MOLECULAR ECOLOGY

Supplementary material:

A remarkable legion of guests: diversity and host specificity of army ant symbionts

Authors: Christoph von Beeren [ab*], Nico Blüthgen [a], Philipp Hoenle [a], Sebastian Pohl [c], Adrian Brückner [d], Alexey K. Tishechkin [e], Munetoshi Maruyama [f], Brian V. Brown [g], John M. Hash [h], W. Eugene Hall [i], Daniel J.C. Kronauer [b*]

Affiliations:

[a] Ecological Networks, Department of Biology, Technical University of Darmstadt, Darmstadt, Germany

[b] Laboratory of Social Evolution and Behavior, The Rockefeller University, New York City, USA
[c] Yale-NUS College, Division of Science, Singapore, Singapore
[d] Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, USA
[e] California Department of Food and Agriculture, Plant Pest Diagnostics Center, Sacramento, USA
[f] The Kyushu University Museum, Hakozaki, Fukuoka, Japan
[g] Natural History Museum of Los Angeles County, Entomology section, Los Angeles, USA
[h] National Museum of Natural History, Smithsonian Institution, Department, Department of Entomology, Washington DC, USA
[i] University of Arizona Insect Collection, Arizona, USA

*Correspondence to:

Christoph von Beeren: cvonbeeren@gmail.com

Daniel J.C. Kronauer: dkronauer@rockefeller.edu

Content:

Supplementary results. Detailed species identification of myrmecophiles.
Figure S1. Clustering of nuclear gene data.
Figure S2. Host range distribution.
Figure S3. Interaction matrix and network modularity of the entire community.
Table S1. Specimen collection information, GenBank accession numbers, and interaction network matrix.
Table S2. PCR primer combinations used in this study.
Table S3. PCR primers used in this study.

MOLECULAR ECOLOGY

Supplementary Results. Detailed species identification of myrmecophiles.

We confirmed previous suspicions that the rove beetle taxon *Vatesus clypeatus* (Wasmann, 1887) (Fig. 1a) in fact constitutes a complex of several closely related and morphologically similar species (Seevers, 1965). Specimens of the two species *Vatesus* cf. *clypeatus* sp. 1 (N = 51 *COI* barcodes) and *Vatesus* cf. *clypeatus* sp. 2 (N = 354 *COI* barcodes) clustered into separate groups in the RAxML tree analysis of mitochondrial DNA barcodes (Fig. 2), but due to their close sequence similarity (minimum inter-cluster p-distance between species = 0.80%, see von Beeren et al. 2016a) were lumped in the same BIN by RESL (BIN: BOLD:ACH6492). However, by using nuclear gene data as well as morphological analysis of male copulatory pieces (aedeagi) and larval coloration we previously demonstrated that these two genetic units represent two species with distinct host spectra (von Beeren et al., 2016a). This case represents an example where standardized RESL-based BINs failed in correctly identifying species, highlighting the need of acquiring characters from multiple additional sources (morphological characters, nuclear gene sequences, host records, etc.) in species identifications.

Here we report the presence of a third species in the *V. clypeatus* species complex at LSBS. We collected five specimens of this species in a single colony of the army ant *E. lucanoides* - a host in which we had not detected *Vatesus* specimens previously. Using the latest species keys by Seevers (1965) the specimens keyed out as member of the *V. clypeatus* species complex. RESL analysis assigned the five individuals to a unique BIN (BIN: BOLD:AEF4336), and the specimens were also recovered as an idiosyncratic cluster in the RAxML tree analysis of *COI* sequences (Fig. 2). The presence of a third species in the *V. clypeatus* complex at LSBS was also supported by nuclear gene data. Each of the three *V.* cf. *clypeatus* species carried a single, idiosyncratic *wg* allele, implying a lack of nuclear gene flow between members of the three major mitochondrial clades (Fig. S1). We denominate this species as '*Vatesus* cf. *clypeatus* sp. 3' (maximum intra-cluster p-distance = 0.48%, range of *COI* sequence lengths: 601bp-642bp; minimum inter-cluster p-distance to other *Vatesus* species = 6.16%, range of *COI* sequence lengths: 159bp-677bp). Diagnostic morphological characters for each species will be included in a taxonomic revision of *Vatesus* beetles, which is currently in progress by CvB.

Specimens identified as the myrmecoid rove beetle *Ecitophya gracillima* Mann, 1925 were assigned to two distinct BINs by RESL, and these two BINs were also recovered as distinct clusters in the RAxML tree analysis (*COI* cluster-I, BIN: BOLD:ADH4433, N = 4; *COI* cluster-II, BOLD:ADH3769, N = 19; Fig. 2). The two clusters differed by a maximum p-distance of 1.67% (see also Fig. 2). In a previous study about myrmecoid beetles based on morphological identification, mitochondrial barcodes, and analysis of two nuclear loci (*wg* and *CAD*), we found no evidence for the existence of two species in the taxon *E. gracillima* (von Beeren et al., 2018). Due to the small *COI* sequence difference between the two BINs and lack of evidence for distinct species from morphological and nuclear gene data, we consider *E. gracillima* to be a single species at LSBS.

Specimens identified as the scuttle fly taxon *Ecitophora comes* Schmitz, 1914 split into three distinct BINs that were also recovered in the RAxML tree analysis (maximum intra-cluster p-distance = 1.09%, minimum inter-cluster p-distance = 13.53%, range of *COI* sequence lengths: 330bp-667bp; Fig. 3). A first morphological inspection did not uncover apparent morphological differences between these three genetic clusters, but certainly a more extensive taxonomic evaluation is needed. Because specimens of each of the three *COI* cluster also had distinct *wg* alleles (Fig. S1), i.e. there seems to be no free nuclear gene flow between these clusters, we treated *E.* cf. *comes* as three distinct species, which we denominate as '*Ecitophora* cf. *comes* sp. 1' (BIN: BOLD:AEB1427, N = 116), '*Ecitophora* cf. *comes* sp. 2' (BIN: BOLD:AEB1425, N = 22), and '*Ecitophora* cf. *comes* sp. 3' (BIN: BOLD:AEB1426, N = 12).

Specimens identified as *Ecitophora pilosula* Borgmeier, 1960 split into two distinct BINs that were also recovered as distinct clusters in the RAxML tree analysis (*COI* cluster-I, BIN: BOLD:AEB3015, N = 79; *COI* cluster-II, BIN: BOLD:AEB3014, N = 11; Fig. 3). The maximum intra-cluster p-distance was 1.37% and the minimum inter-cluster p-distance 13.41% (range of *COI* sequence lengths: 420bp-664bp). Again, we did not detect any apparent diagnostic morphological characters to distinguish *E. pilosula* specimens of the two barcode clusters. However, in contrast to specimens identified as *E.* cf. *comes*, nuclear gene data rather suggested that *E. pilosula* is a single species at LSBS because specimens of the two *COI* clusters shared the same *wg* alleles (Fig. S1). We thus treated *E. pilosula* as a single species in the present study.

Phorid fly specimens that keyed out as *Ecitophora halterata* (Borgmeier 1936) were assigned to two BINs that we also recovered as two distinct clusters in the RAxML tree analysis (*COI* cluster-I, BIN: BOLD:AEB0451, N = 2; *COI* cluster-II, BIN: BOLD:AEB0450, N = 1; Fig. 3). The two BINs differed by 5.02% p-distance. We only successfully amplified the *wg* gene fragment II of specimens belonging to *COI* cluster I so that nuclear gene data cannot help us here in disentangling species boundaries (Fig. S1). As we did not detect morphological differences between the specimens of the two clusters, we treated *E. halterata* as a single species at LSBS.

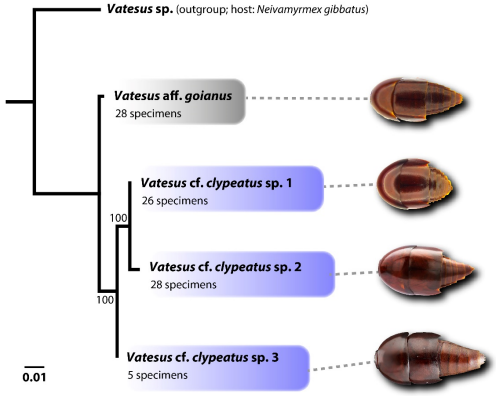
The specimens identified as *Ecitophora varians* Borgmeier, 1960 were assigned to three BINs that were also recovered as three clusters in the RAxML tree (*COI* cluster-I, BIN: BOLD:AEA9847, N = 6; *COI* cluster-II, BIN: BOLD:ADA4306, N = 13; *COI* cluster-III, BIN: BOLD:AEB6798, N = 37; Fig. 3). The maximum intra-cluster p-distance was 0.80% and the minimum inter-cluster p-distance 1.67%. The maximum p-distance between these three clusters was 8.74% (see also Fig. 3). Because specimens of the three *COI* clusters shared the same *wg* alleles in both studied *wg* fragments (Fig. S1) and because we did not detect apparent morphological differences between specimens of the three *COI* clusters, we treated *E. varians* as a single species.

MOLECULAR ECOLOGY

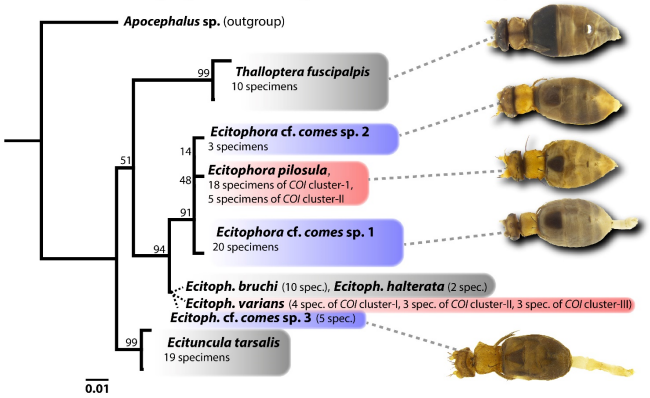
Figure S1. Clustering of nuclear gene data. RAxML clustering of *wingless* (*wg*) gene fragments of *Vatesus* beetles and phorid flies. Grey boxes show cases where morphological identification and *COI* barcode clustering agreed on the presence of a single species. Red and purple boxes highlight cases in which specimens initially identified as a single species split in two or more *COI* clusters. Additional morphological and/or genetic data suggested that those specimens belonged to either a single species (red boxes) or to different species (purple boxes; see also supplementary results). Scale bars show expected nucleotide substitutions per site as inferred by the RAxML algorithm. Bootstrap support values are shown at major nodes (1000 repetitions). We were not able to determine alleles in heterozygous *wg* sequences. Overlapping base peaks in *wg* consensus sequences were accordingly assigned capital letters for ambiguous base pairs according to the terminology of the IUPAC nucleotide code (e.g., R, Y, S). We did not detect any overlapping base peaks in *Vatesus* beetles' *wg* consensus sequences, indicating that all specimens were homozygous. In phorid flies 60 out of 210 analyzed sequences showed overlapping base pair peaks, indicating these specimens were heterozygous at *wg*.

MOLECULAR ECOLOGY

Vatesus beetles



Phorid flies - *wg* gene fragment I (270 bp)



Phorid flies - *wg* gene fragment II (210 bp)

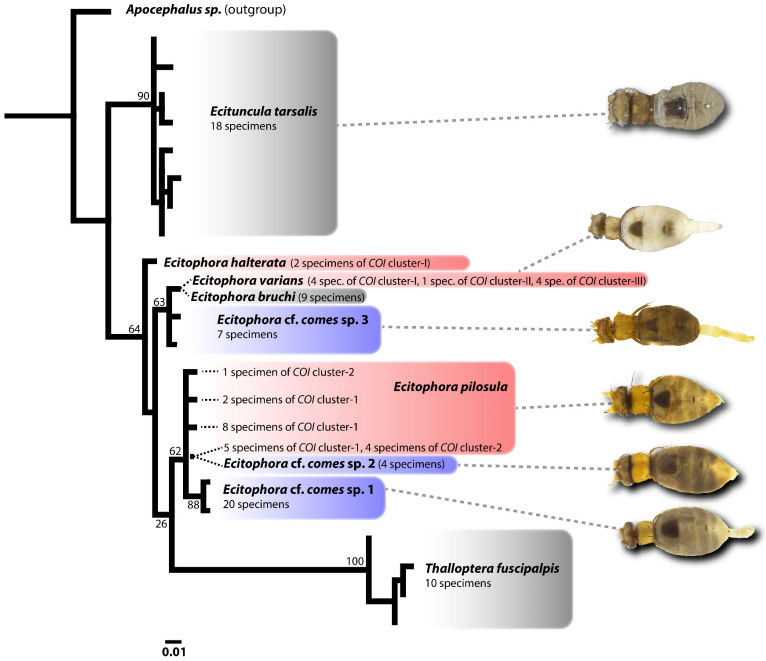
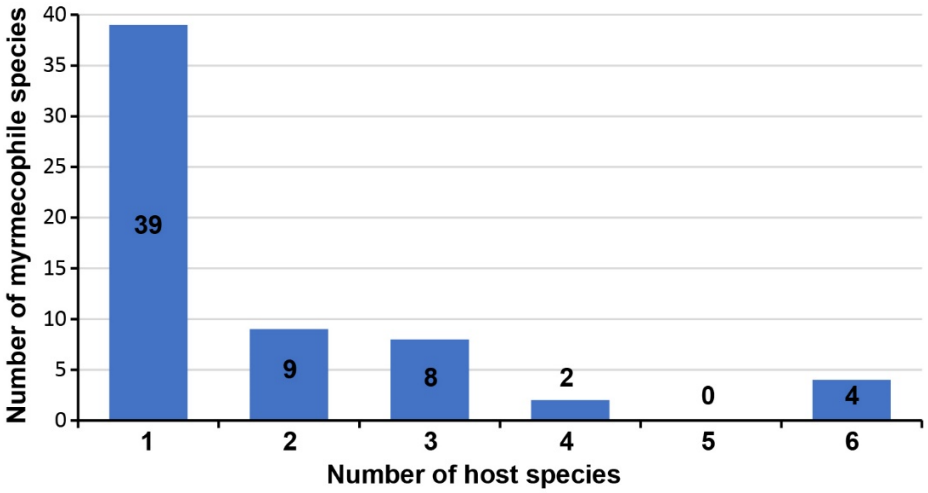
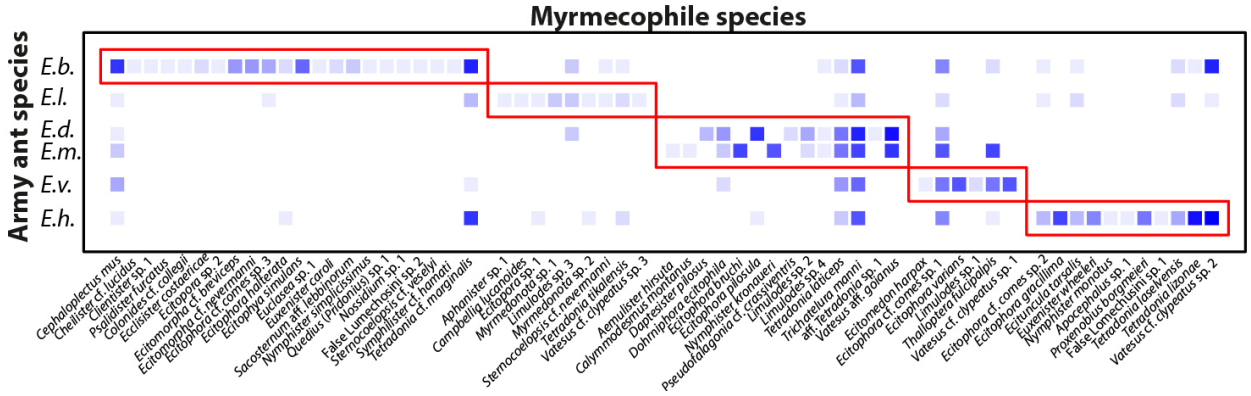


Figure S2. Host range distribution. Histogram visualizing the host distribution range of myrmecophile species of *Eciton* army ants at La Selva Biological Station. The mean number of host species per myrmecophile species was 1.82. See also Table 2 for sample sizes.



MOLECULAR ECOLOGY

Figure S3. Interaction matrix and network modularity of the entire community. Blue squares depict existing associations. Darker blue shading indicates higher link strengths. Modules as detected by QuanBiMo are shown as red boxes (Dormann & Strauss, 2014). Abbreviation of host species are the same as in Table 2. This network is based on 62 myrmecophile species, 2,113 myrmecophile specimens, and 70 *Eciton* colonies (Table 2; network matrix in Table S1).



MOLECULAR ECOLOGY

Table S1. Specimen collection information, GenBank accession numbers, and interaction network matrix. The file can be downloaded as supplementary material on the journal's webpage.

MOLECULAR ECOLOGY

Table S2. PCR primer combinations used in this study. Primer combinations with successful amplification are given for each genus, with most reliable combinations highlighted in bold. Annealing temperatures varied between 45°C and 62°C. *CAD* primer combinations for *Vatesus* and *Tetradonia* beetles were published previously (von Beeren, Maruyama & Kronauer, 2016a,b).

Genus	*COI* and *wg* primer combinations (forward primer/reverse primer)
Aemulister	*COI*: **dgLCO1490/dgHCO2198**
Aphanister	*COI*: **LCO1490/HCO2198**, dgLCO1490/dgHCO2198, LCO_Ecc2/HCO_Ecc1, LCO_Ecc_Nym1/HCO_Ecc1
Apocephalus	*COI*: **LCO1490/HCO2198**, *wg*: wg550F/wgAbr
Calymmodesmus	*COI*: **LCO_milli**/HCO2198, LCO1490/HCO2198, dgLCO1490/dgHCO2198
Campbellia	*COI*: **LCO1490/HCO2198**
Cephaloplectus	*COI*: **LCO1490/HCO2198**, LCO-ce002/HCO2198, dgLCO1490/dgHCO2198, LCO1490/HCO_ce001
Cheilister	*COI*: **LCO_Ecc_Nym1/HCO_Ecc1**, LCO_Ecc_Nym1/HCO2198, LCO_Ecc2/HCO_Ecc2
Clientister	*COI*: **LCO_Ecc_Nym1/dgHCO2198**, LCO1490/dgHCO2198
Colonides	*COI*: **dgLCO1490/dgHCO2198**
Daptesister	*COI*: **LCO1490/HCO2198**, LCO_Ecc_Nym1/HCO2198,
Dinocoryna	*COI*: **LCO1490/HCO2198**
Dorniphora	*COI*: **LCO1490/HCO2198**, dgLCO1490/dgHCO2198, LCO_pho1/HCO2198
Ecclisister	*COI*: **LCO1490/HCO2198** dgLCO1490/HCO2198
Ecitodonia	*COI*: **LCO1490/HCO2198**
Ecitomedon	*COI*: **LCO1490/HCO2198**
Ecitomorpha	*COI*: **LCO1490/HCO2198**, **LepF1/LepR1**, dgLCO1490/dgHCO2198, LCO1490-JJ/HCO2198-JJ, dgLCO1490/CrematoR1, Eciton_F4/Eciton_R4 *wg*: **Wg578F_Tetra/WgAbrZ**, Wg550F/WgAbrZ, Wg578F_Tetr_ecbi/ WgAbrZ, Wg578F_Tetra/WgAbr
Ecitophora	*COI*: **LCO1490/HCO2198**, dgLCO1490/dgHCO2198, LCO1490/dgHCO2198, LepF1/LepR1, LCO_pho3/HCO2198, dgLCO1490/CrematoR1, MLepF1/LepR1 *wg*: **Wg550F/WgAbr**, Wg550F/WgAbrZ
Ecitophya	*COI*: **LCO1490/HCO2198**, dgLCO1490/CrematoR1, Eciton_F4/Eciton_R4
Ecituncula	*COI*: **LCO1490/HCO2198**, dgLCO1490/dgHCO2198, LepF1/LepR1, MLepF1/LepR1; *wg*: **Wg550F/WgAbr**
Euclasea	*COI*: LCO_Ecc2/ HCO2198
Euxenister	*COI*: **LCO1490/HCO2198**, LCO_Ecc1/HCO2198, Eciton_F4/Eciton_R4, LCO_eux001/HCO2198, LCO_Ecc_Nym1/dgHCO2198
False-Lomechusini sp. 1	*COI*: **LCO1490/HCO2198**
False-Lomechusini sp. 2	*COI*: **LCO1490/HCO2198**
Limulodes	*COI*: **LCO1490/HCO2198**, dgLCO1490/HCO2198, LCO1490/HCO2198, LepF1/LepR1
Myrmedonota	*COI*: **LCO1490/HCO2198**, dgLCO1490/dgHCO2198, LCO1490/dgHCO2198, dgLCO1490/HCO2198
Nymphister	*COI*: **LCO1490/HCO2198**, dgLCO1490/HCO2198, dgLCO1490/CrematoR1, LCO_Ecc_Nym1/HCO2198, LCO_Ecc2/dgHCO2198, LCO1490/dgHCO2198
Proxenobius	*COI*: **LCO_grst1/ HCO2198**, LCO_grst2/ HCO2198

Genus	*COI* and *wg* primer combinations (forward primer/reverse primer)
Pseudofalagonia	*COI*: **LCO1490/HCO2198**
Quedius (subgenus *Pridonius*)	*COI*: **MLepF1/LepR1**, LCO1490/HCO2198, dgLCO1490/dgHCO2198, LepF1/LepR1, dgLCO1490/CrematoR1
Sacosternum	*COI*: **LCO1490/HCO2198**, LCO_Ecc_Nym1/dgHCO2198, LCO_Ecc2/HCO_Ecc1, LCO_Ecc2/HCO2198
Sternocoelopsis	*COI*: **LCO1490/HCO2198**, LCO_Ecc_Nym1/HCO2198
Symphilister	*COI*: **LCO1490/HCO2198**
Tetradonia	*COI*: **LCO1490/HCO2198**, **LepF1/LepR1**, LCO_Tetr/HCO2198, dgLCO1490/CrematoR1, Eciton_F4/Eciton_R4, LepF1/LepR1, dgLCO1490/dgHCO2198, LCO1490/dgHCO2198 *wg*: **Wg578_Tetra/WgAbrZ**, Wg550F/WgAbrZ, Wg550F/WgAbr, , Wg578F/WgAbrZ
Thalloptera	*COI*: **LCO1490/HCO2198**, dgLCO1490/dgHCO2198 *wg*: **Wg550F/WgAbr**, Wg550F/WgAbrZ
Trichatelura	*COI*: **LCO1490/HCO2198**, LCO1490/dgHCO2198, dgLCO1490/CrematoR1, dgLCO1490/dgHCO2198 *wg*: **Wg550F/WgAbrZ**
Vatesus	*COI*: **LCO1490/HCO2198**, LepF1/LepR1, dgLCO1490/CrematoR1, Eciton_F4/Eciton_R4, LepF1/MLepR1, Vablock05_F/HCO2198, dgLCO1490/dgHCO2198, MLepF1/LepR1, LCO1490/dgHCO2198, dgLCO1490/HCO2198 *wg:* **Wg550F/WgAbrZ**, Wg550F/WgAbr

MOLECULAR ECOLOGY

Table S3. PCR primers used in this study. Primers used to amplify *CAD* were published previously (von Beeren et al., 2016a,b).

Primer name	Locus	Reading direction	Primer sequence (5' –3')	Source
COI_Eciton_F4	*COI*	forward	CHGGWGCWGGWACAGGATGAACAGT	this study
COI_Eciton_R4	*COI*	reverse	AGTATAGTRATWGCHCCYGCTARWACTGG	this study
CrematoR1	*COI*	reverse	GGRTCTCCYCCTCCDGMDGGRTC	Hoenle et al., 2019
dgHCO2198	*COI*	reverse	TAAACTTCAGGGTGACCAAARAAYCA	Meyer, 2003
dgLCO1490	*COI*	forward	GGTCAACAAATCATAAAGAYATYGG	Meyer, 2003
HCO_ce002	*COI*	reverse	AATAAATGTTGNTATAAAATAGGNT	this study
HCO_Ecc1	*COI*	reverse	AAWAGRTGTTGRTATARAATAGGGTC	this study
HCO2198	*COI*	reverse	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al., 1994
HCO2198-JJ	*COI*	reverse	AWACTTCVGGRTGVCCAAARAATCA	Astrin & Stüben, 2008
LCO_ce2	*COI*	forward	AACCTTATATTTTATTTTTGGAGCCT	this study
LCO_Ecc_Nym1	*COI*	forward	AACYTTATAYTTTATCTTTGGRGCTTG	this study
LCO_Ecc1	*COI*	forward	AACYTTATAYTTTATCTTTGGNGCWT	this study
LCO_Ecc2	*COI*	forward	GCAGGAATAGTAGGAACATCTCTTAG	this study
LCO_Euxe	*COI*	forward	ACYTTRTAYTTYATCTTYGGWGCATGAGCC	this study
LCO_milli	*COI*	forward	AACTTTGTATTTGATTTTTGGTTCTTG	this study
LCO_pho1	*COI*	forward	WWCHYTWTAYTTYATYTTYGGDKCWTGRGC	this study
LCO_Tetr	*COI*	forward	TATTTYATCTTTGGAAGATGRGCAG	this study
LCO1490	*COI*	forward	GGTCAACAAATCATAAAGATATTGG	Folmer et al., 1994
LCO1490-JJ	*COI*	forward	CHACWAAYCATAAAGATATYGG	Astrin & Stüben, 2008
LCOgrst_01	*COI*	forward	TTTATATTTCATTTTCGGTTCATGG	this study
LCOgrst_02	*COI*	forward	GAATAGTAGGAACTTCCCTC	this study
LepF1	*COI*	forward	ATTCAACCAATCATAAAGATATTGG	Hebert et al., 2004
LepR1	*COI*	reverse	TAAACTTCTGGATGTCCAAAAAATCA	Hebert et al., 2004
MLepF1	*COI*	forward	GCTTTCCCACGAATAAATAATA	Hajibabaei et al., 2005
MLepR1	*COI*	reverse	CCTGTTCCAGCTCCATTTTC	Hajibabaei et al., 2006
Vablock05_F	*COI*	forward	ACTTATTCGTGCTGAAYTAGGAAA	this study
Wg550F	*wg*	forward	ATGCGTCAGGARTGYAARTGYCAYGGYATGTC	Wild & Maddison, 2008
Wg578F	*wg*	forward	TGCACNGTGAARACYTGCTGGATG	Wild & Maddison, 2008
Wg578F_Tetr_Ecbi	*wg*	forward	TGCACGGTGAAGACSTGCTGGATG	this study
WG578F_Tetradonia	*wg*	forward	TGCACGGTGAAGACCTGCTGGATG	this study
WgAbR	*wg*	reverse	ACYTCGCAGCACCARTGGAA	Wild & Maddison, 2008
WgAbrZ	*wg*	reverse	CACTTNACYTCRCARCACCARTG	Wild & Maddison, 2008

MOLECULAR ECOLOGY

Supplemental references

Astrin, J.J. & Stüben, P.E. (2008). Phylogeny in cryptic weevils: molecules, morphology and new genera of western Palaearctic Cryptorhynchinae (Coleoptera: Curculionidae). *Invertebrate Systematics*, 22:503–522. doi: 10.1071/IS07057

Dormann, C.F. & Strauss, R. (2014). A method for detecting modules in quantitative bipartite networks. *Methods in Ecology and Evolution*, 5:90–98. doi: 10.1111/2041-210X.12139.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial *cytochrome c oxidase subunit I* from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3:294–299.

Hajibabaei, M., Dewaard, J.R., Ivanova, N.V., Ratnasingham, S., Dooh, R.T., Kirk, S.L., … Hebert, P.D. (2005). Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360:1959–1967.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W. & Hebert, P.D.N. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America*, 103:968–971.

Hebert, P.D.N, Penton, E.H., Burns, J.M., Janzenm D,H. & Hallwachsm W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, 101:14812–14817. doi: 10.1073/pnas.0406166101.

Hoenle, P.O., Blüthgen, N., Brückner, A., Kronauer, D.J.C., Fiala, B., Donoso, D.A., … von Beeren, C. (2019). Species-level predation network uncovers high prey specificity in a Neotropical army ant community. *Molecular Ecology*, 28:2423–2440.

Meyer, C.P. (2003). Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. *Biological Journal of the Linnean Society*, 79:401–459.

Seevers, C.H. (1965). The systematics, evolution and zoogeography of staphylinid beetles, associated with army ants (Coleoptera, Staphylinidae). *Fieldiana Zoology*, 47:137–351.

von Beeren, C., Brückner, A., Maruyama, M., Burke, G., Wieschollek, J. & Kronauer, D.J.C. (2018). Chemical and behavioral integration of army ant-associated rove beetles–a comparison between specialists and generalists. *Frontiers in Zoology*, 15:8.

von Beeren, C., Maruyama, M. & Kronauer, D.J.C. (2016a). Cryptic diversity, high host specificity and reproductive synchronization in army ant-associated *Vatesus* beetles. *Molecular Ecology*, 25:990–1005. doi: 10.1111/mec.13500.

von Beeren, C., Maruyama, M. & Kronauer, D.J.C. (2016b). Community sampling and integrative taxonomy reveal new species and host specificity in the army ant-associated beetle genus *Tetradonia* (Coleoptera, Staphylinidae, Aleocharinae). *PLoS ONE*, 11:e0165056.

Wild, A.L. & Maddison, D.R. (2008). Evaluating nuclear protein-coding genes for phylogenetic utility in beetles. *Molecular Phylogenetics and Evolution*, 48:877–891. doi: 10.1016/j.ympev.2008.05.023.