymenoptera, Ichneumonidae). *Zootaxa*, **3801**, 1–143.
Klopfstein S., Kropf C., Baur H. (2016) *Wolbachia* endosymbionts distort DNA barcoding in the parasitoid wasp genus *Diplazon* (Hymenoptera: Ichneumonidae). *Zoological Journal of the Linnean Society*, **177**, 541–557.
Klopfstein S., Kropf C., Quicke D.L.J. (2010) An evaluation of phylogenetic informativeness profiles and the molecular phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae). *Systematic Biology*, **59**, 226–241.
Klopfstein S., Massingham T., Goldman N. (2017) More on the best evolutionary rate for phylogenetic analysis. *Systematic Biology*, **66**, 769–785.
Klopfstein S., Quicke D.L.J., Kropf C., Frick H. (2011) Molecular and morphological phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae). *Zoologica Scripta*, **40**, 379–402.
Klopfstein S., Ronquist F. (2013) Convergent intron gains in hymenopteran elongation factor-1α. *Molecular Phylogenetics and Evolution*, **67**, 266–276.
Koenen E.J.M., de Vos J.M., Atchison G.W., Simon M.F., Schrire B.D., de Souza E.R., de Queiroz L.P., Hughes C.E. (2013) Exploring the tempo of species diversification in legumes. *South African Journal of Botany*, **89**, 19–30.
Kopylov D.S. (2009) A new subfamily of ichneumons from the lower Cretaceous of Transbaikalia and Mongolia (Insecta: Hymenoptera: Ichneumonidae). (in Russian). *Paleontologicheskii Zhurnal*, **1**, 76–85.
Kopylov D.S. (2010) A new subfamily of ichneumon wasps (Insecta: Hymenoptera: Ichneumonidae) from the Upper Cretaceous of the Russian Far East. *Paleontological Journal*, **4**, 59–69.
Kopylov D.S. (2012) New species of Praeichneumonidae (Hymenoptera, Ichneumonoidea) from the Lower Cretaceous of Transbaikalia. *Paleontological Journal*, **46**, 66–72.
Ksepka D.T., Phillips M.J. (2015) Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. *Annals of the Missouri Botanical Garden*, **100**, 300–328.
Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. (2016) 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.
Lepage T., Bryant D., Philippe H., Lartillot N. (2007) A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution*, **24**, 2669–2680.
Liu L., Pearl D.K., Brumfield R.T., Edwards S.V. (2008) Estimating species trees using multiple-allele DNA sequence data. *Evolution*, **62**, 2080–2091.
Louca S., Doebeli M. (2018) Efficient comparative phylogenetics on large trees. *Bioinformatics*, **34**, 1053–1055.
Mamanova L., Coffey A.J., Scott C.E., Kozarewa I., Turner E.H., Kumar A., Howard E., Shendure J., Turner D.J. (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods*, **7**, 111–118.
Martin M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, **17**, 10–12.
Mayer C., Sann M., Donath A., Meixner M., Podsiadlowski L., Peters R.S., Petersen M., Meusemann K., Liere K., Waegele J.W., Misof B., Bleidorn C., Ohl M., Niehuis O. (2016) BaitFisher: a software package ofr multispecies target DNA enrichment probe design. *Molecular Biology and Evolution*, **33**, 1875–1886.

McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T. (2013a) A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One*, **8**, e54858.
McCormack J.E., Hird S.M., Zellmer A.J., Carstens B.C., Brumfield R.T. (2013b) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.
McKellar R.C., Kopylov D.S., Engel M.S. (2013) Ichneumonidae (Insecta: Hymenoptera) in Canadian Late Cretaceous amber. *Fossil Record*, **16**, 217–227.
Meusemann K., von Reumont B.M., Simon S., Roeding F., Strauss S., Kück P., Ebersberger I., Walzl M., Pass G., Breuers S., Achter V., Von Haeseler A., Burmester T., Hadrys H., Waegele J.W., Misof B. (2010) A phylogenomic approach to resolve the arthropod tree of life. *Molecular Biology and Evolution*, **27**, 2451–2464.
Meyer M., Kircher M. (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **2010**, 1–20.
Misof B., Liu S.L., Meusemann K., Peters R.S., Donath A., Mayer C., Frandsen P.B., Ware J., Flouri T., Beutel R.G., Niehuis O., Petersen M., Izquierdo-Carrasco F., Wappler T., Rust J., Aberer A.J., Aspock U., Aspock H., Bartel D., Blanke A., Berger S., Bohm A., Buckley T.R., Calcott B., Chen J.Q., Friedrich F., Fukui M., Fujita M., Greve C., Grobe P., Gu S.C., Huang Y., Jermiin L.S., Kawahara A.Y., Krogmann L., Kubiak M., Lanfear R., Letsch H., Li Y.Y., Li Z.Y., Li J.G., Lu H.R., Machida R., Mashimo Y., Kapli P., McKenna D.D., Meng G.L., Nakagaki Y., Navarrete-Heredia J.L., Ott M., Ou Y.X., Pass G., Podsiadlowski L., Pohl H., von Reumont B.M., Schutte K., Sekiya K., Shimizu S., Slipinski A., Stamatakis A., Song W.H., Su X., Szucsich N.U., Tan M.H., Tan X.M., Tang M., Tang J.B., Timelthaler G., Tomizuka S., Trautwein M., Tong X.L., Uchifune T., Walzl M.G., Wiegmann B.M., Wilbrandt J., Wipfler B., Wong T.K.F., Wu Q., Wu G.X., Xie Y.L., Yang S.Z., Yang Q., Yeates D.K., Yoshizawa K., Zhang Q., Zhang R., Zhang W.W., Zhang Y.H., Zhao J., Zhou C.R., Zhou L.L., Ziesmann T., Zou S.J., Li Y.R., Xu X., Zhang Y., Yang H.M., Wang J., Wang J., Kjer K.M., Zhou X. (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, **346**, 763–767.
Murphy N.P., Banks J.C., Whitfield J.B., Austin A.D. (2008) Phylogeny of the parasitic microgastroid subfamilies (Hymenoptera: Braconidae) based on sequence data from seven genes, with an improved time estimate of the origin of the lineage. *Molecular Phylogenetics and Evolution*, **47**, 378–395.
Murphy N.P., Carey D., Castro L.R., Dowton M., Austin A.D. (2007) Phylogeny of the platygastroid wasps (Hymenoptera) based on sequences from the 18S rRNA, 28S rRNA and cytochrome oxidase I genes: implications for the evolution of the ovipositor system and host relationships. *Biological Journal of the Linnean Society*, **91**, 653–669.
Narendran T.C. (2001) Parasitic Hymenoptera and biological control. In: Upadhyay RK, Mukerji KG, Chamola BP editors. Biocontrol potential and its exploitation in sustainable agriculture. New York, Kluwer Academic / Plenum Publishers.
Nicholls J.A., Challis R.J., Mutun S., Stone G.N. (2012) Mitochondrial barcodes are diagnostic of shared refugia but not species in hybridiying oak gallwasps. *Molecular Ecology*, **21**, 4051–4062.
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.
Olsen G.L. (1996) DNArates - maximum likelihood site rate calculation. Published by the author.
Paradis E., Claude J., Strimmer K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
Perkins J.F. (1940) Notes on the synonymy of some genera of European Pimplinae (s.l.) (Hym. Ichneumonidae). *The Entomologist*, **73**, 54–56.
Peters R.S., Krogmann L., Mayer C., Donath A., Gunkel S., Meusemann K., Kozlov A., Podsiadlowski L., Petersen M., Lanfear R., Diez P.A., Heraty J.M., Kjer K.M., Klopfstein S., Meier R., Polidori C., Schmitt T., liu S.L., Zhou X., Wappler T., Rust J., Misof B., Niehuis O. (2017) Evolutionary history of the Hymenoptera. *Current Biology*, **27**, 1013–1018.

Peters R.S., Niehuis O., Gunkel S., Bläser M., Mayer C., Podsiadlowski L., Kozlov A., Donath A., Van Noort S., Liu S., Zhou X., Misof B., Heraty J.M., Krogmann L. (2018) Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea) reveals a history of rapid radiations, convergence, and evolutionary success. *Molecular Phylogenetics and Evolution*, **120**, 286–296.

Petersen M., Meusemann K., Donath A., Dowling D., Liu S., Peters R.S., Podsiadlowski L., Vasilikopoulos A., Zhou X., Misof B., Niehuis O. (2017) Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinformatics*, **18**, 111.

Pham N.T., Broad G.R., Waegele J.W. (2013) A review of the *Theronia* genus-group (Hymenoptera: Ichneumonidae: Pimplinae) from Vietnam with descriptions of five new species. *Journal of Natural History*, **47**, 1501–1538.

Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G., Baurain D. (2011) Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*, **9**, e1000602.

Phillips M.J., Delsuc F., Penny D. (2004) Genome-scale phylogeny and the detection of systematic biases. *Molecular Biology and Evolution*, **21**, 1455–1458.

Prum R.O., Berv J.S., Dornburg A., Field D.J., Townsend J.P., Moriarty Lemmon E., Lemmon A.R. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, **526**, 569–573.

Quicke D.L.J. (2015) Phylogeny and systematics of the Ichneumonidae. In: Quicke DLJ editor. The braconid and ichneumonid parasitoid wasps: biology, systematics, evolution and ecology. Chichester, UK, John Wiley & Sons, Ltd.

Quicke D.L.J., Fitton M.G., Notton D.G., Belshaw R., Broad G.R., Dolphin K. (2000) Phylogeny of the Ichneumonidae (Hymenoptera). A simultaneous molecular and morphological analysis. In: Austin AD, Dowton M editors. Hymenoptera: Evolution, Biodiversity and Biological Control. Canberra, Australia, CSIRO, p. 74–83.

Quicke D.L.J., Laurenne N.M., Fitton M.G., Broad G.R. (2009) A thousand and one wasps: a 28S rDNA and morphological phylogeny of the Ichneumonidae (Insecta: Hymenoptera) with an investigation into alignment parameter space and elision. *Journal of Natural History*, **43**, 1305–1421.

Quicke D.L.J., Lopez-Vaamonde C., Belshaw R. (1999) The basal Ichneumonidae (Insecta, Hymenoptera): 28S D2 rDNA considerations of the Brachycyrtinae, Labeninae, Paxylommatinae and Xoridinae. *Zoologica Scripta*, **28**, 203-210.

R Core Team. (2014) R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

Rago A., Gilbert D.G., Choi J.H., Sackton T.B., Wang X., Kelkar Y.D., Werren J.H., Colbourne J.K. (2016) OGS2: genome re-annotation of the jewel wasp *Nasonia vitripennis*. *BMC Genomics*, **17**, 678.

Reddy S., Kimball R.T., Pandey A., Hosner P.A., Braun M.J., Hackett S.J., Han K.-L., Harshman J., Huddleston C.J., Kingston S., Marks B.D., Miglia K.J.M., W.S., Sheldon F.H., Witt C.C., Yuri T., Braun E.L. (2017) Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Systematic Biology*, **66**, 857–879.

Regier J.C., Shultz J.W., Ganley A.R.D., Hussey A., Shi D., Ball B., Zwick A., Stajich J.E., Cummings M.P., Martin J.W., Cunningham C.W. (2008) Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Systematic Biology*, **57**, 920–938.

Rodríguez-Ezpeleta N., Brinkmann H., Roure B., Lartillot N., Lang B.F., Philippe H. (2007) Detecting and overcoming systematic errors in genome-scale phylogenies. *Systematic Biology*, **56**, 389–399.

Rokas A., Williams B.L., King N., Carroll S.B. (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, **425**, 798–804.

Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. (2012a) A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, **61**, 973–999.

Ronquist F., Teslenko M., Van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. (2012b) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.
Rousse P., Quicke D.L.J., Matthee C.A., Lefeuvre P., Van Noort S. (2016) A molecular and morphological reassessment of the phylogeny of the subfamily Ophioninae (Hymenoptera: Ichneumonidae). *Zoological Journal of the Linnean Society*, **178**, 128–148.
Salichos L., Rokas A. (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, **497**, 327–331.
Santos B. (2017) Phylogeny and reclassification of Cryptini (Hymenoptera, Ichneumonidae, Cryptinae), with implications for ichneumonid higher-level classification. *Systematic Entomology*, **42**, 650–676.
Sharma P.P., Kaluziak S.T., Pérez-Porro A.R., Gonzáles V.L., Hormiga G., Wheeler W.C., Giribet G. (2014) Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution*, **31**, 2963–2984.
Sheng M.-L., Broad G.R., Sun S.-P. (2012) A new genus and species of Collyriinae (Hymenoptera, Ichneumonidae). *Journal of Hymenoptera Research*, **23**, 103–125.
Slater G.S., Birney E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **6**, 31.
Song S., Liu L., Edwards S.V., Wu S. (2012) Resolving conflict in the eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 14942–14947.
Spasojevic T., Broad G.R., Bennett A.M.R., Klopfstein S. (2018a) Ichneumonid parasitoid wasps from the Early Eocene Green River Formation: five new species and a revision of the known fauna (Hymenoptera, Ichneumonidae). *Paläontologische Zeitschrift*, **92**, 35–63.
Spasojevic T., Wedmann S., Klopfstein S. (2018b) Seven remarkable new fossil species of parasitoid wasps (Hymenoptera, Ichneumonidae) from the Eocene Messel Pit. *PLoS One*, **13**, e0197477.
Stamatakis A. (2006) Phylogenetic models of rate heterogeneity: A high performance computing perspective. Proceedings of IPDPS2006. Rhodos, Greece.
Stamatakis A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.
Streicher J.W., Schulte J.A.I., Wiens J.J. (2016) How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Systematic Biology*, **65**, 128–145.
Strimmer K., von Haeseler A. (1997) Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 6815–6819.
Suyama M., Torrents D., Bork P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**, W609–W612.
Swofford D.L. (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Townes H.K. (1957) A revision of the genera of Poemeniini and Xoridini (Hymenoptera, Ichneumonidae). *Proceedings of the Entomological Society of Washington*, **59**, 15–23.
Townes H.K. (1969) The genera of Ichneumonidae, Part 1. *Memoirs of the American Entomological Institute*, **11**, 1–300.
Townes H.K. (1971) The genera of Ichneumonidae, Part 4. *Memoirs of the American Entomological Institute*, **17**, 1-372.
Townsend J.P. (2007) Profiling phylogenetic informativeness. *Systematic Biology*, **56**, 222–231.
Townsend J.P., Su Z., Tekle Y.I. (2012) Phylogenetic signal and noise: Predicting the power of a dataset to resolve phylogeny. *Systematic Biology*, **61**, 835–849.
Virgilio M., Backeljau T., Nevado B., De Meyer M. (2010) Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics*, **11**, 206.

Waegele J.W., Mayer C. (2007) Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. *BMC Evolutionary Biology*, **7**, No. 147.
Wahl D.B. (1986) Larval structures of oxytorines and their significance for the higher classification of some Ichneumonidae (Hymenoptera). *Systematic Entomology*, **11**, 117–127.
Wahl D.B. (1990) A review of the mature larvae of Diplazontinae, with notes on larvae of Acaenitinae and Orthocentrinae and proposal of two new subfamilies (Insecta: Hymenoptera, Ichneumonidae). *Journal of Natural History*, **24**, 27–52.
Wahl D.B., Gauld I.D. (1998) The cladistics and higher classification of the Pimpliformes (Hymenoptera : Ichneumonidae). *Systematic Entomology*, **23**, 265–298.
Waterhouse R.M., Tegenfeldt F., Li J., Zdobnov E.M., Kriventseva E.V. (2013) OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Research*, **41**, D358–D365.
Werren J.H., Richards S., Desjardins C.A., Niehuis O., Gadau J., Colbourne J.K., Group T.N.G.W. (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science*, **327**, 334–348.
Whitfield J.B., Austin A.D., Fernandez-Triana J. (2018) Systematics, biology, and evolution of microgastrine parasitoid wasps. *Annual Review of Entomology*, **63**, 389–406.
Yang Z. (1996) Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology & Evolution*, **11**, 367–372.
Yang Z. (1998) On the best evolutionary rate for phylogenetic analysis. *Systematic Biology*, **47**, 125–133.
Yu D.S., Van Achterberg C., Horstmann K. (2016) Taxapad 2016. Ichneumonoidea 2015 (Biological and taxonomical information), Taxapad Interactive Catalogue Database on flash-drive. Nepean, Ottawa, Canada, www.taxapad.com.
Zaldivar-Riverón A., Shaw M.R., Sáez A.G., Mori M., Belokobylskij S.A., Shaw S.R., Quicke D.L.J. (2008) Evolution of the parasitic wasp subfamily Rogadinae (Braconidae): phylogeny and evolution of lepidopteran host ranges and mummy characteristics. *BMC Evolutionary Biology*, **8**.
Zerbino D.R., Birney E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs *Genome Research*, **18**, 821–829.

1274**Tables**

1275Table 1. Taxon sampling for transcriptomics and hybrid enrichment.

Higher grouping	**(Sub)family**	**Genus**	**Species**	**Specimen ID**	**Data type[1]**	**Country[2]**	**Repos[3]**
Outgroup: Chalcidoidea	Pteromalidae	cf. *Eupelmophotismus*	sp.	Ichn-2467	hybEnr	AU	WC
Outgroup: Evanioidea	Gasteruptiidae	*Gasteruption*	sp.	Ichn-2470	hybEnr	AU	WC
Outgroup: Braconidae	Alysiinae	*Dacnusa*	*sibirica*	1KITE	transcr	DE	-
Outgroup: Braconidae	Aphidiinae	*Aphidius*	*colemani*	1KITE	transcr	DE	-
Outgroup: Braconidae	Aphidiinae	*Diaeretus*	*essigellae*	1KITE	transcr	AU	-
Outgroup: Braconidae	Helconinae	Gen.	sp.	Ichn-2469	hybEnr	AU	-
Outgroup: Braconidae	Macrocentrinae	*Macrocentrus*	*marginator*	1KITE	transcr	SE	-
Outgroup: Braconidae	Microgastrinae	*Cotesia*	*vestalis*	1KITE	transcr	NL	-
Outgroup: Braconidae	Rogadinae	*Aleiodes*	*testaceus*	1KITE	transcr	FR	-
Ophioniformes	Anomaloninae	*Heteropelma*	*amictum*	1KITE	transcr	CH	-
Ophioniformes	Banchinae	*Apophua*	*evanescens*	transcr-SK09	transcr	CH	NMBE
Ophioniformes	Campopleginae	*Campoplex*	*capitator*	Ichn-2519	hybEnr	FR	NMBE
Ophioniformes	Campopleginae	*Hyposoter*	*didymator*	1KITE	transcr	FR	-
Ophioniformes	Cremastinae	*Dimophora*	*kentmartini*	Ichn-2465	hybEnr	AU	WC
Ophioniformes	Ctenopelmatinae	*Xenoschesis*	*fulvicornis*	Ichn-2359	hybEnr	CH	NMBE
Ophioniformes	Mesochorinae	*Astiphromma*	sp.	Ichn-2453	hybEnr	US	KYU
Ophioniformes	Metopiinae	*Colpotrochia*	*cincta*	transcr-SK10	transcr	CH	-
Ophioniformes	Ophioninae	*Leptophion*	*anici*	Ichn-2462	hybEnr	AU	WC
Ophioniformes	Tersilochinae	*Diaparsis*	sp.	Ichn-2449	hybEnr	CH	NMBE
Ophioniformes	Tryphoninae	*Netelia*	cf. *melanura*	1KITE	transcr	SE	-
Ophioniformes	Tryphoninae	*Netelia*	sp.	1KITE	transcr	HN	-
Ichneumoniformes	Alomyinae	*Alomya*	*debellator*	Ichn-2387	hybEnr	CH	NMBE
Ichneumoniformes	Cryptinae	*Buathra*	*laborator*	1KITE	transcr	SE	NMBE
Ichneumoniformes	Ichneumoninae	*Ichneumon*	*albiger*	1KITE	transcr	CH	-

Labeniformes	Labeninae	*Poecilocryptus*	cf. *nigromaculatus*	Ichn-2466	hybEnr	AU	WC
Pimpliformes	Acaenitinae	*Arotes*	*moiwanus*	Ichn-2341	hybEnr	JP	NMBE
Pimpliformes	Acaenitinae	*Coleocentrus*	*excitator*	transcr-SK01	transcr	CH	NMBE
Pimpliformes	Acaenitinae	*Jezarotes*	*tamanukii*	Ichn-2338	hybEnr	JP	NMBE
Pimpliformes	Acaenitinae	*Spilopteron*	*occiputale*	Ichn-2507	hybEnr	US	NMBE
Pimpliformes	Acaenitinae	*Yezoceryx*	sp.	Ichn-2509	hybEnr	TH	KYU
Pimpliformes	Collyriinae	*Collyria*	*trichophthalma*	transcr-SK11	transcr	CH	-
Pimpliformes	Cylloceriinae	*Cylloceria*	*melancholica*	Ichn-2131	hybEnr	SE	NMBE
Pimpliformes	Cylloceriinae	*Cylloceria*	cf. *caligata*	Ichn-2524	hybEnr	US	KYU
Pimpliformes	Diacritinae	*Diacritus*	*aciculatus*	Ichn-2143	hybEnr	SE	NMBE
Pimpliformes	Diacritinae	*Diacritus*	*aciculatus*	Ichn-2522	hybEnr	SE	NMBE
Pimpliformes	Diplazontinae	*Homotropus*	*frontorius*	Ichn-2365	hybEnr	CH	NMBE
Pimpliformes	Diplazontinae	*Sussaba*	*cognata*	Ichn-2513	hybEnr	CH	NMBE
Pimpliformes	Diplazontinae	*Syphophilus*	*tricinctorius*	1KITE	transcr	CH	-
Pimpliformes	Diplazontinae	*Syrphophilus*	*tricinctorius*	Ichn-2364	hybEnr	CH	NMBE
Pimpliformes	Diplazontinae	*Tymmophorus*	*suspiciosus*	Ichn-2460	hybEnr	CH	NMBE
Pimpliformes	Orthocentrinae	*Eusterinx*	*oligomera*	Ichn-2502	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	*Hemiphanes*	*erratum*	Ichn-2523	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	*Hemiphanes*	*performidatum*	Ichn-2499	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	*Megastylus*	*flavopictus*	Ichn-2183	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	*Neurateles*	sp.	Ichn-2526	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	*Orthocentrus*	sp.	Ichn-2500	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	*Picrostigeus*	*debilis*	Ichn-2498	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	*Plectiscidea*	cf. *tenuicornis*	Ichn-2503	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	*Proclitus*	nr. *praetor*	Ichn-2510	hybEnr	SE	NMBE
Pimpliformes	Pimplinae	*Acrodactyla*	*degener*	Ichn-2137	hybEnr	SE	NMBE
Pimpliformes	Pimplinae	*Amazopimpla*	sp.	Ichn-2455	hybEnr	EC	UTU
Pimpliformes	Pimplinae	*Anastelgis*	sp.	Ichn-2454	hybEnr	NI	UTU
Pimpliformes	Pimplinae	*Apechthis*	*compunctor*	transcr-SK05	transcr	CH	NMBE
Pimpliformes	Pimplinae	*Atractogaster*	*semisculptus*	Ichn-2278	hybEnr	SE	SMTP

Pimpliformes	Pimplinae	*Camptotypus*	sp.	Ichn-2434	hybEnr	CG	NMBE
Pimpliformes	Pimplinae	*Delomerista*	*mandibularis*	Ichn-2528	hybEnr	SE	KYU
Pimpliformes	Pimplinae	*Dolichomitus*	*quercicolus*	transcr-SK04	transcr	CH	NMBE
Pimpliformes	Pimplinae	*Dolichomitus*	*imperator*	transcr-SK15	transcr	CH	-
Pimpliformes	Pimplinae	*Echthromorpha*	*intricatoria*	Ichn-2464	hybEnr	AU	WC
Pimpliformes	Pimplinae	*Flavopimpla*	*nigromaculata*	Ichn-2458	hybEnr	TH	KYU
Pimpliformes	Pimplinae	*Liotryphon*	cf. *punctulatus*	Ichn-2361	hybEnr	CH	NMBE
Pimpliformes	Pimplinae	*Nomosphecia*	*pyramida*	Ichn-2093	hybEnr	TH	KYU
Pimpliformes	Pimplinae	*Parema*	*nigrobalteata*	Ichn-2497	hybEnr	TH	KYU
Pimpliformes	Pimplinae	*Perithous*	*scurra*	Ichn-2525	hybEnr	SE	SMTP
Pimpliformes	Pimplinae	*Pimpla*	*rufipes*	transcr-SK03	transcr	CH	-
Pimpliformes	Pimplinae	*Pimpla*	*flavicoxis*	1KITE	transcr	SE	NMBE
Pimpliformes	Pimplinae	*Piogaster*	*pilosator*	Ichn-2144	hybEnr	CH	NMBE
Pimpliformes	Pimplinae	*Polysphincta*	*tuberosa*	Ichn-2518	hybEnr	SE	NMBE
Pimpliformes	Pimplinae	*Reclinervellus*	*nielseni*	Ichn-2505	hybEnr	CZ	NMBE
Pimpliformes	Pimplinae	*Schizopyga*	*frigida*	Ichn-2136	hybEnr	SE	NMBE
Pimpliformes	Pimplinae	*Theronia*	*hilaris*	Ichn-2433	hybEnr	US	NMBE
Pimpliformes	Pimplinae	*Xanthopimpla*	*varimaculata*	Ichn-2448	hybEnr	TH	KYU
Pimpliformes	Pimplinae	*Zatypota*	*percontatoria*	Ichn-2139	hybEnr	SE	NMBE
Pimpliformes	Pimplinae	*Zonopimpla*	sp.	Ichn-2457	hybEnr	GU	UTU
Pimpliformes	Poemeniinae	*Cnastis*	*vulgaris*	Ichn-2354	hybEnr	JP	NMBE
Pimpliformes	Poemeniinae	*Deuteroxorides*	*elevator*	transcr-SK07	transcr	CH	NMBE
Pimpliformes	Poemeniinae	*Eugalta*	cf. *strigosa*	Ichn-2442	hybEnr	TH	KYU
Pimpliformes	Poemeniinae	*Poemenia*	*hectica*	Ichn-2234	hybEnr	SE	NMBE
Pimpliformes	Poemeniinae	*Pseudorhyssa*	*nigricornis*	Ichn-2344	hybEnr	JP	NMBE
Pimpliformes	Rhyssinae	*Epirhyssa*	sp.	Ichn-2074	hybEnr	TH	KYU
Pimpliformes	Rhyssinae	*Rhyssa*	*amoena*	transcr-SK08	transcr	CH	-
Pimpliformes	Rhyssinae	*Triancyra*	*galloisi*	Ichn-2343	hybEnr	JP	NMBE
Xoridiformes	Xoridinae	*Xorides*	*praecatorius*	transcr-SK12	transcr	CH	NMBE

1276

[1] Data type: transcr = transcriptome, hybEnr = hybrid enrichment data

[2] Country of origin of the specimen in ISO 3166-1 alpha-2 codes

[3] Repositories of voucher specimens: KYU = Department of Entomology, University of Kentucky, USA; NMBE = Natural History Museum Bern, Switzerland; SMTP = Swedish Malaise Trap Project, Station Linné on Öland, Sweden; UTU = University of Turku, Finland; WC = Waite Insect collection, University of Adelaide, Australia. A dash indicates that no voucher specimens are available as they were ground for RNA extractions (if a museum collection is mentioned for specimens used for transcriptome sequencing, it refers to a pseudo-voucher/para-or syngenophore, i.e., identical morphospecies collected at the same locality as the specimen whose transcriptome was sequenced).

1285Table 2. Established oligonucleotide primer pairs for Sanger sequencing of eight exons. Gene
1286numbers correspond to the *Nasonia* genome annotation and are preceded by "Nasvi2EG" on
1287Genbank.

Gene	Primer pair	Primer sequence 5'–3'	Annealing temp	Fragment length
014339	F	GGC AAG AGG GTC AAA TTA CA	51–52 °C	142 bp + intron +190 bp
	R	CTC CAC AAT CTT TCC ATG CC		
011390	F	GAT CGT CGA AGT TAC AGC TA	49–50 °C	357 bp
	R	GGA GTG CCG GTC ATA AAG TC		
013087	F1	CTG AAG ACC ATT TCC CTG CG	51–52 °C	389 bp
	R1	GCG ACC TTG GAA GCA TCT TG		
	F2	CAG CAG CAA CGT GGA ATG G	51 °C	413 bp
	R2	GCG ACC TTG GAA GCA TCT TG		
001408	F	ACC GGT GGT CAC GAT TC	51 °C	377 bp
	R	CCA CAA TTT GAA TGT TCG ATC		
003698	F	CGA GCA GAG AGA GTA CAT G	51–52 °C	536 bp
	R	GGA AGA AGT CGT ATT TGA GGA A		
010812	F	GAT TAT CGT CCG TAC TTC AC	48–49 °C	140 bp + intron + 205 bp
	R	TAT CTC TCG AGG GGA ATC AT		
008805	F	GAG GAT GTC TCG AGC GC	52–53 °C	328 bp
	R	CTA GCG AAA TCA TAA GCG TAT C		
021461	F	GGA GTC GAC ATA AAC GGA TA	50–51 °C	404 bp
	R	GTG ATG TAG CCG TGT CTG AT		

1288

1289

1290

1291

1292

1293

1294Table 3. Summary of gene choice criteria.

Criterion[1]		Mean / median	95% range	#genes[2]
Stationarity	$p > 0.05$	0.365 / 0.212	0.000–0.999	449
Clocklikeness	igrvar < 0.015	0.045 / 0.017	0.004–0.067	305

Root correct	bootstrap > 0.5	0.52 / 0.52	0.04–0.90	380
#SiteRates	> 80	117.7 / 115.5	73–178	683
Taxonomic bootscore	for final ranking	259.2 / 263.5	137–369	—

[1]Criteria: Stationarity = homogeneity of nucleotide composition across taxa, tested by Chi-square test, with genes with p-values above 0.05 retained. Clocklikeness = variance of IGR relaxed-clock model in single-gene analysis, genes retained above 0.015. Root correct = bootstrap support for placement of *Xorides* as sister to all other Ichneumonidae, gene retained if it was above 50%. #SiteRates = number of sites evolving at near-optimal rate (according to Klopfstein *et al.*, 2017: between 0.1 and 0.4 substitutions from root to tip), genes retained if they had more than 80. Taxonomic bootscore = sum of bootstrap support for 12 groupings chosen as unequivocal in both our and previous analyses.

[2]Number of genes that would have been retained under each of the criteria if applied individually.

Table 4. Number of transitions between idiobiont and koinobiont parasitoid strategy as inferred from trees from the Bayesian posterior sample or from the ML bootstrap analyses, along with the average of the estimated probabilities that the ancestor of Pimpliformes was an idiobiont.

Analysis	Number of transitions				Probability (MRCA=idiobiont)
	mean	media n	min	max	
PhyloBayes (AA)	5.0	5	5	6	0.99
MrBayes (AA)	5.8	6	5	6	0.18
RAxML (AA)	5.5	6	5	7	0.58
RAxML (12)	5.2	5	5	7	0.98
RAxML (123)	5.1	5	5	6	0.99

Figure legends

Figure 1. Maximum likelihood trees from concatenated sequences of 723 genes present in all transcriptome taxa. Support values are given as percentages, with maximum support indicated by an asterisk. The tree on the left is the result of the analysis of partitioned amino acid sequences, with support values representing non-parametric bootstraps and 10-gene jackknifing. The tree on the right resulted from the analysis of the nucleotide data, including all three codon positions. Support values represent bootstrap support from all three codon positions, first and second, and percentages of the 10-gene jackknifing datasets recovering the nodes, respectively. The only difference in the tree based on first and second codon positions was the placement of Rhyssa as sister to ((*Collyria* + *Coleocentrus*) + *Dolichomitus*).

Figure 2. Consensus tree resulting from the Bayesian analysis of the amino acid sequence data under an empirical profile mixture model as implemented in PhyloBayes, with support values added from the partitioned Bayesian analysis in MrBayes, maximum likelihood analysis in RAxML, and of the nucleotide analyses in RAxML. Support values represent Bayesian posterior probabilities and bootstrap supports, respectively. Note the very short branch lengths at the base of the Pimpliformes. Of taxa with an asterix after the name, we have analysed the transcriptomes, *Nasonia vitripennis* was included based on its published data. Current subfamily classification of the pimpliform genera is indicated by coloured taxon labels, which start with three letters corresponding to the subfamily name or to the family (upper-case letters) name in the outgroups as follows: Aca = Acaenitinae, Alo = Alomyinae, Ano = Anomaloninae, Ban = Banchinae, BRA = Braconidae, Cam = Campopleginae, Col = Collyriinae, Cre = Cremastinae, Cry = Cryptinae, Cte = Ctenopelmatinae, Cyl = Cylloceriinae, Dia = Diacritinae, Dip = Diplazontinae, GAS = Gasteruptiidae, Ich = Ichneumoninae, Lab = Labeninae, Mes = Mesochorinae, Met = Metopiinae, Oph = Ophioninae, Ort = Orthocentrinae, Pim = Pimplinae, Poe = Poemeniinae, PTE = Pteromalidae, Rhy = Rhyssinae, Ter = Tersilochinae, Try = Tryphoninae, Xor = Xoridinae.

Figure 3. Phylogenetic uncertainty at the base of Pimpliformes shown by comparison of consensus phylogenies from the Bayesian analyses (a, c), and maximum likelihood trees from the ML analyses (d, e, f). Support values in the form of posterior probabilities (a, c) or bootstrap percentages (d, e, f) are only shown if lower than 1.0 or 100%, respectively.

1342Ancestral states of idiobiosis versus koinobiosis are shown as pies next to nodes, while stars
1343indicate state transitions as inferred under maximum parsimony. Panel (b) shows one of the
1344trees from the PhyloBayes analysis that shows an alternative reconstruction, with the
1345reconstruction of the MRCA of Pimpliformes in favour of it being a koinobiont (arrow), as is
1346also the case in panel (c). Outgroups and additional taxa within Ophioniformes and
1347Ichneumoniformes were suppressed here for clarity of the figures.

1348

1349 Figure 4. Tree resolution as measured as the average bootstrap support in an ML
1350analysis of all three codon positions plotted against the number of genes sampled (x-axis not
1351in scale). Boxplots show the variation in twenty (for 2–30 genes) and 10 replicates (for 40–90
1352genes), while lines represent the results from various analyses of genes sampled, non-
1353randomly, after ranking according to three criteria.