**Supplementary Material**

**Table. 1.** Summary regarding properties of each partition. Values were obtained from the GARLI (Zwickl, 2006) console output. m = number of sequences, n = number of total sites.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Gene partition | *m* | *N* | constant characters | uninformative variable | parsimony informative |
|
| CO1 (1) | 74 | 538 | 297 | 53 | 188 |
| CO1 (2) | 74 | 538 | 400 | 50 | 88 |
| CO1 (3) | 74 | 538 | 77 | 59 | 402 |
| 28S | 69 | 1854 | 1577 | 107 | 170 |
| 16S | 57 | 434 | 226 | 47 | 161 |
| 12S | 49 | 149 | 78 | 24 | 47 |

1.1 *Phylogenetic methodology*

 Gaps were constantly treated as missing data. Molecules were analyzed alone, in combination with morphology and behavior, or combined solely with morphology. Data were analyzed using parsimony (MP), maximum-likelihood (ML) and Bayesian methods due to the array of advantages and disadvantages of each method. Hence, a total of nine different phylogenetic constructions are presented.

 Model testing for gene partitions was implemented in jModelTest 2 (Darriba *et al*., 2012) to determine the most appropriate model to employ for ML and Bayesian analysis. The results are summarized in Table 2. Alignment for CO1 data included a 75th taxon: *Discoelius zonalis*. *D. zonalis* was excluded from the primary phylogenetic analysis because only a portion of CO1 sequence was available. Akaike Information Criterion (AIC) scores were calculated to evaluate which partitioning scheme was best (Table 3). CO1 was partitioned based on codon positions, and the third nucleotide position was included in all analyses.

**Table. 2.** Summary of jModelTest 2 results for static alignments including *Discoelius zonalis*. Scores for Akaike Information Criterion (AIC) uncorrected and corrected (AICc) and Bayesian Information Criterion (BIC) are indicated along with the best model for each criterion.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Partition | AIC | AICc | BIC | AIC | AICc | BIC |
| 28S, 69taxa/1858sites | 12437.86 | 12462.95 | 13244.84 | GTR+I+G | GTR+I+G | GTR+I+G |
| 12S, 49taxa/149sites | 3034.20 | 3486.29 | 3345.29 | TVM+I+G | HKY+G | HKY+I+G |
| 16S, 57taxa/434sites | 9489.01 | 9585.52 | 9985.43 | GTR+I+G | GTR+I+G | TVM+I+G |
| CO1, 1614 sites 75 taxa | 54397.47 | 54432.00 | 55248.54 | GTR+I+G | GTR+I+G | GTR+I+G |
| CO1, pos. 1 (1)  | 14278.90 | 14411.47 | 14956.38 | GTR+I+G | GTR+I+G | GTR+I+G |
| CO1, pos. 2 (2)  | 5677.62 | 5808.18 | 6350.82 | TVM+I+G | TVM+I+G | TVM+I+G |
| CO1, pos. 3 (3)  | 31857.61 | 31982.22 | 32515.54 | TPM1uf+I+G | HKY+G | HKY+G |
| CO1, pos. 1+2 (12) | 20248.37 | 20303.16 | 21035.37 | GTR+I+G | GTR+I+G | GTR+I+G |

*Parsimony* – In order to construct a topology based upon the optimality criterion of maximum parsimony, the program TNT was used (Goloboff *et al*., 2008). The best trees were obtained by implementing a search strategy of 1000 random additions, each with 40 rounds of parsimony ratchet, 20 rounds of tree drifting and 30 rounds of fusing. *Scolia verticalis* was set as the outgroup. Characters labeled additive by Pickett and Carpenter (2010) were treated as additive in the simultaneous analysis. When behavior was excluded, all data were treated as non-additive. Each bootstrap replicate consisted of 10 random additions, each with 40 rounds of parsimony ratchet, 20 rounds of tree drifting and 30 rounds of fusing. Limited bootstrap replicates were done because of time constraints.

*Maximum Likelihood* – ML analyses were performed in GARLI (Zwickl, 2006). All analyses had random starting trees. Termination conditions were enforced and the following settings used: genthreshfortopoterm = 20000, scorethreshforterm = 0.001. The models were set as unlinked and with subset specific rates. The models implemented were: GTR+I+G for the first codon of CO1, TVM+I+G for the second codon of CO1, TPM1uf+I+G for the third codon of CO1, GTR+I+G for 28S, GTR+I+G for 16S and HKY+I+G for 12S. When adding the morphological matrix, an Mkv (Lewis, 2001) unordered model was specified to it. The modweight entry was modified for each run according to the formula: 0.0005 x (# of subsets + 1). Outgroup taxa (Table 5) were forcefully constrained. The rest of the values were left as defaults according to the configuration file templates specific to large data sets provided with GARLI (Zwickl, 2006). Several replicates were done to ensure that the best possible topology was being recovered. Sometimes suboptimal trees are produced due to searches becoming trapped in local branch length optimizations (Zwickl, 2006). For our data this was not the case; the vast majority of replicates produced topological outputs that were identical.

*Bayesian –* Bayesian analyses were performed with MrBayes 3.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; Ronquist *et al*., 2012).Priors were left flat and all other defaults were also unchanged unless noted. All models were unlinked using lset commands and allowed to evolve with independent rates of evolution using the command “prset ratepr=variable”, which is equivalent to the GARLI (Zwickl, 2006) “subsetspecificrates = 1” entry. All DNA partitions were assigned the GTR+I+G model, except 12S was assigned the simpler HKY+I+G model (see justification section). Morphological data was assigned an Mkv model (Lewis, 2001) with gamma rate heterogeneity. Again, outgroups were forcefully constrained. A relative burn-in of 25% was set and 2 runs with 4 chains each were performed. All analyses were allocated 3 million generations and all reached a standard deviation of split frequencies below 0.01.

 The Alignments used are included as a second supplement. This file can be viewed using notepad or Mesquite; I recommend opening with Mesquite.

**Table. 3.** Akaike Information Criterion (AIC) scores for different partitioning schemes of CO1. LnL scores are derived from GARLI (Zwickl, 2006). The number of free parameters is represented by k. These values were obtained with inclusion of *Discoelius zonalis*. Brackets denote that an unlinked and separate model with subset specific rates was applied to that partition.

|  |  |  |  |
| --- | --- | --- | --- |
| CO1 Partition Scheme | Garli LnL score | *k* | AIC |
| (1)(2)(3/HKY+G) | -25840.487 | 24 | 51728.974 |
| (12)(3/HKY+G) | -25995.6977 | 15 | 52021.3954 |
| (1,2,3)GTR+I+G | -27025.7652 | 10 | 54071.5304 |
| (1)(2)(3/TPM1uf+I+G) | -25836.9048 | 26 | 51725.8096 |



**Fig. 1.** Strict consensus of 6 equally parsimonious trees (L = 9415) obtained from simultaneous analysis of all molecular data, morphology and behavioral characters. Bootstrap values from 500 replicates indicated upon branches if ≥ 50. Subfamily taxa are represented by the following coloring scheme: Green = Euparagiinae; Blue = Masarinae; Red = Stenogastrinae; Orange = Eumeninae; Brown = Polistinae; Cyan = Vespinae.

 

**Fig. 2.** Strict consensus of 6 equally parsimonious trees (L = 9271) obtained from simultaneous analysis of all molecular data, morphology but excluding behavioral characters. Bootstrap values from 100 replicates indicated upon branches if ≥ 50. Subfamily taxa are represented by the following coloring scheme: Green = Euparagiinae; Blue = Masarinae; Red = Stenogastrinae; Orange = Eumeninae; Brown = Polistinae; Cyan = Vespinae.



**Fig. 3.** Strict consensus of 2 equally parsimonious trees (L = 8610) obtained from analysis of molecular data only. Bootstrap values from 250 replicates indicated upon branches if ≥ 50. Subfamily taxa are represented by the following coloring scheme: Green = Euparagiinae; Blue = Masarinae; Red = Stenogastrinae; Orange = Eumeninae; Brown = Polistinae; Cyan = Vespinae.



**Fig. 4.** Bayesian phylogeny derived from simultaneous analysis of molecules, morphology and behavior. Posterior probability (PP) indicated on branches; branches with PP below 0.5 collapsed. Subfamily taxa are represented by the following coloring scheme: Green = Euparagiinae; Blue = Masarinae; Red = Stenogastrinae; Orange = Eumeninae; Brown = Polistinae; Cyan = Vespinae.



**Fig. 5.** Bayesian phylogeny derived from simultaneous analysis of molecules and morphology but excluding behavior. Posterior probability (PP) indicated on branches; branches with PP below 0.5 collapsed. Subfamily taxa are represented by the following coloring scheme: Green = Euparagiinae; Blue = Masarinae; Red = Stenogastrinae; Orange = Eumeninae; Brown = Polistinae; Cyan = Vespinae.



**Fig. 6.** Bayesian phylogeny derived from analysis of molecules alone. Posterior probability (PP) indicated on branches; branches with PP below 0.5 collapsed. Subfamily taxa are represented by the following coloring scheme: Green = Euparagiinae; Blue = Masarinae; Red = Stenogastrinae; Orange = Eumeninae; Brown = Polistinae; Cyan = Vespinae.



**Fig. 7.** Maximum likelihood tree (-ln = 42623.0886) derived from simultaneous analysis of molecules, morphology and behavior (unordered). Subfamily taxa are represented by the following coloring scheme: Green = Euparagiinae; Blue = Masarinae; Red = Stenogastrinae; Orange = Eumeninae; Brown = Polistinae; Cyan = Vespinae.



**Fig. 8.** Maximum likelihood tree (-ln = 41938.6760) derived from simultaneous analysis of molecules and morphology but excluding behavior (unordered). Subfamily taxa are represented by the following coloring scheme: Green = Euparagiinae; Blue = Masarinae; Red = Stenogastrinae; Orange = Eumeninae; Brown = Polistinae; Cyan = Vespinae.



**Fig. 9.** Maximum likelihood tree (-ln = 38671.1889) derived from analysis of molecules only. Subfamily taxa are represented by the following coloring scheme: Green = Euparagiinae; Blue = Masarinae; Red = Stenogastrinae; Orange = Eumeninae; Brown = Polistinae; Cyan = Vespinae.

**Table. 4.** Summary of selected relationships. Y = yes, the relationship is supported; N = no, the relationship is not supported. P.C. 2010 represents results of Pickett and Carpenter (2010). ML = Maximum Likelihood; MP = Maximum Parsimony. Bayesian posterior probabilities converted into percentages. Support values listed in the order of ML, Bayesian and MP.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Relationships | Molecules only (ML,Bayes,MP) | Simultaneous Analysis(ML,Bayes,MP) | All – behavior (ML,Bayes,MP) | P.C. 2010  |
| Monophyly of… |  |  |  |  |
| Euparagiinae | Y,Y,Y (-,100,100) | Y,Y,Y(-,100,100) | Y,Y,Y (-,100,100) | Y (100) |
| Masarinae | Y,Y,Y (-,100,74) | Y,Y,Y (-,100,95) | Y,Y,Y (-,100,93) | Y (100) |
| Eumeninae | Y,Y,Y (-,97.4,68) | Y,Y,Y (-,97.5,79) | Y,Y,Y (-,99.6,55) | Y (100) |
| Stenogastrinae | Y,Y,Y (-,100,99) | Y,Y,Y(-,100,100) | Y,Y,Y (-,100,100) | Y (100) |
| Polistinae | N,N,N | Y,Y,Y (-,100,99) | Y,Y,Y (-,100,96) | Y (100) |
| Vespinae | Y,Y,Y (-,100,100) | Y,Y,Y(-,100,100) | Y,Y,Y (-,100,100) | Y (100) |
| Ropalidiini | N,N,N | Y,Y,Y (-,100,91) | Y,Y,Y (-,100,91) | Y (100) |
| Mischocyttarini | Y,Y,Y (-,100,96) | Y,Y,Y(-,100,100) | Y,Y,Y (-,100,100) | Y (100) |
| Epiponini | Y,Y,Y (-,100,92) | Y,Y,Y (-,100,99) | Y,Y,Y (-,100,98) | Y (100) |
| Polistini | Y,Y,Y (-,100,94) | Y,Y,Y(-,100,100) | Y,Y,Y (-,100,100) | Y (100) |
| (Epiponini + Polistini) | Y,Y,Y (-,70.2,66) | N,Y,Y (-,61.8, ≥50) | N,Y,Y (-,62.5,56) | Y (95) |
| (Euparagiinae + Masarinae) | Y,Y,Y (-,99.7,50) | Y,N,N (-,-,-) | Y,Y,N (-.50.5,-) | N |
| (Provespa + Vespa) | Y,Y,Y(-,94.4,≥50) | N,N,Y (-,-,59) | N,N,Y (-,-,52) | Y (98) |
| *Ceramius*  | N,N,N | N,N,N | N,N,N | N |
| *Liostenogaster* | N,N,Y (-,-,53) | N,N,Y (-,-,89) | N,N,Y (-,-,86) | Y (99) |
| *Polybia* | N,N,N | Y,Y,Y(-,100,≥50) | Y,Y,N (-,99.6,-) | N |
| *Eumenes* s.l. | Y,Y,Y (-,100,99) | Y,Y,Y(-,100,100) | Y,Y,Y (-,100,100) | Y (100) |
| **Social groups** | **N,N,N** | **Y,Y,Y (-,100,87)** | **Y,Y,N (-,96.6,-)** | **Y (100)** |

ML analysis on just mitochondrial DNA (mtDNA) data places *B. freyi* within the Eumeninae. Using only 28S data and performing an ML analysis: *Belonogaster* came out monophyletic; Polistinae and Vespinae were sisters. Upon BLAST searching the *B. freyi* 16S data I found that sequence similarity was highest with eumenine species. Also the CO1 sequence is drastically different from the other *Belonogaster* sequences. Thus, the odd placement of *B. freyi* is due to false sequence data. The result of Polistinae paraphyly should be ignored.

|  |  |
| --- | --- |
| Taxa**Table. 5.** List of taxa and loci sampled in this study. A \* indicates samples which were collected personally. Unpbl.C indicates unpublished sequence data obtained from J.M. Carpenter. A ^ denotes material that is available but not used in this analysis due to failed incorporation into the alignment.  | GenBank Accession NumbersCOI 28S 16S 12S |
| Outgroup |  |  |  |  |
| *Rhopalosoma nearcticum* | EU567206 | GQ374724 | N/A | N/A |
| *Scolia verticalis* | GU596820 | EF013060 | N/A^ | GU596556 |
|  |  |  |  |  |
| Euparagiinae |  |  |  |  |
| *Euparagia desertorum* | GU596821 | GU596726 | GU596598 | GU596509 |
| *Euparagia scutellaris* | GU596949 | GU596727 | GU596692 | GU596810 |
|  |  |  |  |  |
| Masarinae |  |  |  |  |
| *Ceramiopsis gestroi* | GU596822 | GU596714 | GU596604 | GU596588 |
| *Ceramius beyeri* | GU596825 | GU596715 | GU596603 | N/A |
| *Ceramius lichtensteinii* | GU596823 | GU596716 | N/A | N/A |
| *Ceramius micheneri* | GU596824 | GU596717 | GU596602 | GU596504 |
| *Ceramius richardsi* | GU596826 | GU596718 | GU596601 | GU596505 |
| *Ceramius socius* | GU596827 | GU596719 | N/A | GU596506 |
| *Jugurtia braunsiella* | GU596828 | GU596728 | GU596600 | GU596511 |
| *Trimeria rhachiphora* | GU596829 | GU596812 | GU596599 | GU596557 |
|  |  |  |  |  |
| Eumeninae |  |  |  |  |
| *Abispa ephippium* | NC011520 | JF510017 | EU302588 | N/A^ |
| *Ancistroceroides alastoroides* | GU596832 | GU596703 | GU596687 | N/A |
| *Ancistrocerus antilope* | JN934263 | EF190743 | N/A | N/A |
| *Ancistrocerus waldenii* | GU596831 | GU596703 | GU596606 | N/A |
| *Discoelius zonalis* | JN934345.1 | N/A | N/A | N/A |
| *Eumenes fraternus* | EU649440 | EF190749 | N/A | N/A |
| *Eumenes tripunctatus* | GU596834 | GU596725 | GU596607 | GU596567 |
| *Euodynerus megaera* | EU649462 | EF190753 | N/A | N/A |
| *Euodynerus quadrifasciatus* | JN934337 | N/A | N/A | N/A |
| *Katamenes dimidiatus* | GU596835 | GU596729 | N/A | GU596512 |
| *Monobia quadridens* | GU596833 | GU596750 | GU596696 | GU596621 |
| *Parancistrocerus perennis* | GU596830 | GU596751 | GU596605 | GU596522 |
| *Symmorphus cristatus* | EU649667 | EF190750 | N/A | N/A |
| *Synagris calida* | Unpbl.C + \* | Unpbl.C | Unpbl.C | Unpbl.C |
| *Synagris cornuta cornuta* | Unpbl.C | Unpbl.C | Unpbl.C | Unpbl.C |
| *Zeta mendozanum* | GU596836 | GU596818 | GU596608 | GU596561 |
|  |  |  |  |  |
| Stenogastrinae |  |  |  |  |
| *Anischnogaster iridipennis* | GU596837 | GU596705 | N/A | N/A |
| *Anischnogaster laticeps* | GU596838 | GU596706 | N/A | N/A |
| *Liostenogaster flavolineata* | GU596839 | GU596730 | N/A^ | GU596568 |
| *Liostenogaster vechti* | GU596840 | GU596731 | N/A^ | N/A |
| *Parischnogaster mellyi* | GU596841 | GU596752 | GU596697 | GU596523 |

|  |  |
| --- | --- |
| Taxa**Table 5** (continued) | GenBank Accession NumbersCOI 28S 16S 12S |
| Polistinae |  |  |  |  |
| *Agelaia pallipes* | GU596948 | GU596702 | GU596686 | GU596571 |
| *Apoica pallens* | GU596895 | GU596707 | GU596614 | N/A |
| *Apoica strigata* | GU596897 | GU596709 | GU596616 | GU596572 |
| *Belonogaster clypeata clypeata* | Unpbl.C | Unpbl.C | Unpbl.C | Unpbl.C |
| *Belonogaster freyi* | Unpbl.C | Unpbl.C | Unpbl.C | Unpbl.C |
| *Belonogaster grisea* | Unpbl.C | Unpbl.C | Unpbl.C | Unpbl.C |
| *Belonogaster juncea juncea* | Unpbl.C | Unpbl.C | Unpbl.C | Unpbl.C |
| *Metapolybia aztecoides* | GU596903 | GU596732 | GU596622 | GU596589 |
| *Metapolybia cingulata* | GU596904 | GU596733 | GU596623 | GU596575 |
| *Mischocyttarus carinulatus* | GU596852 | GU596735 | GU596626 | GU596514 |
| *Mischocyttarus cearensis* | GU596853 | GU596736 | GU596627 | GU596515 |
| *Mischocyttarus deceptus* | GU596851 | GU596737 | GU596624 | GU596513 |
| *Mischocyttarus drewseni gigas* | GU596907 | GU596738 | GU596628 | GU596516 |
| *Mischocyttarus flavitarsis* | GU596908 | GQ374721 | GU596629 | GU596563 |
| *Mischocyttarus latior* | GU596909 | GU596740 | GU596630 | N/A^ |
| *Mischocyttarus lemoulti* | GU596911 | GU596742 | GU596632 | GU596519 |
| *Mischocyttarus melanarius* | GU596913 | GU596743 | GU596634 | GU596590 |
| *Mischocyttarus paraguayensis* | GU596916 | GU596746 | GU596636 | N/A |
| *Polistes annularis* | GU596854 | GU596754 | GU596639 | GU596524 |
| *Polistes dominula* | GU596870 | GU596769 | GU596651 | GU596549 |
| *Polistes fuscatus* | GU596874\* | GU596772 | GU596653 | GU596534 |
| *Polistes japonicus* | GU596877 | GU596776 | GU596657 | GU596566 |
| *Polistes pacificus* | GU596885 | GU596785 | GU596665 | GU596543 |
| *Polybia belemensis* | GU596924 | GU596794 | GU596673 | N/A |
| *Polybia occidentalis occidentalis* | GU596931 | GU596799 | GU596678 | N/A |
| *Polybia raui* | GU596932 | GU596800 | GU596698 | N/A |
| *Polybia sericea* | GU596935 | GU596801 | GU596679 | N/A |
| *Polybioides melainus* | GU596937 | N/A | GU596681 | GU596583 |
| *Protonectarina sylveirae* | GU596938 | GU596803 | GU596682 | GU596584 |
| *Protopolybia emortualis* | GU596939 | GU596804 | GU596683 | N/A |
| *Protopolybia scutellaris* | GU596941 | GU596805 | GU596685 | GU596593 |
| *Pseudopolybia vespiceps* | GU596944 | GU596809 | GU596699 | GU596554 |
|  |  |  |  |  |
| Vespinae |  |  |  |  |
| *Dolichovespula adulterina* | Unpbl.\* | N/A | N/A | N/A |
| *Dolichovespula alpicola* | Unpbl.\* | N/A | N/A | N/A |
| *Dolichovespula arenaria* | GU596847\* | GU596723 | GU596691 | GU596508 |
| *Dolichovespula maculata* | GU596843 | GU596724 | GU596694 | N/A |
| *Provespa anomola* | GU596849 | GU596806 | GU596609 | GU596550 |
| *Provespa barthelemyi*  | GU596850 | GU596807 | GU596610 | GU596551 |
| *Vespa mandarinia* | GU596946 | GU596813 | GU596700 | GU596570 |
| *Vespa orientalis* | GU596844 | GU596814 | GU596611 | GU596558 |
| *Vespula germanica* | GU596845 | GU596815 | GU596612 | GU596595 |
| *Vespula pensylvanica* | Unpbl.\* | N/A | FJ821513 | N/A |

*References*

Darriba, D., Taboada, GL., Doallo, R. and Posada, D. 2012. jModelTest 2: more models, new

heuristics and parallel computing. *Nat. Methods.* 9: 772.

Goloboff, P.A., Farris, J.S. and Nixon, K.C. 2008. TNT, a free program for phylogenetic

analysis. *Cladistics*. 24: 774-786.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic

trees. *Bioinformatics*. 17: 754-755.

Lewis, P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological

character data. *Syst. Biol.* 50: 913-925.

Pickett, K.M. and Carpenter, J.M. 2010. Simultaneous Analysis and the Origin of Eusociality in

the Vespidae (Insecta: Hymenoptera). *Arthropod Syst. Phylogeny.* 68: 3-33.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference

under mixed models. *Bioinformatics*. 19: 1572-1574.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B.,

Liu, L., Suchard, M.A. and Huelsenbeck, J.P. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61:539-542.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.