SUPPORTING INFORMATION

Partner choice and fidelity stabilize coevolution in a Cretaceous-age defensive symbiosis

Martin Kaltenpoth[a,b,1], Kerstin Roeser-Mueller[b], Sabrina Koehler[a], Ashley Peterson[c,d], Taras Y. Nechitaylo[a], J. William Stubblefield[e], Gudrun Herzner[b], Jon Seger[d], Erhard Strohm[b]

[a]Insect Symbiosis Research Group, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany.

[b]Department of Zoology, University of Regensburg, 93040 Regensburg, Germany.

[c]Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112, USA.

[d]Department of Biology, University of Utah, Salt Lake City, UT 84112, USA.

[e]Fresh Pond Research Institute, Cambridge, MA 02140, USA.

[1]To whom correspondence should be addressed. Email: mkaltenpoth@ice.mpg.de.

Contents

DNA extraction, PCR and sequencing of host genes

DNA was extracted either from insect thoraces or, to allow for later morphological determination of single specimens, from three legs. The MasterPure™ Complete DNA and RNA Purification Kit (Epicentre, Madison, WI, USA) was used for DNA isolation according to the manufacturer's instructions. PCR amplifications were performed on a TGradient Thermocycler (Biometra, Göttingen, Germany), in final reaction volumes of 12.5 µl, composed of 1 µl genomic DNA extract, 1 µl of each primer (10 µM), 1.5 µl dNTP-Mix (2 mM; Fermentas, St. Leon-Rot, Germany), 1.25 µl Peqlab reaction buffer (200 mM Tris-HCl (pH 8.55 at 25 °C), 160 mM $(NH_4)_2SO_4$, 0.1% Tween 20, 20 mM $MgCl_2$) and 0.5 units SAWADY Taq DNA polymerase (Peqlab, Erlangen, Germany). Cycle parameters were as follows: 3 min initial denaturation at 94°C, followed by 35 cycles of 94°C for 40 sec, the primer-specific annealing temperature for 40 sec, 72°C for 40 sec (or 90 sec for longer fragments), and a final extension of 4 min at 72°C. Primer sequences and references are listed in Table S2, details on primer combinations, annealing temperatures and the corresponding fragment lengths are summarised in Table S3. Prior to sequencing, PCR products were purified with the peqGOLD MicroSpin Cycle-Pure Kit (Peqlab Biotechnologie GmbH, Erlangen, Germany) following the manufacturer's protocol. Sequencing was done commercially at Seqlab Sequence Laboratories (Göttingen, Germany).

Partial sequences of six different genes were obtained, all of which have previously been shown to be useful for phylogenetic analyses in Hymenoptera (1-4): A fragment of the subunit 1 of the mitochondrial cytochrome oxidase gene (*coxI*; 841 bp) was amplified and sequenced, as well as a fragment of the ribosomal 28S gene (*28S*; 865 bp). Additionally, the following four single-copy nuclear genes were used: Wingless (*wnt*, comprising of 378 bp cds), long-wavelength rhodopsin (*lwrh*, comprising of 608 bp of cds and 156 bp ncs), arginine kinase (*argK*, with 825 bp cds and 111 bp ncs), and elongation factor 1α (*ef1a*, including 1,041 bp cds and 696 bp ncs). The listed fragment lengths are those of the processed sequences used for the phylogenetic analyses. Primer sequences and PCR conditions for amplification of the host genes are given in Tables S2 and S3. Outgroup sequences for *Apis*, *Bembix,* and *Bicyrtes* could be obtained from the NCBI database. Accession numbers for all sequences are given in Table S1.

Reconstruction of the host phylogeny

Sequences were aligned using BioEdit 7.0.5.3 (5) and SeaView 4.2.6 (6). All alignments were checked and corrected manually. Open reading frames and intron / exon boundaries were identified by comparison with published coding sequences for *Apis mellifera* (*lwrh*: BK005514.1; *argK* AF023619.1; *ef1a*: NM_001014993.1) or via a blast search against non-redundant sequences in the NCBI database. As substitution rates and patterns can differ greatly between coding (cds) and non-coding sequences (ncs), we split the dataset into nine partitions: *28S*, *coxI*, *wnt*, *lwrh*-cds, *lwrh*-ncs, *argK*-cds, *argK*-ncs, *ef1a*-cds, and *ef1a*-ncs. Due to high substitution rates, the non-coding sequences could only be reliably aligned within the Philanthini species. Therefore, we coded the intron sequences of all outgroup taxa as missing data and thus excluded them from the analyses.

In a first step, we reconstructed nine separate gene trees using fast likelihood inferences with the software RAxML v7.0.4 (7-9) corresponding to the nine partitions determined above. Maximum likelihood (ML) searches were conducted with the rapid hill-climbing algorithm (7) under the General Time-Reversible model with four gamma parameters GTR+G (10-12). Support values (100 bootstrap steps) were calculated for each node and topologies were manually compared among the gene trees. Because none of the strongly supported nodes were different, we combined all loci in one supermatrix.

Additionally, searches for a saturation effect within one of the three codon positions were conducted for the genes *wnt*, *coxI*, *lwrh*, *argK*, and *ef1a* by calculating homoplasy indices (HI) for each codon position and gene separately. The software PAUP* 4.0 beta (13) was used for these analyses. The homoplasy index of the third codon position of the genes *coxI* and *lwrh* (HI(*coxI*)=0.66, HI(*lwrh*)=0.46) were higher compared to the first and second positions (HI: CO-1st=0.52, *coxI*-2nd=0.25, *lwrh*-1st=0.34, *lwrh*-2nd=0.18). Therefore, we excluded the third codon positions of the genes *coxI* and *lwrh* from further analyses, or we used the translated amino acid sequences (stated for each analysis).

In a next step, multiple independent analyses with different data partitioning strategies (1-4) were performed to test for the robustness of the phylogenetic reconstructions: (1) unpartitioned, (2) four partitions with combined nuclear introns, exons and mitochondrial sequences separately, plus *28S* sequences, (3) nine partitions with single genes separately and splitting coding and non-coding sequence parts, (4) complete random partitioning in 9 partitions; all analyses were conducted with excluded third codon positions of the genes *coxI* and *lwrh* and also with base sequences translated into amino acid sequences. The best fitting evolutionary model for the amino acid-translated sequences (*coxI*, *lwrh*) was inferred with ProtTest v1.4 (14). The CPREV model showed the highest fit for *lwrh*, and the MTREV for *coxI*. From these different runs, we chose the tree with the highest likelihood for presentation. Bootstrap support values were obtained through a full non-parametric bootstrap inference with 10,000 replicates, carried out separately with RAxML.

Bayesian inferences were run with the program MrBayes 3.1.2 (15-17). The searches were also conducted under the GTR+G model with four rate categories. We ran each analysis for 10,000,000 generations and sampled trees every 1,000 generations. We checked if the standard deviation of split frequencies was consistently less than 0.01, and we used a "Burnin" of 20%, i.e. the first 20% of the sampled trees were discarded. We computed 50% majority rule consensus trees for each analysis with posterior probability values for every node. Different partition schemes (1-4) were analyzed as well (see above). However, mixed data sets consisting of DNA and protein sequences cannot be analyzed in MrBayes, so only nucleotide sequences were used, and third codon positions were excluded for *coxI* and *lwrh*.

Equal weighted maximum-parsimony (MP) analyses were performed using the program PAUP* 4.0 beta (13). We used a heuristic search and TBR (tree-bisection-reconnection) for branch swapping. Bootstrap supports were obtained from 1,000 independent replicates. The third codon positions of *coxI* and *lwrh* were excluded for all MP analysis as well. Further, MP analyses were only conducted for the partition schemes (1) and (3). Since all three analyses (ML, Bayesian, and MP) yielded very similar tree topologies, the results were combined for visualization (Fig. S1).

Dating of the host phylogeny

Divergence time estimations were inferred using BEAST v1.7.5 (18). MCMC analyses with HKY and GTR nucleotide substitution models (empirical or estimated base frequencies, various site heterogeneity models [none, G, I+G]) were conducted under a strict clock (using a single rate of sequence evolution across the phylogeny) and an uncorrelated lognormal relaxed clock model (allowing variable substitution rates; 19). In each analysis, 25 million steps were performed, and trees were sampled every 2,500 steps. To estimate the influence of partitioning, analyses were conducted with the partitioned (9 gene partitions, codon partitioning (1+2, 3) for *argK, ef1a* and *wnt*) as well as with the unpartitioned dataset (3rd codon positions excluded for *coxI* and *lwrh* in both datasets due to saturation). The phylogenetic tree from the ML analysis (see previous section) was used as the starting tree in all analyses. In some of the analyses, the starting tree was fixed by removing the tree priors from the BEAST input file (see Table S4).

Four calibration points were included in the initial dating analysis: (A) The age of the Bembicinae with oldest fossils known from Florissant beds in Colorado (*Psammaecius sepultus,* originally described as *Hoplisus sepultus* by Cockerell (20), reviewed by Pulawski and Rasnitsyn (21) and transferred to the extant bembicin genus *Psammaecius*), which date back to the latest Eocene (~34.1 Mya) (22), (B) the age of the oldest *Cerceris* fossil from late Stampian (*Cerceris berlandi*, ~30 Mya) shales in France (23), (C) the age of the oldest *Philanthus* fossils from Colorado (*Philanthus saxigenus* and *Prophilanthus destructus,* ~34.1 Mya) (24, 25), and (D) the root age was calibrated based on earlier phylogenetic analyses (26, 27). Minimum age constraints for the

Bembicinae and the *P. saxigenus* fossil were modelled with lognormal distributions (mean±SD=34.1±0.5, offset=20.0 for both fossils). The age of the *Cerceris* fossil was used to place a hard lower boundary on the age of the Cercerini+Aphilanthopini clade (uniform distribution, minimum=30.0, maximum=1,000.0; or lognormal distribution with mean±SD=30.0±0.75, offset=20.0). As the phylogenetic relationship of Crabronidae subfamilies and bees ("Apidae" *sensu lato*) is still controversial (26, 28, 29), we did not enforce monophyly of the Crabronidae (Philanthinae+Bembicinae).

Five compression fossils described from 1906 to 1944 have been assigned to the Philanthinae by earlier authors: *Prophilanthus destructus* (20), *Philanthus saxigenus* (24), *Philoponites clarus* (30), *Philanthus annulatus* (25), and *C. berlandi* (23). No recent publication has reviewed the systematic affinities of these specimens. In our view, only the *Cerceris* specimen is clearly assignable to Philanthinae. Timon-David's (23) description and illustration leave no doubt that the specimen belongs to the philanthine tribe Cercerini. When it comes to the other four specimens, however, no structures are described or illustrated that would convincingly associate them with the Philanthinae, much less the tribe Philanthini. Therefore, the dating analyses were also repeated excluding the *Philanthus* fossil calibration point.

The root of the tree was modelled with a normal distribution with mean±SD=140.0±10.0, since both the divergence of Sphecidae from the other Apoidea and that of Crabronidae and bees have been estimated to the period of 130-150 Mya (26, 27). This time period coincides with the estimated rise of the angiosperms. Due to their tight association with angiosperms, bees and crabronid wasps have likely evolved with or after the origin of angiosperms (26, 31). However, to assess the effect of root age on the divergence estimates, we additionally performed analyses without a root age prior.

Evaluation and comparison of the models was performed using Tracer v1.5 (32). Bayes factors (BF) were computed for comparison of marginal likelihood values, and $\log_{10}BF>100$ were interpreted as decisive evidence for differences in model performance. A summary of model parameters and results of the dating analyses are given in Table S4. For visualization of the results, the maximum clade credibility tree was inferred with TreeAnnotator (18), using a burnin of 1,000 and a posterior probability limit of 0.5. The consensus tree was visualized with FigTree v1.3.1 (33), including highest posterior density (HPD) intervals (Fig. 1 and S2-S4). Due to the unclear systematic position of the putative Philanthini fossils, the analyses excluding this calibration point were displayed (Fig. 1 and S2-S4). It should be noted, however, that the analyses including the Philanthini fossils yielded identical tree topologies and very similar age estimates (see Table S4).

Among the tested evolutionary models (GTR, GTR+I+G, HKY, HKY+G, HKY+I+G), assessment of convergence and Tracer v1.5 (32) evaluation of Bayes factors revealed the HKY+G and HKY+I+G as the best models (Table S4). Across all models, the partitioned dataset (nine gene partitions, and codon partitioning [1+2, 3] for *argK, ef1a* and *wnt*) consistently yielded better likelihood scores than the non-partitioned dataset, and the uncorrelated lognormal relaxed clock model outperformed the strict clock model. Despite some minor topological discrepancies within the Philanthini (i.e. the placement of *Trachypus boharti, Philanthus albopilosus*, and *Philanthus ventilabris*, see Fig. S2 and S4), both HKY+G and HKY+I+G models consistently yielded age estimates of 64.7 to 68.7 Mya (lower boundary) to 102.0 to 107.5 Mya (upper boundary) for the age of the beewolf-*Streptomyces* symbiosis, regardless of whether the input tree was fixed to the ML input tree or not. Furthermore, estimates for the symbiosis age changed only slightly when a uniform distribution was used to model the ancestral age of the Cercerini+Aphilanthopini instead of a lognormal distribution, or when the putative *Philanthus* and *Prophilanthus* fossils or the root calibration was omitted, respectively (60.1 to 68.3 Mya for the lower and 92.3 to 110.5 for the upper boundary). Omitting both calibration points, however, resulted in low performance of the HKY+G model and yielded considerably lower age estimates for the symbiosis (39.4 to 56.3 Mya for the lower and 62.8 to 86.4 Mya for the upper boundary).

DNA extraction, PCR and sequencing of CaSP genes

Genomic DNA was extracted from whole beewolf antennae according to a standard phenol-chloroform extraction protocol (34) or with the MasterPure™ Complete DNA and RNA Purification Kit (Epicentre Biotechnologies) according to the manufacturer's instructions. The presence of '*Candidatus* Streptomyces philanthi' in the antennae was confirmed by diagnostic PCR using the specific 16S rDNA primer Strep_phil_fwd3 in combination with the general actinomycete primer Act-A19 as described earlier (35). Almost complete 16S rDNA sequences of many '*Ca*. S. philanthi' ecotypes had already been sequenced earlier (35-37). The 16S rDNA of additional specimens was amplified with the primers fD1 and Spa-2R, and sequenced bi-directionally with fD1 and rP2 (Tables S2 and S3). PCR amplifications were performed on a Biometra® T-Gradient Thermocycler or on a VWR Gradient Thermocycler in a total reaction volume of 25 µl containing 2 µl of template, 1x PCR buffer (10 mM Tris-HCl, 50 mM KCl, 0.08% Nonidet P40), 2.5 mM $MgCl_2$, 240 µM dNTPs, 20 pmol of each primer, and 1 U of Taq DNA polymerase (MBI Fermentas). Cycle parameters were as follows: 3 min. at 94°C, followed by 32 cycles of 94°C for 40 sec., 65°C for 1 min., and 72°C for 1 min., and a final extension time of 4 min. at 72°C.

Parts of the elongation factor Tu and the elongation factor G as well as the intergenic spacer region (collectively referred to as *fus-tuf* in the following) of '*Ca*. S. philanthi' were amplified by using the primer pairs EF-Tu-1F/EF-Tu-2R and EF-Tu-3F/EF-Tu-3R, respectively, and sequenced using the same primers (Tables S2 and S3). The primer pairs gyrB-F1/gyrB-R3 and gyrB-F3/gyrB-R10 amplified overlapping fragments of the gyrase B gene (*gyrB*) of the endosymbionts that could be sequenced by using the same primers. Additionally, a 627 bp fragment of gyrase A (*gyrA*) was amplified using primers gyrA-5F/gyrA-5R and sequenced unidirectionally using primer gyrA-5F (Tables S2 and S3). PCR reaction mixtures were the same as described for the amplification of the 16S rDNA. Cycle parameters were as follows: 3 min. at 94°C, followed by 35 cycles of 94°C for 40 sec., 65°C (*fus-tuf* primers) or 62°C (*gyrB* primers) or 60°C (*gyrA* primers) for 40 sec., and 72°C for 40 sec., and a final extension time of 4 min. at 72°C. Sequencing was done in the Department of Entomology at the Max Planck Institute for Chemical Ecology (Jena, Germany) or commercially by SEQLAB Sequence Laboratories (Göttingen, Germany).

Symbiont phylogenetic analysis

For the phylogenetic analysis, 16S rRNA, *gyrA*, *gyrB*, and *fus-tuf* sequences of all *Streptomyces* species for which fully sequenced or good draft genomes were available were retrieved from the NCBI database. Additionally, cultures of three closely related strains (based on 16S rRNA, *Streptomyces ramulosus* DSM 40100*, Streptomyces abikoensis* DSM 40831, and *Streptomyces mutabilis* DSM 40169) were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany), and the four gene fragments were sequenced as described above.

All protein-coding sequences were assembled and aligned based on their translated amino acid sequences using Geneious Pro 5.4 (38). The 16S rRNA gene sequences were imported into ARB and aligned against closely related *Streptomyces* sequences based on the secondary structure prediction (39). The alignments were concatenated in BioEdit 7.0.5.3 (5). The concatenated alignment consisted of a total of 4653 bp (1391 bp of 16S rDNA, 639 bp of *fus*, 930 bp of *tuf*, 249 bp of *fus-tuf* intergenic spacer, 765 bp of *gyrB*, and 549 bp of *gyrA*). Accession numbers for all symbiont and other Actinobacteria sequences are given in Tables S6 and S7, respectively.

Approximately-maximum-likelihood trees were reconstructed with FastTree 2.1 using the GTR model (40). Local support values were estimated with the Shimodaira-Hasegawa test based on 1,000 resamples without reoptimizing the branch lengths for the resampled alignments (40). Additionally, a maximum likelihood tree was reconstructed using PHYML (41) as implemented in Geneious Pro 5.4 (38). The GTR+I+G model was chosen, the transition/transversion ratio was set to 4 (fixed), and both the proportion of invariable sites and the gamma distribution parameter were estimated. Bootstrap values were obtained from a search with 1,000 replicates.

Bayesian inferences were run with the program MrBayes 3.1.2 (15-17), with the concatenated alignment split into six partitions: 16S rRNA, *gyrA, gyrB*, *fus, tuf*, and the *fus-tuf* intergenic spacer. The searches were conducted under the GTR+I+G model. We ran each analysis for 20,000,000 generations and sampled trees every 1,000 generations. A "burnin" of 25% was used, i.e. the first 25% of the sampled trees were discarded. We checked if the standard deviation of split frequencies was consistently lower than 0.01. We computed a 50% majority rule consensus tree with posterior probability values for every node. Since the phylogenetic trees reconstructed with the three different methods were topologically very similar, the results were combined into a single figure (Fig. S8).

To gain more comprehensive insights into within-species patterns of symbiont phylogenetic relationships, we sequenced *gyrA* for the symbionts of 109 beewolf individuals across 41 species (for accession numbers see Table S6). We aligned the sequences as described above and used FastTree 2.1 for phylogenetic reconstruction, with the same settings as for the concatenated alignment (Fig. S9).

Detection of opportunistic bacteria in Philanthini antennae

In a few cases, bacteria other than CaSP could be found in the antennae of female beewolves. To assess the incidence of CaSP across beewolf species, 338 specimens of 34 different Philanthini species were screened for the presence of CaSP by diagnostic PCR using the specific 16S rDNA primer Strep_phil_fwd3 in combination with the general actinomycete primer Act-A19 as described earlier (35). Additionally, since bacteria of the genus *Amycolatopsis* were detected repeatedly, and notably in the only two specimens of *Philanthus* cf. *basalis* investigated, the same specimens were screened for the presence of *Amycolatopsis* by using the specific primer Amy_16S_1F in combination with the actinobacterial primer Act-A19 (Table S3). Antennal specimens that were negative for both CaSP and *Amycolatopsis* were tested with the general actinobacterial primer pair Act-S20/Act-A19 (Table S3) and subsequently with the general eubacterial primers EUB933F-GC (5'-CGCCCGCCGCGCGGCGGGCGGGGCGGGGGCAC-GGGGGGCACAAGCGGTGGAGCATGTGG-3') and EUB1387R (5'-GCCCGGGAACGTAT-TCACCG-3') (42, 43). Amplification products of the actinobacterial PCR were sequenced directly, whereas those of the eubacterial PCR were separated by temperature-gradient gel electrophoresis (TGGE) prior to sequencing as described earlier (44). Briefly, TGGE gels (50ml) were prepared with a final concentration of 8% polyacrylamide (60:1), 8M urea, 0.1X TBE buffer and 2% glycerol, and polymerized on polybond films (Biometra) by adding 110µl TEMED (N,N,N',N'-tetramethylethan-1,2-diamine) and 40µl ammoniumpersulfate (50%). After electrophoresis for 18 hours at 150V with a temperature gradient from 40°C to 50°C on a TGGE Maxi System (Biometra), gels were stained with silver nitrate as described previously (44). Bands were excised using a sterile scalpel, and the DNA was re-eluted overnight at 4°C in 50µl LowTE buffer (1mM Tris, 0.1mM EDTA). Excised bands as well as amplicons from the *Amycolatopsis*- and actionobacterial PCRs were sequenced and compared against the NCBI database using BLASTn.

Diagnostic PCRs for *T. boharti* antennae consistently yielded positive results for both CaSP and *Amycolatopsis*. *Amycolatopsis* PCR products were sequenced and turned out to stem from CaSP, indicating that the Amy_16S_1F primer successfully amplified the *T. boharti* CaSP strain despite two mismatches in the primer binding site (as opposed to 3-5 mismatches for all other CaSP strains, 1-2 of which are located towards the 3'-end of the primer). Hence*, T. boharti* specimens that yielded positive PCRs for both CaSP and *Amycolatopsis* were assumed to harbor pure cultures of CaSP, and only CaSP-negative specimens were subsequently screened with *Amycolatopsis*, general actinobacterial, and general eubacterial primers.

Sequences of actinobacterial 16S rRNA from beewolf antennae (NCBI accession numbers KC607731-KC607747) were aligned to the SILVA-ARB SSU database (45) using the SINA aligner (46) and imported into ARB (39). The most closely related strains for each beewolf isolate as well as representative strains of the actinobacterial genera containing isolates were selected for phylogenetic analysis. Furthermore, CaSP strains were included as a reference. The alignment was exported from ARB, and an approximately-maximum-likelihood tree was reconstructed with FastTree 2.1 using the GTR model (40). Local support values were estimated with the Shimodaira-Hasegawa test based on 1,000 resamples without reoptimizing the branch lengths for the resampled alignments (40) (Fig. S5).

Two sequences from antennae of *Philanthus triangulum* and one from *T. boharti* were assigned by BLAST to Proteobacteria (*Serratia* and *Wolbachia*) or Tenericutes (*Spiroplasma*) (NCBI accession numbers KF922849-KF922851). As these sequences probably represent systemic infections of the hosts, including the antennal hemolymph, rather than specialized colonization of the antennal gland reservoirs, they were excluded from phylogenetic analyses.

Localization of *Amycolatopsis* in the antennal gland reservoirs of *Philanthus* cf*. basalis*

To exclude the possibility of contamination and confirm that the *Amycolatopsis* sequences originated from bacteria within the antennal gland reservoirs of *P*. cf. *basalis*, we performed fluorescence *in-situ* hybridization (FISH) on the second antenna of a *P.* cf. *basalis* individual that was positive for *Amycolatopsis,* based on the PCR results for the first antenna. The antenna was fixated in 95% ethanol, embedded in cold-polymerizing resin (Technovit 8100, Heraeus Kulzer) and used for FISH as described earlier (36, 37). The specific fluorescent probes Cy3-SPT177 (specific for 'Ca. S. philanthi', see 35) and Cy3-Amy_16S (specific to *Amycolatopsis*; complementary to primer Amy_16S_1F) as well as the general eubacterial probe Cy3-EUB338 (47) were used to stain the bacteria within the antennal gland reservoirs (Table S3, Fig. S6). To confirm the specificity of the Cy3-Amy_16S probe, an antenna of a female *P. triangulum* specimen was prepared for FISH in the same way and stained with the same probes (Fig. S7).

Production of AGS by beewolf females with CaSP and with other Actinobacteria, respectively

Field-collected female beewolves were reared in observation cages as described previously (48) and provided with honey and bees *ad libitum*. Freshly constructed brood cells were checked for the presence of the white antennal gland secretion (AGS) containing the symbiotic bacteria. In the observation cages, the AGS is usually visible with the unaided eye after secretion by the female beewolf to the ceiling of the brood cell (48). Six females did not apply AGS to any of their brood cells (AGS-), whereas the AGS was regularly found in brood cells of all other females (AGS+ females) (Table S8). The AGS- and seven randomly selected AGS+ females were sacrificed, and RNA and DNA were extracted from the antennae using the MasterPureTM Complete DNA and RNA Purification Kit (Epicentre, Madison, WI, USA) according to the manufacturer's instructions. DNA extracts were screened with CaSP- (Strep_phil_fwd3/Act-A19) and actinobacteria-specific (Act-S20 /Act-A19) primer. Products from Act-PCRs were sequenced bidirectionally with or withour prior subcloning (using the StrataClone PCR Cloning Kit, Agilent Technologies, La Jolla, CA, USA, according to the manufacturer's instructions). Sequences were aligned with the '*Ca*. S. philanthi triangulum' 16S rRNA sequence to check for similarity and compared with the NCBI database using BLASTn. Sequences that were distinct from CaSP were included in the phylogenetic analyses described above (see Fig. S5).

Detection of CaSP in sand surrounding *P. triangulum* nests

Total DNA was extracted from microorganisms present in the sand of used *P. triangulum* observation cages, following separation by Nicodenz® gradient centrifugation as described previously (49). Briefly, six sand samples (30 g each) were filled up to 50 ml with disruption buffer (0.2 M NaCl, 50 mM Tris-HCl pH 8.0) and thoroughly mixed. Large sand particles were sedimented by centrifugation at 100 × g for five minutes at room temperature. The supernatant was transferred into the tubes with Nicodenz® and cells were separated from sand particles at 10,000 × g for 20 min at 4 °C. Cells were collected from the surface of Nicodenz®, washed three times with PBS and finally, total DNA was extracted with the SoilMaster™ DNA Extraction Kit (Epicentre). The quality of extracted DNA was checked by 1% agarose gel electrophoresis and PCR with the general eubacterial 16S rRNA primers fD1 and rP2. The DNA extracts from the six samples were pooled for bTEFAP.

BTEFAP was done commercially by Research and Testing Laboratory (Lubbock, TX, USA). In total, 8665 reads were generated using primers Gray28F (5'-GAGTTTGATCNTGGCTCAG-3') and Gray519r (5'-GTNTTACNGCGGCKGCTG-3') (50, 51). Generation of the sequencing library was established through one-step PCR with 30 cycles, using a mixture of Hot Start and HotStar high-fidelity Taq polymerases (Qiagen). Sequencing extended from Gray28F, using a Roche 454 FLX instrument with Titanium reagents and procedures at Research and Testing Laboratory (RTL), based upon RTL protocols (http://www.researchandtesting.com). All low-quality reads (quality cut-off = 25) and sequences <200 bp or >600 bp were removed following sequencing, which left 7,123 sequences for subsequent analysis. Processing of the high-quality reads was performed using QIIME (52). The sequences were denoised using the denoiser algorithm (53) and subsequently clustered into operational taxonomic units (OTUs) using multiple OTU picking with cdhit (54) and uclust (55) with 97% similarity cut-offs. For each OTU, one representative sequence was extracted (the most abundant) and aligned to the Greengenes core set (available from http://greengenes.lbl.gov/) using PyNast (56), with the minimum sequence identity per cent set to 75. Taxonomy was assigned using RDP classifier (57), with a minimum confidence to record assignment set to 0.80. For visualization of the results, OTUs were combined based on phylum-level taxonomic affiliation. To assess the number of CaSP reads within the sample, all high-quality sequences were compared to the '*Ca*. S. philanthi triangulum' 16S reference sequence (GenBank accession number DQ375802) by BLAST, and identical sequences were counted using a custom-made Perl script (Fig. S10). This provided a conservative estimate for the number of CaSP sequences in the samples, as it excluded highly similar sequences containing even low numbers of sequencing errors.

Table S1: Collection localities and GenBank accession numbers for beewolf specimens used to reconstruct the host phylogeny.

Species	Specimen no.	Sex	Collection locality	Wingless	NCBI accession number LWRh	EF1a	28s	ArgK	COI
Philanthinus quattuordecimpunctatus	TU-EY-E027	Male	Turkey	JN74246	KF975656	KJ556972	JN674300	JQ083489	JQ040297
Philanthus albopilosus	USA-E56	Male	USA	JN674198	KF975607	KJ556924	JN674251	JQ083441	JQ040264
Philanthus barbatus	USA-E18	Male	USA	JN674199	KF975608	KJ556925	JN674252	JQ083442	JQ040265
Philanthus barbiger	UT-E15	Male	USA	JN674200	KF975609	KJ556926	JN674253	JQ083443	JQ040266
Philanthus basilaris	UT-E6	Male	USA	JN74202	KF975611	KJ556928	JN674255	JQ083445	JQ040268
Philanthus bicinctus	USA-E29	Male	USA	JN74203	KF975612	KJ556929	JN674256	JQ083446	-
Philanthus bilunatus	USA-BS34	Male	USA	JN74204	KF975613	KJ556930	JN674257	JQ083447	-
Philanthus capensis	SA-E62	Male	South Africa	JN74205	KF975614	KJ566218	JN674258	JQ083448	JQ040269
Philanthus cf. basalis	IN-E035	Male	India	JN674201	KF975610	KJ556927	JN674254	JQ083444	JQ040267
Philanthus coarctatus	MO-1	Female	Oman	JN74206	KF975615	KJ556931	JN674259	JQ083449	-
Philanthus coronatus	m1	Male	Germany	JN74207	KF975616	KJ556932	JN674260	JQ083450	JQ040270
Philanthus crabroniformis	USA-E10	Male	USA	JN74208	KF975617	KJ556933	JN674261	JQ083451	JQ040271
Philanthus crotoniphilus	USA-E39	Male	USA	JN74209	KF975618	KJ556934	JN674262	JQ083452	-
Philanthus fuscipennis	SA-E69	Male	South Africa	JN74210	KF975619	KJ556935	JN674263	JQ083453	JQ040272
Philanthus gibbosus	UT-E188	Male	USA	JN74211	KF975620	KJ556936	JN674264	JQ083454	-
Philanthus gloriosus	USA-E60f	Male	USA	JN74212	KF975621	KJ556937	JN674265	JQ083455	JQ040273
Philanthus histrio	SA-E58	Male	South Africa	JN74213	KF975622	KJ556938	JN674266	JQ083456	JQ040274
Philanthus inversus	USA-E53b	Male	USA	JN74214	KF975623	KJ556939	JN674267	JQ083457	-
Philanthus lepidus	CAN-E1	Male	Canada	JN74215	KF975624	KJ556940	JN674268	JQ083458	-
Philanthus loefflingi	SA-E13	Male	South Africa	JN74216	KF975625	KJ556941	JN674269	JQ083459	JQ040275
Philanthus melanderi	SA-E79	Male	South Africa	JN74217	KF975626	KJ556942	JN674270	JQ083460	JQ040276
Philanthus multimaculatus	UT-E76	Male	USA	JN74218	KF975627	KJ556943	JN674271	JQ083461	JQ040277
Philanthus occidentalis	CAL-Eth4	Male	USA	JN74219	KF975628	KJ556944	JN674272	JQ083462	JQ040278
Philanthus pacificus	USA-E19	Male	USA	JN74220	KF975629	KJ556945	JN674273	JQ083463	JQ040279
Philanthus parkeri	UT-E45	Male	USA	JN74221	KF975630	KJ556946	JN674274	JQ083464	JQ040280
Philanthus politus	JS-32a		USA	JN74222	KF975631	KJ556947	JN674275	JQ083465	-
Philanthus psyche	UT-E154/ *JS-A	Male	USA	JN74223	KF975632	KJ556948*	JN674276	JQ083466	JQ040281
Philanthus pulchellus	SP-001	Male	Spain	JN74224	KF975633	KJ556949	JN674277	JQ083467	JQ040282
Philanthus pulcher	USA-E8b	Female	USA	JN74225	KF975634	KJ556950	JN674278	JQ083468	JQ040283
Philanthus pulcherrimus	IN-E064	Male	India	JN74226	KF975635	KJ556951	JN674279	JQ083469	JQ040284
Philanthus rugosus	SA-E23	Male	South Africa	JN74227	KF975636	KJ556952	JN674280	JQ083470	JQ040285
Philanthus rutilus	JS-32			JN74228	KF975637	KJ556953	JN674281	JQ083471	-
Philanthus sanbornii	m28	Male	USA	JN74229	KF975638	KJ556954	JN674282	JQ083472	JQ040286
Philanthus serrulatae	JS-63	Female	USA	JN74230	KF975639	KJ556955	JN674283	-	-
Philanthus solivagus	USA-BS36	Male	USA	JN74231	KF975640	KJ556956	JN674284	JQ083473	-
Philanthus sp. CAL	CAL-Eth14	Male	USA	JN74233	KF975642	KJ556958	JN674286	JQ083475	-
Philanthus sp. IN-E010	IN-E010	Male	India	JN74232	KF975641	KJ556957	JN674285	JQ083474	JQ040287
Philanthus tarsatus	JS-44	Male	USA	JN74234	KF975643	KJ556959	JN674287	JQ083476	-
Philanthus triangulum	N14/ *JS-B	Male	Germany	JN74235	KF975644	KJ556960*	JN674288	JQ083477	JQ040288
Philanthus triangulum diadema	SA-E8	Male	South Africa	JN74236	KF975645	KJ556961	JN674289	JQ083478	JQ040289
Philanthus turneri	SA-E116	Female	South Africa	JN74237	KF975646	KJ556962	JN674290	JQ083479	JQ040290
Philanthus ventilabris	USA-E50	Male	USA	JN74238	KF975647	KJ556963	JN674291	JQ083480	JQ040291
Philanthus venustus	Ph02	Male	Greece	JN74239	KF975648	KJ556964	JN674292	JQ083481	-
Philanthus zebratus	USA-E25	Male	USA	JN74240	KF975649	KJ556965	JN674293	JQ083482	JQ040292
Trachypus boharti	BR-002	Female	Brasil	JN74250	KF975650	KJ556966	JN674294	JQ083483	JQ040293
Trachypus denticollis	JS-11		Chile	JN74241	KF975651	KJ556967	JN674295	JQ083484	-
Trachypus elongatus	BR-E032	Male	Brasil	JN74242	KF975652	KJ556968	JN674296	JQ083485	JQ040294
Trachypus flavidus	BR-E067	Male	Brasil	JN74243	KF975653	KJ556969	JN674297	JQ083486	JQ040295
Trachypus patagonensis	BR-E092	Female	Brasil	JN74244	KF975654	KJ556970	JN674298	JQ083487	JQ040296
Trachypus spec.	JS-52		Chile	JN74245	KF975655	KJ556971	JN674299	JQ083488	-
Aphilanthops foxi	CAL-Eth10	Male	USA	JN74247	KF975657	KJ556973	JN674301	JQ083490	JQ040298
*Bembix amoena/ *B. troglodytes*	-		-	EU367331.1	-	EU367212.1	EU367154.1	-	EF203767.1*
Bicyrtes ventralis	-		-	-	DQ116701.1	AY585161	AY654458.1	-	-
Cerceris rybiensis/Eucerceris	*Cerc1/**Cerc2/***JS-C	Female	Germany/USA	* JN74248	KF975658 **	KJ556974***	AY654460.1	JQ083491**	-
Clypeadon laticinctus	UT-E177/ *BS32a/ **JS-D	Female	USA	JN74249	KF975659 *	KJ556975**	JN674302	JQ083492	JQ040299
Apis mellifera	-		-	AY703618.1	U26026.1	NM_001014993.1	AY703551.1	NM_001011603.1	AF214668.1

Table S2: Primers used for the amplification and sequencing of host and symbiont genes, and probes for the fluorescence in-situ hybridization to detect CaSP and *Amycolatopsis* in beewolf antennae.

	Target organism	Target sequence	Primer/probe name	5'-3' Sequence	Fwd/rev	5'-mod.	Target taxon	Reference
Primers	**Host**	Wingless	beewgfor	TGCACNGTSAAGACCTGYTGGATGAG	fwd	-	Apoidea	Danforth et al. 2004
			Lepwg2a	ACTICGCARCACCARTGGAATGTRCA	rev	-	Apoidea	Brower & DeSalle 1998, Danforth et al. 2004
		LWRh	LWRH_Rev1744	GCDGCTCGRTAYTTHGGATG	rev	-	Philanthinae	this study
			LWRhFor4_N	GAGAARAAYATGCGNGARCAAGC	fwd	-	Philanthinae	this study (modified from Danforth et al. 2004)
			LWRhFor1	AATTGCTATTAYGARACNTGGGT	fwd	-	Apoidea	Mardulyn & Cameron 1999, Danforth et al. 2004
			LWRhRev1	ATATGGAGTCCANGCCATRAACCA	rev	-	Apoidea	Mardulyn & Cameron 1999, Danforth et al. 2004
		EF1a	For1deg	GYATCGACAARCGTACSATYG	fwd	-	Apoidea	Danforth et al. 2003
			F2Rev1	AATCAGCAGCACCTTTAGGTGG	rev	-	Apoidea	Danforth et al. 2003
			HaF2for	GGGYAAAGGWTCCTTCAARTATGC	fwd	-	Apoidea	Danforth et al. 1999
			Cho10	ACRGCVACKGTYTGHCKCATGTC	rev	-	Apoidea	Danforth et al. 2003
		ArgK	ArgK_Loretta	TGATCGATGATCACTTCCTTTTCAA	fwd	-	Philanthinae	this study
			ArgK_fwd2	GACAGCAARTCTCTGCTGAAGAA	fwd	-	Apoidea	Kawakita et al. 2003
			ArgK_KLTrev2	GATKCCATCRTDCATYTCCTTSACRGC	rev	-	Apoidea	www.danforthlab.entomology.cornell.edu/resources.html
		COI / COII	CO_fwd1	TGGAGCHTCWTTYAGATTAATAATYCG	fwd	-	Philanthinae	this study
			CO_rev2	TCCWCCAATWGTRAATAATAARAYA	rev	-	Philanthinae	this study
			CO_LCO	GGTCAACAAATCATAAAGATATTGG	fwd	-	insects	Folmer et al. 1994
			CO_Ben	GCWACWACRTAATAKGTATCATG	rev	-	insects	Kronauer et al. 2004
		28s rRNA	28s_3665F	AGAGAGAGTTCAAGAGTACGTG	fwd	-	Apoidea	Cameron & Mardulyn 2001
			28s_4749R	GTTACACACTCCTTAGCGGA	rev	-	Apoidea	Danforth et al. 2006
	Symbiont	16S rRNA	fD1	AGAGTTTGATCCTGGCTCAG	fwd	-	Eubacteria	Weisburg et al. 1991
			rP2	ACGGCTACCTTGTTACGACTT	rev	-	Eubacteria	Weisburg et al. 1991
			Spa-2R	KTTCGCTCGCCRCTAC	rev	-	Eubacteria	Hain et al. 1997
			Act-S20	CGCGGCCTATCAGCTTGTTG	fwd	-	Actinobacteria	Stach et al. 2003
			Act-A19	CCGTACTCCCCAGGCGGGG	rev	-	Actinobacteria	Stach et al. 2003
			Strep_phil_fwd3	CATGGTTRGTGGTGGAAAGC	fwd	-	*Ca*. S. philanthi	Kaltenpoth et al. 2006
			Amy_16S_1F	CCTGTACTTTGGGATAAGCCT	fwd	-	*Amycolatopsis*	this study
			Amytop_16S_3R	CCTCTGTACCAGCCATTGTAG	rev	-	*Amycolatopsis*	this study
		EF-Tu	EF-Tu-1F	ATYACCAAGGTGCTGCACG	fwd	-	*Ca*. S. philanthi	this study
			EF-Tu-3F	TTCAAGGTCGAGGCCAACG	fwd	-	*Ca*. S. philanthi	this study
			EF-Tu-2R	GCCACCCTCGTCCTTSGAS	rev	-	*Ca*. S. philanthi	this study
			EF-Tu-3R	GCACCGGTGATCATGTTCTT	rev	-	*Ca*. S. philanthi	this study
		gyrB	gyrB-F1	GAGGTCGTGCTGACCGTGCTGCA	fwd	-	*Ca*. S. philanthi	Hatano et al. 2003
			gyrB-F3	TTCGTGAAGTACCTGAACTCG	fwd	-	*Ca*. S. philanthi	this study
			gyrB-R3	SAGCTTGACCGAGATGATCG	rev	-	*Ca*. S. philanthi	this study
			gyrB-R10	CGACTTGCGGATGATGTCC	rev	-	*Ca*. S. philanthi	this study
		gyrA	gyrA-5F	AACCTGCTGGCCTTCCAG	fwd	-	*Ca*. S. philanthi	this study
			gyrA-5R	AACGCCCATGGTGTCACG	rev	-	*Ca*. S. philanthi	this study
Probes	**Symbiont**	16S rRNA	SPT177	CACCAACCATGCGATCGGTA	rev	Cy3 or Cy5	*Ca*. S. philanthi	Kaltenpoth et al. 2005
			Amy_16S	AGGCTTATCCCAAAGTACAGG	rev	Cy3	*Amycolatopsis*	this study
			EUB338	GCTGCCTCCCGTAGGAGT	rev	Cy3	Eubacteria	Amann et al. 1990

Table S3: Primer combinations and PCR conditions used for amplification and sequencing of host and symbiont genes.

Target organism	Target sequence	Forward primer	Reverse primer	PCR cycle number	Annealing temp. (°C)	Fragment length (bp)	Sequencing primers Forward	Sequencing primers Reverse
Host	28s	28s_3665F	28s_4749R	35	62.9	1080	28s_3665F	28s_4749R
	Opsin	LWRhFor1	LWRhRev1	35	58.5	650	LWRhFor1	-
		LWRhFor4_N	LWRH_Rev1744	35	53.8	800	-	Rev1744
		LWRhFor1	LWRH_Rev1744	35	53.8	1200	LWRhFor1	LWRH_Rev1744
	wingless	beewgFor	Lepwg2a	35	65.6	450	beewgFor	-
	ArgK	ArgK_fwd2	ArgK_KLTrev2	35	50.5	1200	ArgK_fwd2	ArgK_KLTrev2
		ArgK_Loretta	ArgK_KLTrev2	35	53.0	700	ArgK_Loretta	-
	COI / COII	CO_LCO	CO_Ben	35	49.0	1100	CO_LCO	CO_Ben
		CO_fwd1	CO_rev2	35	52.8	1000	CO_fwd1	CO_rev2
	EF1a	For1deg	F2Rev1	35	56.8	1300	For1deg	F2Rev1
		HaF2for	Cho10	35	58.0	1700	-	Cho10
Symbiont	16S rDNA	fD1	Spa-2R	32	65.0	2090	fD1	rP2
	EF-Tu	EF-Tu-1F	EF-Tu-2R	35	65.0	870	EF-Tu-1F	-
		EF-Tu-3F	EF-Tu-3R	35	65.0	1220	EF-Tu-3F	EF-Tu-3R
	gyrB	gyrB-F1	gyrB-R3	35	62.0	740	gyrB-F1	-
		gyrB-F3	gyrB-R10	35	62.0	440	-	gyrB-R10
	gyrA	gyrA-5F	gyrA-5R	35	60.0	630	gyrA-5F	-
Diagnostic PCRs	CaSP 16S	Strep_phil_fwd3	Act-A19	35	68.0	684	-	-
	Amy 16S	Amy_16S_1F	Amytop_16S_3R	35	65.0	1108	-	-
	Amy 16S	Amy_16S_1F	Act-A19	35	65.0	742	Amy_16S_1F	Act-A19

Table S4: Model parameters and results of the phylogenetic dating analyses using BEAUti and BEAST. For each analysis, 25 million steps were performed with tree sampling every 2500 steps, and a burnin of 1,000 and a posterior probability limit of 0.5 were used for tree reconstruction.

	Gene	Codon	Substitution	Base			Age priors				Tracer analysis		Age of symbiosis (mya)		Philanthus	Displayed
Goal	**partitions**	**partitioning**	**model**	**frequencies**	**Clock model**	**Starting tree**	**Root**	**Bembicinae**	**Cerceris+Aphil+Clyp [1]**	**Philanthus+Trachypus**	**ESS**	**marg. likelihood**	**lower**	**upper**	**monophyletic?**	**in figure**
Model and	9	(1+2, 3)	GTR	estimated	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	*bad*	*-27614.681*	78.0	109.7	no	
clock choice	9	(1+2, 3)	GTR+I+G	empirical	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	*bad*	-26399.144	65.5	103.0	no	
and	1	no	GTR+I+G	empirical	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	*-27956.219*	67.6	101.9	no	
optimization	9	(1+2, 3)	GTR+I+G	empirical	strict	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	*bad*	-26426.987	44.5	87.3	yes	
	9	(1+2, 3)	HKY	empirical	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	*-27942.536*	80.5	109.3	no	
	9	(1+2, 3)	HKY	empirical	strict	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	*-28078.763*	69.7	107.6	yes	
	9	(1+2, 3)	HKY+G	empirical	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	-26600.493	67.8	103.2	no	
	9	**(1+2, 3)**	**HKY+G**	**estimated**	**relaxed uncorrlogn**	**user-specified (ML)**	**normal: 140 + 10**	**lognormal: 34.1 + 0.5, off: 20**	**lognormal: 30 + 0.75, off: 20**	**lognormal: 34.1 + 0.5, off: 20**	**good**	**-26548.450**	**68.0**	**103.9**	**no**	
	9	**(1+2, 3)**	**HKY+G**	**estimated**	**relaxed uncorrlogn**	**user-specified (ML), fixed**	**normal: 140 + 10**	**lognormal: 34.1 + 0.5, off: 20**	**lognormal: 30 + 0.75, off: 20**	**lognormal: 34.1 + 0.5, off: 20**	**good**	**-26547.967**	**68.7**	**107.5**	**no**	
	9	(1+2, 3)	HKY+I+G	empirical	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	-26475.777	65.8	102.9	no	
	9	**(1+2, 3)**	**HKY+I+G**	**estimated**	**relaxed uncorrlogn**	**user-specified (ML)**	**normal: 140 + 10**	**lognormal: 34.1 + 0.5, off: 20**	**lognormal: 30 + 0.75, off: 20**	**lognormal: 34.1 + 0.5, off: 20**	**good**	**-26425.094**	**64.7**	**102.0**	**no**	
	9	**(1+2, 3)**	**HKY+I+G**	**estimated**	**relaxed uncorrlogn**	**user-specified (ML), fixed**	**normal: 140 + 10**	**lognormal: 34.1 + 0.5, off: 20**	**lognormal: 30 + 0.75, off: 20**	**lognormal: 34.1 + 0.5, off: 20**	**good**	**-26424.060**	**66.1**	**106.0**	**no**	
	1	no	HKY+I+G	empirical	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	*-28075.427*	66.6	100.8	no	
	9	(1+2, 3)	HKY+I+G	empirical	strict	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	*no convergence*					
	9	(1+2, 3)	HKY+I+G	estimated	strict	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	*-26576.153*	44.0	85.9	yes	
Effect of age	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	-26548.450	**68.0**	**103.9**	no	
priors	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML), fixed	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	-26547.967	**68.7**	**107.5**	no	
(HKY+G	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	uniform: 30 - 1000	lognormal: 34.1 + 0.5, off: 20	*bad*	-26548.839	71.1	109.1	no	
model)	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20		good	-26548.386	**61.8**	**96.7**	no	
	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML), fixed	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20		good	-26548.299	**65.3**	**104.8**	no	
	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	uniform: 30 - 1000		good	-26549.133	**67.1**	**105.2**	no	**Fig. S2**
	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML), fixed	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	uniform: 30 - 1000		good	-26547.686	**68.3**	**110.0**	no	**Fig. 1 + S3**
	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML), fixed	normal: 130 + 10	lognormal: 34.1 + 0.5, off: 20	uniform: 30 - 1000		good	-26530.420	62.8	103.5	no	
	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML), fixed	normal: 120 + 10	lognormal: 34.1 + 0.5, off: 20	uniform: 30 - 1000		good	-26530.465	56.7	93.9	no	
	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML)		lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	-26548.936	**61.1**	**92.3**	no	
	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML)		lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20		*bad*	-26548.688	41.8	64.1	no	
	9	(1+2, 3)	HKY+G	estimated	relaxed uncorrlogn	user-specified (ML)		lognormal: 34.1 + 0.5, off: 20	uniform: 30 - 1000		*bad*	-26549.694	56.3	86.4	no	
Effect of age	9	(1+2, 3)	HKY+I+G	estimated	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	-26425.094	**64.7**	**102.0**	no	
priors	9	(1+2, 3)	HKY+I+G	estimated	relaxed uncorrlogn	user-specified (ML), fixed	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	-26424.060	**66.1**	**106.0**	no	
(HKY+I+G	9	(1+2, 3)	HKY+I+G	estimated	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	uniform: 30 - 1000	lognormal: 34.1 + 0.5, off: 20	*bad*	-26425.009	68.9	109.4	no	
model)	9	(1+2, 3)	HKY+I+G	estimated	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20		good	-26425.090	**60.9**	**97.1**	no	**Fig. S4**
	9	(1+2, 3)	HKY+I+G	estimated	relaxed uncorrlogn	user-specified (ML), fixed	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20		good	-26424.516	**61.9**	**101.9**	no	
	9	(1+2, 3)	HKY+I+G	estimated	relaxed uncorrlogn	user-specified (ML)	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	uniform: 30 - 1000		*bad*	-26425.221	64.5	104.6	no	
	9	(1+2, 3)	HKY+I+G	estimated	relaxed uncorrlogn	user-specified (ML), fixed	normal: 140 + 10	lognormal: 34.1 + 0.5, off: 20	uniform: 30 - 1000		good	-26424.123	**65.2**	**110.5**	no	
	9	(1+2, 3)	HKY+I+G	estimated	relaxed uncorrlogn	user-specified (ML)		lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20	lognormal: 34.1 + 0.5, off: 20	good	-26425.166	**60.1**	**92.8**	no	
	9	(1+2, 3)	HKY+I+G	estimated	relaxed uncorrlogn	user-specified (ML)		lognormal: 34.1 + 0.5, off: 20	lognormal: 30 + 0.75, off: 20		good	-26425.043	**39.4**	**62.8**	no	
	9	(1+2, 3)	HKY+I+G	estimated	relaxed uncorrlogn	user-specified (ML)		lognormal: 34.1 + 0.5, off: 20	uniform: 30 - 1000		ok	-26425.557	**56.0**	**86.3**	no	

Table S5: Infection prevalence of CaSP across 34 different species of beewolves, as revealed by diagnostic PCRs for CaSP. Diagnostic PCRs for *Amycolatopsis* and general PCRs for Actinobacteria and Eubacteria were used to detect other bacterial symbionts in beewolf antennae.

		Antennal symbionts					CaSP infection rate	
Species	**Number of specimens (total)**	**CaSP**	**CaSP+others (co-infection)**	**other Actinobacteria**	**other bacteria**	**no bacteria**	**of all (%)**	**of colonized antennae (%)[1]**
Philanthus albopilosus	3	2		1			67	67
Philanthus barbiger	28	27				1	96	100
Philanthus cf. *basalis*	2	0		2			0	0
Philanthus basilaris	25	25					100	100
Philanthus bicinctus	3	3					100	100
Philanthus capensis	1	1					100	100
Philanthus coarctatus	1	1					100	100
Philanthus coronatus	1	1					100	100
Philanthus crabroniformis	2	2					100	100
Philanthus fuscipennis	5	5					100	100
Philanthus gibbosus	2	2					100	100
Philanthus gloriosus	6	6					100	100
Philanthus histrio	1	1					100	100
Philanthus inversus	1	1					100	100
Philanthus lepidus	1	1					100	100
Philanthus loefflingi	6	4	1	1		1	83	100
Philanthus melanderi	3	2				1	67	100
Philanthus multimaculatus	19	17				2	89	100
Philanthus pacificus	3	3					100	100
Philanthus parkeri	36	36					100	100
Philanthus psyche	15	15					100	100
Philanthus pulchellus	2	2					100	100
Philanthus pulcher	4	4					100	100
Philanthus rugosus	4	4					100	100
Philanthus triangulum	68	53	2	6	2	7	81	93
Philanthus triangulum diadema	7	6			1		86	100
Philanthus turneri	1	1					100	100
Philanthus ventilabris	6	6					100	100
Philanthus venustus	2	2					100	100
Philanthus zebratus	2	2					100	100
Trachypus boharti	68	66		1	1		97	99
Trachypus elongatus	5	5					100	100
Trachypus patagonensis	2	2					100	100
Philanthinus quattuordecimpunctatus	3	3					100	100
Total	**338**	**311**	**3**	**11**	**4**	**12**	**93**	98

[1]excluding all antennae without Actinobacteria ("no bacteria" and "other bacteria"), as the latter probably represented systemic infections with *Wolbachia, Spiroplasma,* or *Serratia*

Table S6: Collection localities and GenBank accession numbers for beewolf specimens used to reconstruct the symbiont phylogeny.

					GenBank accesion numbers			
Genus	**Species**	**Specimen no.**	**Sex**	**Collection locality**	**16S**	**EF-G/-Tu**	**gyrB**	**gyrA**
Philanthus	***albopilosus***	UT-E116	female	Utah, USA	KC607720	KC607680	KC607639	KC607532
	barbatus	USA-BS-39	female	USA				KC607533
	barbiger	UT-E8	female	Utah, USA	DQ375779	KC607681	KC607640	KC607538
		UT-E290	female	Utah, USA				KC607535
		UT-E295	female	Utah, USA				KC607536
		UT-E296	female	Utah, USA				KC607537
		USA-E55	female	Utah, USA				KC607534
	basilaris	UT-E3	female	Utah, USA	DQ375780	KC607682	KC607641	KC607540
		UT-E4	female	Utah, USA	KC607721	KC607683	KC607642	KC607545
		UT-E1	female	Utah, USA				KC607539
		UT-E349	female	Utah, USA				KC607543
		UT-E333	female	Utah, USA				KC607541
		UT-E334	female	Utah, USA				KC607542
		UT-E350	female	Utah, USA				KC607544
	bicinctus	USA-E32	female	Utah, USA	DQ375781	KC607684	KC607643	KC607546
	bilunatus	USA-BS-33	female	USA	KC607722	KC607685	KC607644	KC607547
	capensis	SA-E56	female	WCP, South Africa	DQ375782	KC607686	KC607645	KC607548
	coarctatus	coarct2	female	Oman	DQ375783	KC607687	KC607646	KC607549
	coronatus	coronat	female	Germany	DQ375784	KC607688	KC607647	KC607550
	crabroniformis	USA-E20	female	Wyoming, USA	DQ375785	KC607689	KC607648	KC607551
	crotoniphilus	USA-BS-40	female	USA	DQ375786			KC607552
	fuscipennis	SA-E19	female	ECP, South Africa		KC607690	KC607649	KC607553
		SA-E45	female	ECP, South Africa				KC607554
		SA-E37	female	ECP, South Africa	DQ375787			
	gibbosus	gib1	female	Utah, USA		KC607691		
		gib4	female	Utah, USA	DQ375788		KC607650	KC607555
		UT-E196	female	Utah, USA				KC607556
		UT-E284	female	Utah, USA				KC607557
		WI-003	female	Wisconsin, USA				KC607558
		WI-004	female	Wisconsin, USA				KC607559
	gloriosus	USA-E59a	female	Utah, USA		KC607692	KC607651	KC607560
		USA-E59c	female	Utah, USA	DQ375789			
		UT-E71	female	Utah, USA				KC607561
		UT-E72	female	Utah, USA				KC607562
	histrio	SA-E57	female	WCP, South Africa	DQ375790	KC607693	KC607652	KC607563
	inversus	UT-E50	female	Utah, USA	DQ375791	KC607694	KC607653	KC607564
		UT-E90	female	Utah, USA				KC607565
	lepidus	lep1	female	USA	DQ375792	KC607695		
		lep3	female	USA			KC607654	KC607566
	loefflingi	SA-E40	female	ECP, South Africa		KC607696	KC607655	KC607567
		SA-E41	female	ECP, South Africa				KC607568
		SA-E52	female	WCP, South Africa	DQ375793			
	melanderi	SA-E87	female	WCP, South Africa	KC607723	KC607697	KC607656	KC607569
	multimaculatus	USA-E1a	female	Utah, USA		KC607698	KC607657	
		USA-E1d	female	Utah, USA	DQ375794			
		USA-E1c	female	Utah, USA				KC607570
		UT-E25	female	Utah, USA				KC607572
		UT-E102	female	Utah, USA				KC607571
		UT-E254	female	Utah, USA				KC607573
	pacificus	UT-E221	female	Wyoming, USA	DQ375795	KC607699	KC607658	KC607574
	parkeri	USA-E43-1	female	Utah, USA	DQ375796	KC607700	KC607659	KC607575
		UT-E23	female	Utah, USA				KC607576
		UT-E24	female	Utah, USA				KC607577
		UT-E300	female	Utah, USA				KC607578
	politus	USA-BS-29	female	USA	DQ375797	KC607701	KC607660	KC607579
	psyche	USA-E44-1	female	Utah, USA	DQ375798	KC607702	KC607661	
		USA-E44-2	female	Utah, USA				KC607580
	pulchellus	SP-002	female	Spain	KC607724	KC607703	KC607662	KC607581
	pulcher	USA-E6	female	Wyoming, USA	DQ375799	KC607704	KC607663	KC607582
	rugosus	SA-E29	female	ECP, South Africa	DQ375800	KC607705	KC607664	KC607583
	sanbornii	USA-BS-37	female	USA	KC607725	KC607706	KC607665	KC607584
	solivagus	USA-BS-35	female	USA	KC607726	KC607707	KC607666	KC607585

Table S6 continued: Collection localities and GenBank accession numbers for beewolf specimens used to reconstruct the symbiont phylogeny.

					GenBank accesion numbers			
Genus	**Species**	**Specimen no.**	**Sex**	**Collection locality**	**16S**	**EF-G/-Tu**	**gyrB**	**gyrA**
Philanthus	***tarsatus***	USA-BS-25	female	USA	DQ375801	KC607708	KC607667	KC607586
	triangulum diadema	SA-E1	female	KZN, South Africa		KC607710	KC607669	KC607587
		SA-E20	female	ECP, South Africa	DQ375803			KC607591
		SA-E46	female	ECP, South Africa				KC607592
		SA-E65	female	WCP, South Africa				KC607593
		SA-E89	female	WCP, South Africa				KC607594
		SA-E115	female	WCP, South Africa				KC607590
	triangulum	S1_Ant	female	Germany		KC607709	KC607668	
		Ant7	female	Germany	DQ375802			
		S4_Ant	female	Germany				KC607599
		U3_Ant	female	Ukraine				KC607602
		D10_Ant	female	Germany				KC607588
		D11_Ant	female	Germany				KC607589
		F80_Ant	female	Germany				KC607595
		F85_Ant	female	Germany				KC607596
		F88_Ant	female	Germany				KC607597
		F90_Ant	female	Germany				KC607598
		R1_Ant	female	Germany	KC607733			
		R2_Ant	female	Germany	KC607734			
		TU-M019	female	Turkey				KC607600
		TU-M065	female	Turkey				KC607601
	turneri	SA-E116	female	WCP, South Africa	KC607727	KC607711	KC607670	KC607603
	ventilabris	UT-E70	female	Utah, USA	DQ375803	KC607712	KC607671	KC607605
		UT-E91	female	Utah, USA				KC607606
		UT-E92	female	Utah, USA				KC607607
		UT-E164	female	Utah, USA	KC607728	KC607713	KC607672	KC607604
	venustus	ven1	female	Greece	DQ375804			KC607608
		ven2	female	Greece		KC607714	KC607673	
	zebratus	USA-BS-30	female	USA	DQ375805	KC607715	KC607674	KC607609
	cf. basalis	IN-E038	female	Karnataka, India	KC607738			
		IN-E043	female	Karnataka, India	KC607739			
Trachypus	***boharti***	BR-003	female	Brasil	GU721170	KC607716	KC607675	KC607610
		BR-M001	female	Brasil				KC607611
		BR-M002	female	Brasil				KC607612
		BR-M004	female	Brasil				KC607613
		BR-M011	female	Brasil				KC607614
		BR-M019	female	Brasil				KC607615
		BR-M130	female	Brasil				KC607616
		BR-M132	female	Brasil				KC607617
		BR-M133	female	Brasil				KC607618
		BR-M135	female	Brasil				KC607619
		BR-M136	female	Brasil				KC607620
		BR-M139	female	Brasil				KC607621
		BR-M140	female	Brasil				KC607622
		BR-M141	female	Brasil				KC607623
		BR-M142	female	Brasil				KC607624
		BR-M143	female	Brasil				KC607625
		BR-M144	female	Brasil				KC607626
		BR-M145	female	Brasil				KC607627
		BR-M149	female	Brasil				KC607628
		BR-M151	female	Brasil				KC607629
	denticollis	W-Ant1	female	Chile	GU721171	KC607717	KC607676	KC607630
	elongatus	BR-M083	female	Brasil	KC607729	KC607718	KC607677	KC607632
		BR-M091	female	Brasil				KC607631
		BR-M167	female	Brasil				KC607633
		BR-M168	female	Brasil				KC607634
		BR-M170	female	Brasil				KC607635
	patagonensis	BR-M084	female	Brasil				KC607636
		BR-M092	female	Brasil	KC607730	KC607719	KC607678	KC607637
Philanthinus	***quattuordecimpunctatus***	TU-EY-E021	female	Turkey	JN104609	KC607679	KC607638	JN104610
		TU-EY-E022	female	Turkey	JN104609			JN104610
		TU-EY-E023	female	Turkey	JN104609			JN104610
Total	**42**	**124**			**48**	**41**	**41**	**109**

Table S7: GenBank accession numbers for actinobacterial sequences included in the phylogenetic analyses.

Genus	Species	Strain	16S	gyrB	gyrA	EF-G/-Tu
Frankia	*alni*	ACN14a	NC_008278	NC_008278	NC_008278	NC_008278
Streptomyces	*abikoensis (=luteoverticillatus)*	DSM 40831	KC954556	KC954562	KC954559	KC954568/KC954565
Streptomyces	*albus*	J1074	AJ621602[(1)]	NZ_DS999645	NZ_DS999645	NZ_DS999645
Streptomyces	*auratus*	AGR0001	-	NZ_JH725387	NZ_JH725387	NZ_JH725387
Streptomyces	*avermitilis*	MA4680	NC_003155	NC_003155	NC_003155	NC_003155
Streptomyces	*bingchenggensis*	BCW1	NC_016582	NC_016582	NC_016582	NC_016582
Streptomyces	*cattleya*	NRRL 8057	NC_016111	NC_016111	NC_016111	NC_016111
Streptomyces	*clavuligerus*	ATCC 27064	NZ_CM001015	NZ_CM001015	NZ_CM001015	NZ_CM001015
Streptomyces	*coelicolor*	A3(2)	NC_003888	NC_003888	NC_003888	NC_003888
Streptomyces	*flavogriseus*	ATCC 33331	NC_016114	NC_016114	NC_016114	NC_016114
Streptomyces	*ghanaensis*	ATCC 14672	AJ781384[(1)]	NZ_DS999641	NZ_DS999641	NZ_DS999641
Streptomyces	*griseoflavus*	Tu4000	AJ781322[(1)]	NZ_GG657758	NZ_GG657758	NZ_GG657758
Streptomyces	*griseus subsp. griseus*	NBRC 13350	NC_010572	NC_010572	NC_010572	NC_010572
Streptomyces	*griseus*	XylebKG1	NZ_GL877172	NZ_GL877172	NZ_GL877172	NZ_GL877172
Streptomyces	*hygroscopicus subsp. jinggangensis*	5008	NC_017765	NC_017765	NC_017765	NC_017765
Streptomyces	*hygroscopicus*	ATCC 53653	EF408736	NZ_ACEX01000401	ACEX01000401	NZ_ACEX01000355
Streptomyces	*lividans*	TK24	AY039029	NZ_GG657756	NZ_GG657756	NZ_GG657756
Streptomyces	*mutabilis*	DSM 40169	KC954557	KC954563	KC954560	KC954569/KC954566
Streptomyces	*pristinaespiralis*	ATCC 25486	-	NZ_CM000950	NZ_CM000950	NZ_CM000950
Streptomyces	*ramulosus*	DSM 40100	KC954558	KC954564	KC954561	KC954570/KC954567
Streptomyces	*roseosporus*	NRRL 11379	NZ_ABYX01000136	NZ_ABYX01000145	ABYX01000145	NZ_ABYX01000157
Streptomyces	*scabiei*	87.22	NC_013929	NC_013929	NC_013929	NC_013929
Streptomyces	*sp.*	SPB78	-	NZ_GG657742	NZ_GG657742	NZ_GG657742
Streptomyces	*sp.*	SPB74	-	NZ_GG770539	NZ_GG770539	NZ_GG770539
Streptomyces	*sp.*	SirexAAE	NC_015953	NC_015953	NC_015953	NC_015953
Streptomyces	*sp.*	Tu6071	NZ_CM001165	NZ_CM001165	NZ_CM001165	NZ_CM001165
Streptomyces	*sp.*	C	-	NZ_ACEW01000329	ACEW01000329	NZ_ACEW01000364
Streptomyces	*sp.*	Mg1	-	NZ_ABJF01000426	ABJF01000426	NZ_ABJF01000117
Streptomyces	*sviceus*	ATCC 29083	AB184559[(1)]	NZ_CM000951	NZ_CM000951	NZ_CM000951
Streptomyces	*venezuelae*	ATCC 10712	NC_018750	NC_018750	NC_018750	NC_018750
Streptomyces	*violaceusniger*	Tu 4113	NC_015957	NC_015957	NC_015957	NC_015957
Streptomyces	*viridochromogenes*	DSM 40736	AB045858[(1)]	NZ_ACEZ01000135	ACEZ01000135	NZ_ACEZ01000155
Streptosporangium	*roseum*	DSM 43021	NC_013595	NC_013595	NC_013595	NC_013595

[(1)] sequence from another strain of the same species used, because 16S rRNA sequence for the same strain was not available

Table S8: Antennal symbionts of field-collected beewolf females (*Philanthus triangulum*) applying (AGS+) or not applying (AGS-) visible amounts of antennal gland secretion to their brood cells under laboratory conditions.

Specimen number	Age (days)	AGS visible	Number of brood cells	Brood cells with visible AGS number	Brood cells with visible AGS proportion (%)	Antennal symbionts (diagnostic PCRs and sequencing)	BLAST identity
10b	51	+	39	29	74.4	CaSP	
12a	56	+	48	42	87.5	CaSP	
15d	57	+	36	30	83.3	CaSP	
24c	24	+	9	7	77.8	CaSP	
25c	46	+	12	12	100	CaSP	
29c	36	+	32	26	81.3	CaSP	
04c	53	+	18	11	61.1	*Rhodococcus baikonurensis,*	100%
						Nocardioides simplex	100%
08a	18	-	2	0	0	none detected	
29b	60	-	35	0	0	*Streptomyces pluricolorescens*	99%
04b	63	-	17	0	0	*Streptomyces flavofuscus*	100%
19a	31	-	5	0	0	*Streptomyces ramulosus*	99%
10c	44	-	23	0	0	*Streptomyces rochei*	99%
15c	34	-	20	0	0	Streptomyces phaeochromogenes	99%

Table S9: Symbiont establishment and transmission success upon artificial infection with native (CaSP) and non-native (Amy) symbionts.

Experimental infection treatment	Individual	Life span (days)	Antennal symbionts PCR CaSP	Antennal symbionts PCR Amy	Antennal symbionts FISH CaSP	Antennal symbionts FISH Amy	Number of brood cells	Brood cells with visible AGS (%)	Number of cocoons tested	GC-MS antibiotics-positive (%)	Diagnostic PCRs CaSP pos (%)	Diagnostic PCRs Amy pos (%)
CaSP	**HT-W07**	21	+	-	+	-	7	57.1	6	50.0	33.3	16.7
	HT-W10	38	+	-	+	-	20	80	20	70.0	55.0	0.0
	HT-W15	38	+	-	N/A	N/A	39	58.9	34	88.2	61.8	35.3
	HT-W21	31	+	-	+	-	8	62.5	8	75.0	50.0	0.0
	HT-W32	42	+	-	-	(+)	17	64.7	14	85.7	57.1	42.9
Amy	**HT-W01**	41	-	+	-	+	6	0	6	0.0	50.0	16.7
	HT-W11	24	-	+	-	(+)	19	0	18	0.0	27.8	16.7
	HT-W14	45	+	+	+	-	34	0	25	20.0	4.0	8.0
	HT-W18	13	-	-	-	+	15	0	14	0.0	7.1	7.1
	HT-W23	11	-	+	-	(+)	20	0	14	0.0	14.3	0.0
	HT-W29	22	-	+	-	-	8	0	7	0.0	0.0	14.3

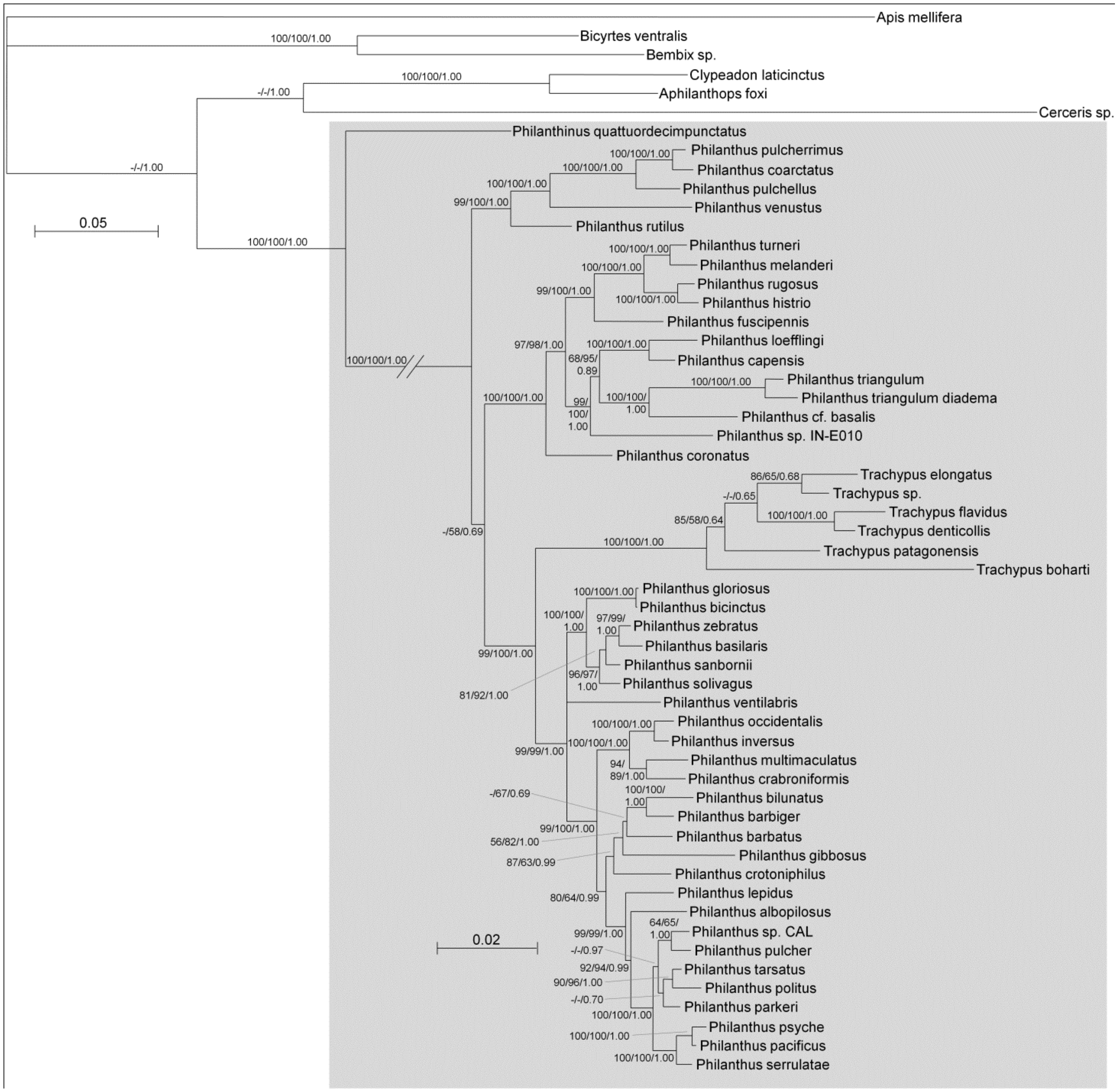

Figure S1: Reconstruction of phylogenetic relationships among Philanthini digger wasps. The phylogeny is based on a concatenated alignment of 5521 bp of *28S, lwrh*, *argK*, *wnt*, *ef1a*, and *coxI*. Bootstrap values (>50%) from maximum-parsimony (MP, 1,000 replicates) and maximum likelihood (ML, 10,000 replicates) analyses as well as Bayesian posterior probabilities (>0.5) are provided at the nodes. Taxa with antennal *Streptomyces* symbionts are highlighted with grey background. Scale bars represent substitutions per site.

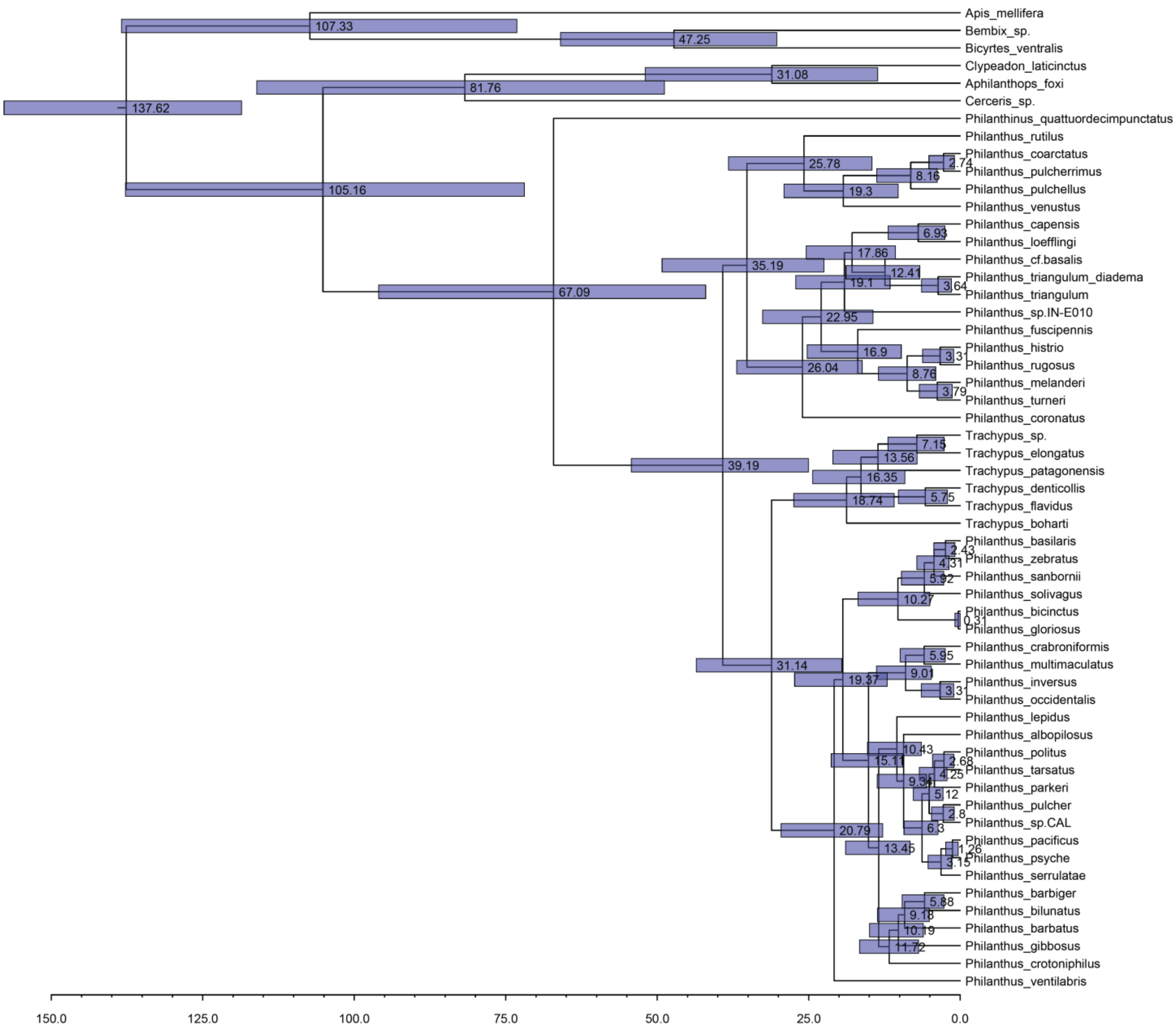

Figure S2: Dated phylogeny of the Philanthinae. Phylogenetic tree with the highest clade credibility resulting from BEAST analyses under the uncorrelated lognormal clock model, based on the combined, partitioned 6-gene-data set (codon partitioning [1+2, 3] for the protein-coding genes), using the HKY+G substitution model and the ML tree from the host phylogenetic analyses as the stating tree. Node ages are shown in million years ago (Mya) with their 95% HPD interval bars (equivalent to 95% confidence intervals). The fossils of *Cerceris berlandi* (~30 Mya, used to calibrate the age of the Cercerini+Aphilanthopini with a uniform distribution with minimum 30 Mya) and *Psammaecius sepultus* (~34.1 Mya, used to calibrate the age of the Bembecini by a lognormal distribution with mean±SD=34.1±0.5, offset=20.0) as well the age of the root (modelled with a normal distribution with mean±SD=140.0±10.0 based on earlier phylogenetic analyses) were used for age calibration. The scale represents divergence time in Mya.

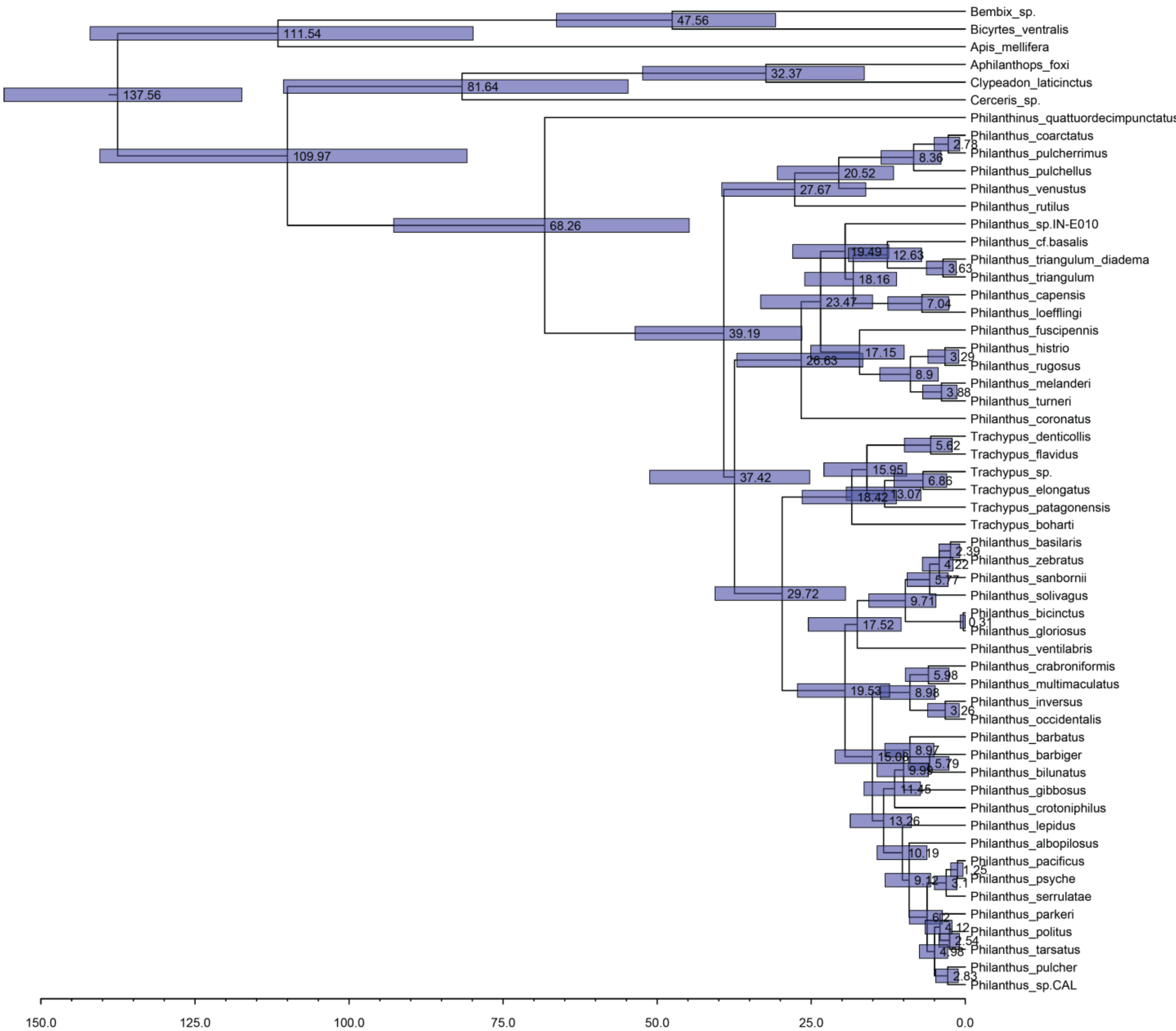

Figure S3: Dated phylogeny of the Philanthinae. Phylogenetic tree with the highest clade credibility resulting from BEAST analyses under the uncorrelated lognormal clock model, based on the combined, partitioned 6-gene-data set (codon partitioning [1+2, 3] for the protein-coding genes), using the HKY+G substitution model. The tree topology was fixed to the ML tree from the host phylogenetic analyses. Node ages are shown in million years ago (Mya) with their 95% HPD interval bars (equivalent to 95% confidence intervals). The same calibration points as in Fig. S2 were used. The scale represents divergence time in Mya.

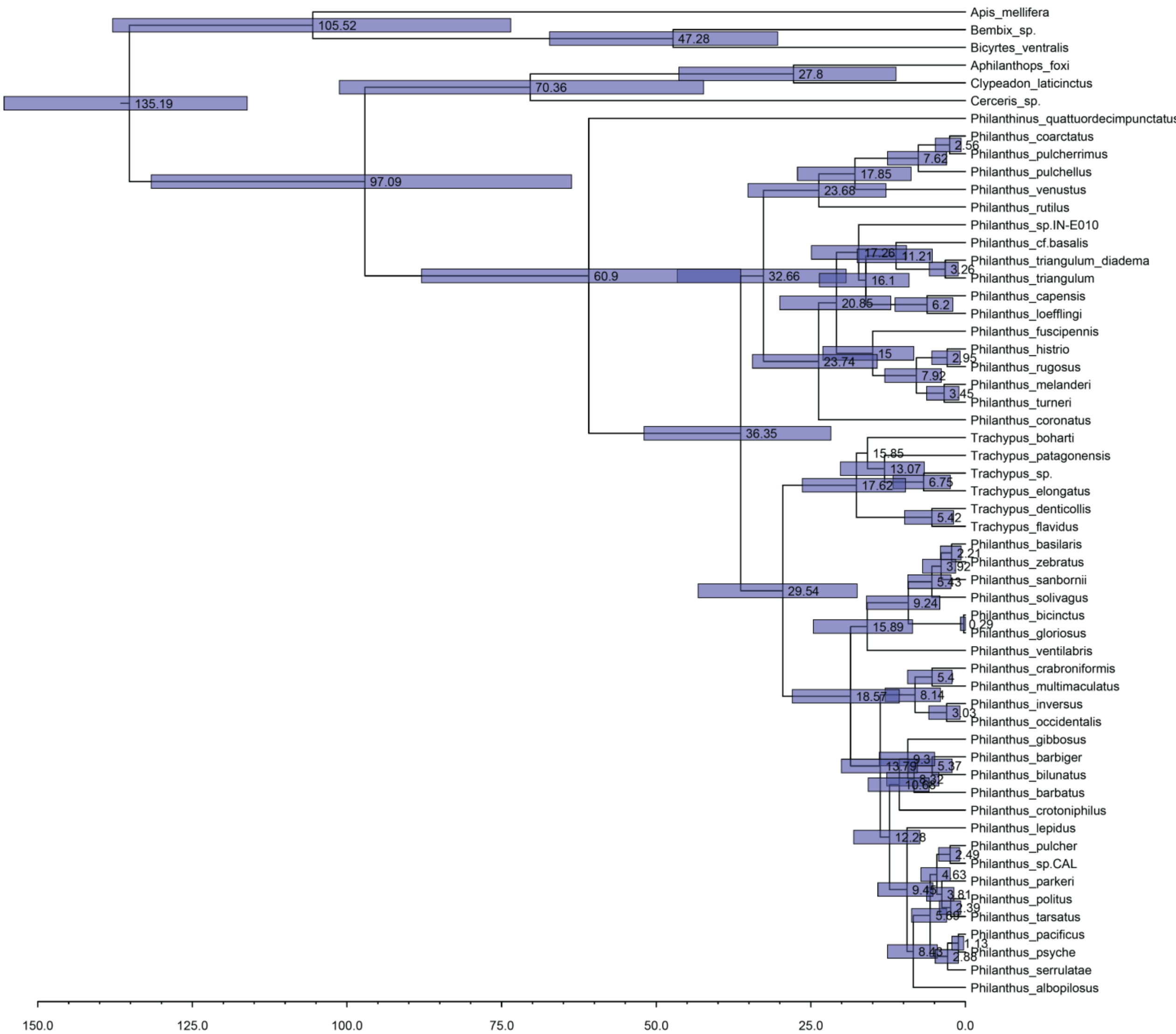

Figure S4: Dated phylogeny of the Philanthinae. Phylogenetic tree with the highest clade credibility resulting from BEAST analyses under the uncorrelated lognormal clock model, based on the combined, partitioned 6-gene-data set (codon partitioning [1+2, 3] for the protein-coding genes), using the HKY+G+I substitution model and the ML tree from the host phylogenetic analyses as the stating tree. Node ages are shown in million years ago (Mya) with their 95% HPD interval bars (equivalent to 95% confidence intervals). The same calibration points as in Fig. S2 were used. The scale represents divergence time in Mya.

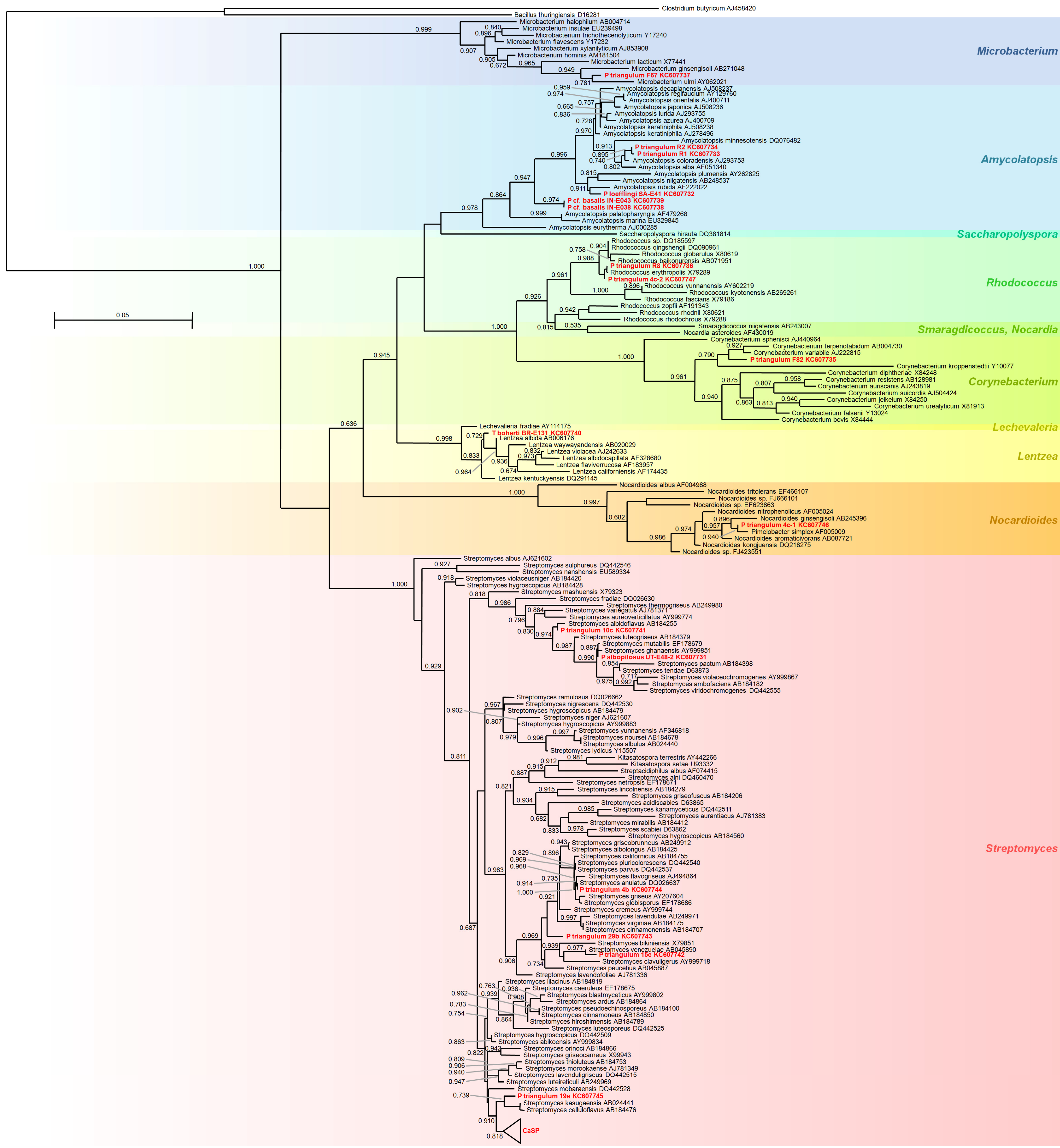

Figure S5: Phylogenetic placement of Actinobacteria detected in beewolf antennae based on 16S rRNA sequence data. While the majority of the infected antennae of all individuals harbored bacteria within the CaSP symbiont clade (98%), other strains within *Streptomyces, Amycolatopsis, Microbacterium, Nocardioides, Corynebacterium, Lentzea*, and *Rhodococcus* were occasionally recorded. The approximately-maximum-likelihood tree was reconstructed with FastTree based on a secondary-structure guided alignment of 16S rRNA sequences with the SINA aligner. Sequences obtained from beewolf antennae are highlighted in bold red font. Scale bar represents substitutions per site.

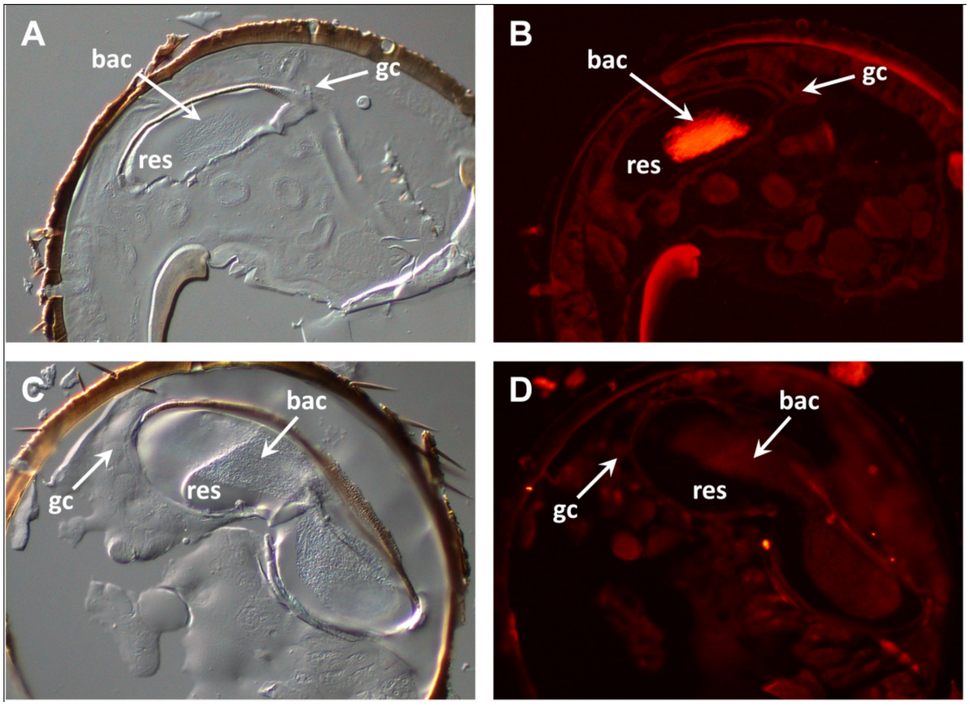

Figure S6: Replacement of CaSP symbionts by *Amycolatopsis* in antennae of a female *Philanthus* cf. *basalis.* (**A**) and (**C**) Differential interference contrast micrographs of antennal cross-sections. (**B**) and (**D**) Fluorescent micrographs of the same areas, after staining with the *Amycolatopsis*-specific probe Amy_16S-Cy3 (**B**) or the CaSP-specific probe SPT177-Cy3 (**D**). bac=bacteria, res=antennal gland reservoir, gc=gland cells.

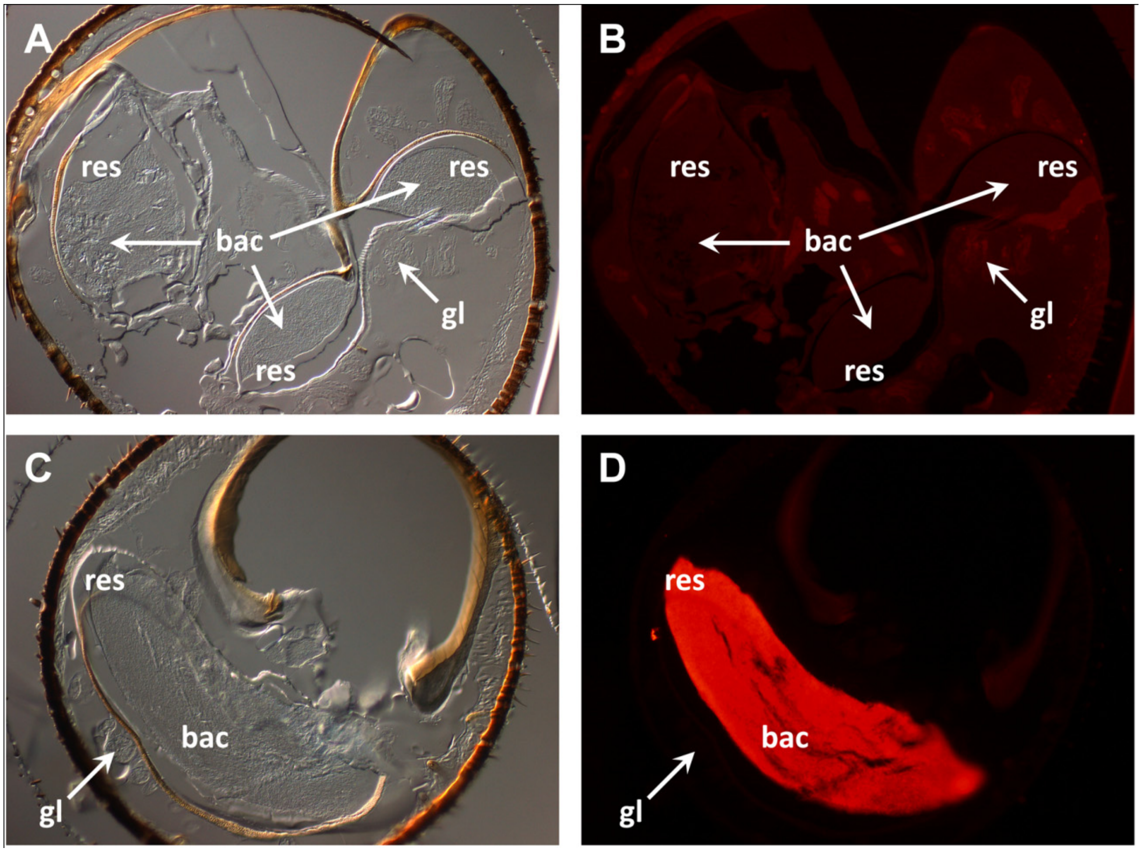

Figure S7: FISH of CaSP symbionts in the antennae of a female *Philanthus triangulum,* demonstrating specificity of the probe Amy_16S-Cy3 for *Amycolatopsis*. (A) and (C) Differential interference contrast micrographs of antennal cross-sections. (B) and (D) Fluorescent micrographs of the same areas, after staining with the *Amycolatopsis*-specific probe Amy_16S-Cy3 (B) or the CaSP-specific probe SPT177-Cy3 (D). bac= bacteria, res = antennal gland reservoir, gl = gland cells.

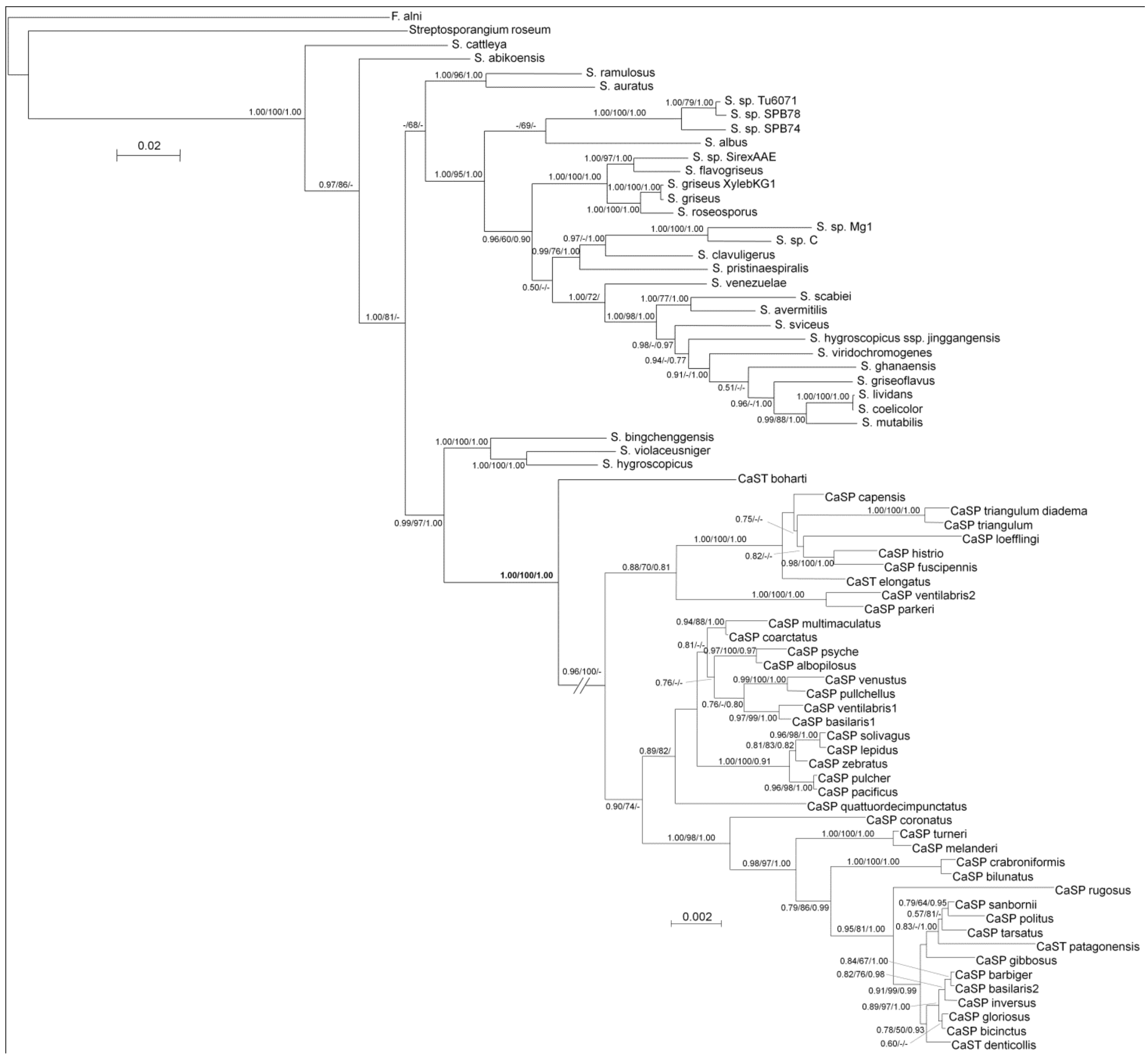

Figure S8: Phylogenetic relationships among beewolf symbionts. The phylogeny was reconstructed using Bayesian and maximum likelihood models, based on the concatenated alignment of 16S rRNA, gyrA, gyrB, and EF-Tu/G (4653 aligned bp). Values at the nodes are local support values from the FastTree analysis (GTR model), bootstrap values for the PHYML analysis (Geneious), and Bayesian posteriors, respectively. Scale bars represent substitutions per site.

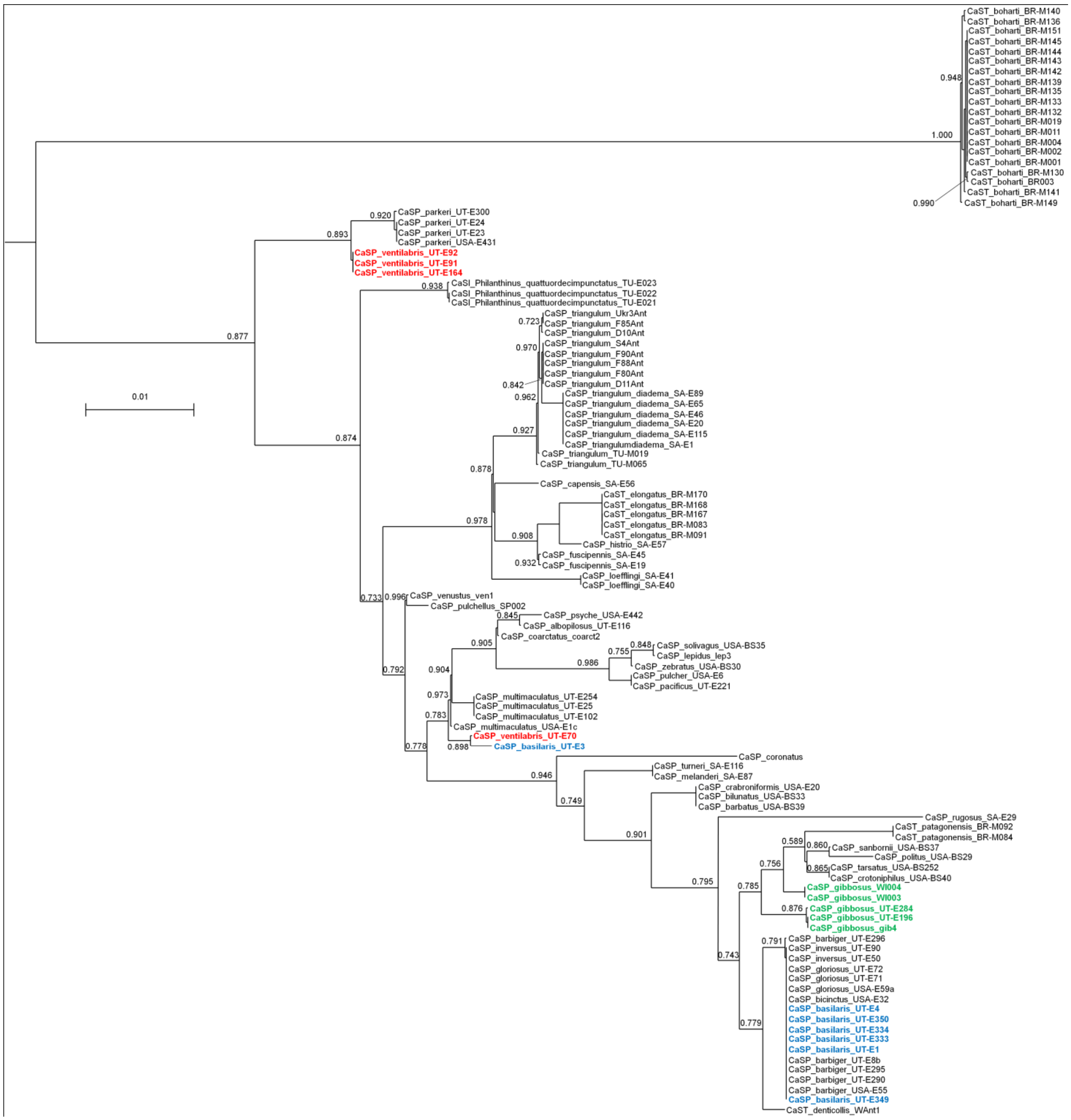

Figure S9: Phylogenetic relationships among symbionts of 109 beewolf individuals across 41 species. The phylogeny was reconstructed using FastTree (GTR model), based on partial gyrase A (gyrA) sequences. Numbers at the nodes represent local support values. Host species with individuals carrying symbionts in different clades are highlighted in different colors. Scale bar represents substitutions per site.

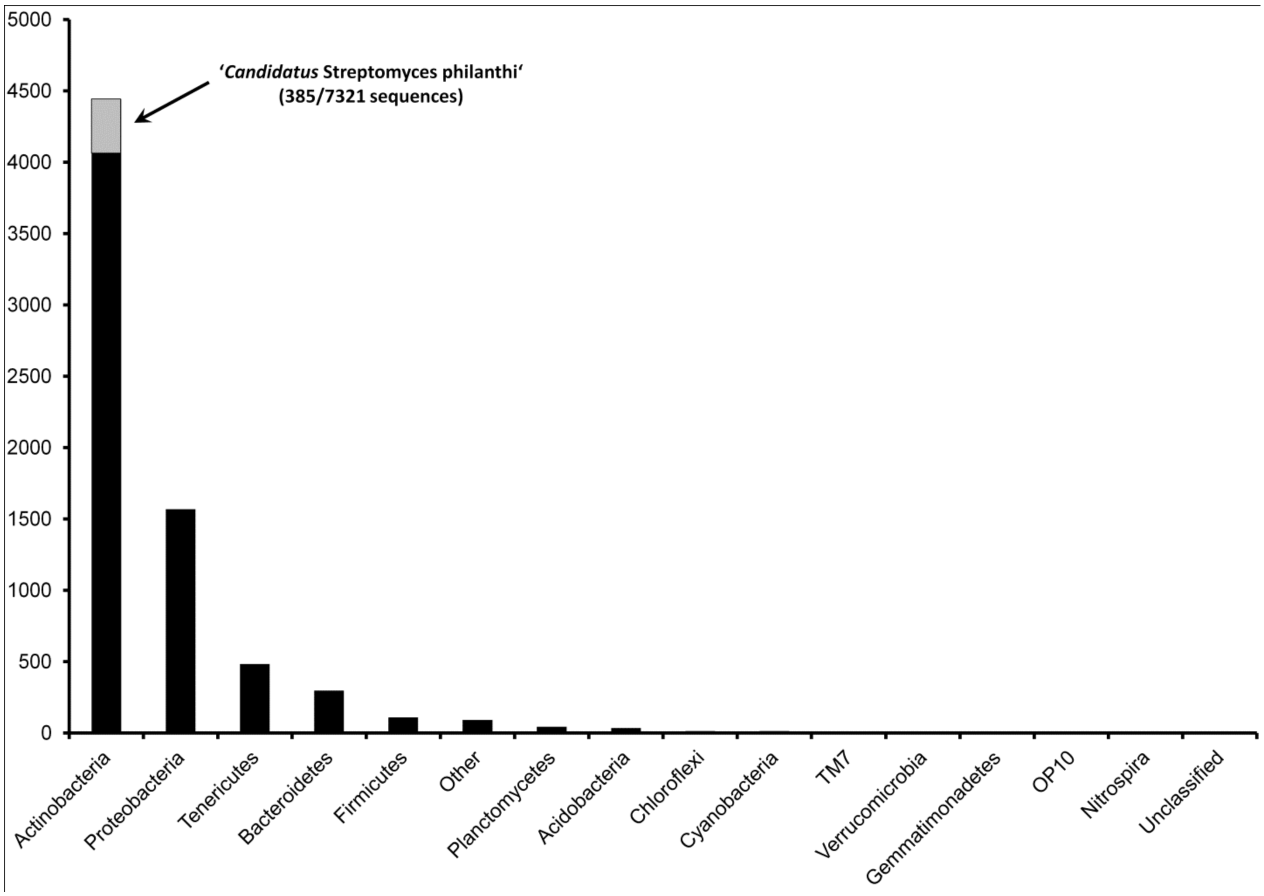

Figure S10: Detection of CaSP in sand from beewolf rearing cages. The microbial community composition was determined by bacterial tag-encoded FLX amplicon sequencing (bTEFAP) of bacterial 16S rRNA. After quality control, denoising and OTU picking (cdhit and uclust), OTUs were combined based on phylum-level taxonomic affiliation. The proportion of '*Candidatus* Streptomyces philanthi triangulum' 16S reads is highlighted in grey.

Supplementary references

1. Cameron SA & Mardulyn P (2001) Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera : Apinae). *Syst Biol* 50(2):194-214.
2. Danforth BN, Brady SG, Sipes SD, & Pearson A (2004) Single-copy nuclear genes recover Cretaceous-age divergences in bees. *Syst Biol* 53(2):309-326.
3. Kawakita A*, et al.* (2003) Evolution and phylogenetic utility of alignment gaps within intron sequences of three nuclear genes in bumble bees (*Bombus*). *Mol Biol Evol* 20(1):87-92.
4. Ramirez SR, Roubik DW, Skov C, & Pierce NE (2010) Phylogeny, diversification patterns and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae). *Biol J Linnean Soc* 100(3):552-572.
5. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95-98.
6. Gouy M, Guindon S, & Gascuel O (2010) SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27(2):221-224.
7. Stamatakis A, Ludwig T, & Meier H (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21(4):456-463.
8. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688-2690.
9. Stamatakis A, Hoover P, & Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* 57(5):758-771.
10. Tavaré S (1986) Some probabilistic and statistical problems on the analysis of DNA sequences. *Lectures in mathematics in the life sciences*, (American Mathematical Society), Vol 17, pp 57-86.
11. Yang Z (1993) Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol* 10:1396-1401.
12. Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J Mol Evol* 39:306-314.
13. Swofford DL (2003) *PAUP*. Phylogenetic analysis using parsimony (and other methods), Version 4* (Sinauer Associates, Sunderland, Massachussetts).
14. Abascal F, Zardoya R, & Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104-2105.
15. Huelsenbeck JP & Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8):754-755.
16. Huelsenbeck JP, Ronquist F, Nielsen R, & Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294(5550):2310-2314.
17. Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572-1574.
18. Drummond AJ & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7.

19. Drummond AJ, Ho SYW, Phillips MJ, & Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4(5):699-710.
20. Cockerell TDA (1906) Fossil Hymenoptera from Florissant, Colorado. *Bull Mus Comp Zool Harvard Coll* 50:31-58.
21. Pulawski WJ & Rasnitsyn AP (1980) The taxonomic position of *Hoplisus sepultus* from the lower oligocene of Colorado. *Polskie Pismo Entomologiczne* 50(3):393-396.
22. Evanoff E, McIntosh WC, & Murphey PC (2001) Stratigraphic summary and 40AR/39AR geochronology of the Florissant Formation, Colorado. *Proc Denver Mus Nat Sci* 4(1):1-16.
23. Timon-David J (1944) Insectes fossiles de l'Oligocene inferieur des Camoins (Bassin de Marseille). I.- Dipteres brachyceres. II. Hymenopteres [Fossil insects of the lower Oligocene in Camoins (Bassin de Marseille). I. Diptera, Brachycera. II. Hymenoptera]. *Bull Soc Entomol Fr* 48:40-45. French
24. Rohwer SA (1909) Three new fossil insects from Florissant, Colorado. *Am J Sci* 28(168):533-536.
25. Theobald N (1937) Les insectes fossiles des terrains oligocenes de France [The fossil insects of oligocene soils in France]. *Mem Soc Sci Nancy* 473:1-473. French
26. Cardinal S & Danforth BN (2013) Bees diversified in the age of eudicots. *P Roy Soc B-Biol Sci* 280(1755).
27. Grimaldi D & Engel MS (2005) *Evolution of the insects* (Cambridge University Press, New York) pp 1-755.
28. Debevec AH, Cardinal S, & Danforth BN (2012) Identifying the sister group to the bees: a molecular phylogeny of Aculeata with an emphasis on the superfamily Apoidea. *Zool Scr* 41(5):527-535.
29. Melo GAR (1999) Phylogenetic relationships and classification of the major lineages of Apoidea (Hymenoptera), with emphasis on the crabronid wasps. *Univ Kansas Nat Hist Mus* 14:1-55.
30. Cockerell TDA (1915) British fossil insects. *Proc US Natl Mus* 49(2119):469-499.
31. Brady SG, Larkin L, & Danforth BN (2009) Bees, ants, and stinging wasps (Aculeata). *The timetree of life.*, eds Hedges SB & Kumar S), pp 264-269.
32. Rambaut A & Drummond AJ (2007) Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer.
33. Rambaut A (2010) FigTree v1.3.1: Tree figure drawing tool. Available from http://tree.bio.ed.ac.uk/software/figtree.
34. Sambrook J, Fritsch EF, & Maniatis T (1989) *Molecular cloning: a laboratory manual* (Cold Spring Harbor Laboratory Press, New York) 2nd ed. Ed.
35. Kaltenpoth M*, et al.* (2006) *'Candidatus* Streptomyces philanthi', an endosymbiotic streptomycete in the antennae of *Philanthus* digger wasps. *Int J Syst Evol Microbiol* 56(6):1403-1411.
36. Kaltenpoth M, Schmitt T, Polidori C, Koedam D, & Strohm E (2010) Symbiotic streptomycetes in antennal glands of the South American digger wasp genus *Trachypus* (Hymenoptera, Crabronidae). *Physiol Entomol* 35(2):196-200.
37. Kaltenpoth M, Yildirim E, Gürbüz MF, Herzner G, & Strohm E (2012) Refining the roots of the beewolf-*Streptomyces* symbiosis: Antennal symbionts in the rare genus *Philanthinus* (Hymenoptera, Crabronidae). *Appl Environ Microbiol* 78(3):822-827.
38. Drummond AJ*, et al.* (2011) Geneious v5.4 (Biomatters Ltd., Auckland, New Zealand).

39. Ludwig W, *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* 32(4):1363-1371.

40. Price MN, Dehal PS, & Arkin AP (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5(3):e9490.

41. Guindon S & Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5):696-704.

42. Iwamoto T, *et al.* (2000) Monitoring impact of in situ biostimulation treatment on groundwater bacterial community by DGGE. *FEMS Microbiol Ecol* 32(2):129-141.

43. Kawai M, *et al.* (2002) 16S ribosomal DNA-based analysis of bacterial diversity in purified water used in pharmaceutical manufacturing processes by PCR and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* 68(2):699-704.

44. Stoll S, Gadau J, Gross R, & Feldhaar H (2007) Bacterial microbiota associated with ants of the genus *Tetraponera*. *Biol J Linnean Soc* 90(3):399-412.

45. Pruesse E, *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35(21):7188-7196.

46. Pruesse E, Peplies J, & Gloeckner FO (2012) SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28(14):1823-1829.

47. Amann RI, *et al.* (1990) Combination of 16S ribosomal RNA targeted oligonucleotide probes with flow-cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* 56(6):1919-1925.

48. Strohm E & Linsenmair KE (1995) Leaving the cradle - how beewolves (*Philanthus triangulum* F.) obtain the necessary spatial information for emergence. *Zoology* 98(3):137-146.

49. Berry AE, Chiocchini C, Selby T, Sosio M, & Wellington EMH (2003) Isolation of high molecular weight DNA from soil for cloning into BAC vectors. *FEMS Microbiol Lett* 223(1):15-20.

50. Ishak HD, *et al.* (2011) Bacterial diversity in *Solenopsis invicta* and *Solenopsis geminata* ant colonies characterized by 16S amplicon 454 pyrosequencing. *Microb Ecol*:1-11.

51. Sun Y, Wolcott RD, & Dowd SE (2011) Tag-encoded FLX amplicon pyrosequencing for the elucidation of microbial and functional gene diversity in any environment. *Methods Mol Biol* 733:129-141.

52. Caporaso JG, *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335-336.

53. Reeder J & Knight R (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat Methods* 7(9):668-669.

54. Li W & Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658-1659.

55. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460-2461.

56. Caporaso JG, *et al.* (2010) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26(2):266-267.

57. Wang Q, Garrity GM, Tiedje JM, & Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73(16):5261-5267.

58. Weisburg WG, Barns SM, Pelletier DA, & Lane DJ (1991) 16s ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 173(2):697-703.
59. Hain T, WardRainey N, Kroppenstedt RM, Stackebrandt E, & Rainey FA (1997) Discrimination of *Streptomyces albidoflavus* strains based on the size and number of 16S-23S ribosomal DNA intergenic spacers. *Int J Syst Bacteriol* 47(1):202-206.
60. Hatano K, Nishii T, & Kasai H (2003) Taxonomic re-evaluation of whorl-forming Streptomyces (formerly Streptoverticillium) species by using phenotypes, DNA-DNA hybridization and sequences of gyrB, and proposal of Streptomyces luteireticuli (ex Katoh and Aral 1957) corrig., sp nov., nom. rev. *Int J Syst Evol Microbiol* 53:1519-1529.
61. Kaltenpoth M, Gottler W, Herzner G, & Strohm E (2005) Symbiotic bacteria protect wasp larvae from fungal infestation. *Curr Biol* 15(5):475-479.
62. Danforth BN, Conway L, & Ji SQ (2003) Phylogeny of eusocial Lasioglossum reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera : Halictidae). *Syst Biol* 52(1):23-36.
63. Danforth BN, Sauquet H, & Packer L (1999) Phylogeny of the bee genus Halictus (Hymenoptera : Halictidae) based on parsimony and likelihood analyses of nuclear EF-1 alpha sequence data. *Mol Phylogenet Evol* 13(3):605-618.
64. Danforth BN, Sipes S, Fang J, & Brady SG (2006) The history of early bee diversification based on five genes plus morphology. *Proc Natl Acad Sci USA* 103(41):15118-15123.
65. Mardulyn P & Cameron SA (1999) The major opsin in bees (Insecta: hymenoptera): A promising nuclear gene for higher level phylogenetics. *Mol Phylogenet Evol* 12(2):168-176.
66. Folmer O, Black M, Hoeh W, Lutz R, & Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3(5):294-299.
67. Kronauer DJ, Holldobler B, & Gadau J (2004) Phylogenetics of the new world honey ants (genus *Myrmecocystus*) estimated from mitochondrial DNA sequences. *Mol Phylogenet Evol* 32(1):416-421.
68. Brower AVZ & DeSalle R (1998) Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. *Insect Mol Biol* 7(1):73-82.
69. Stach JEM, Maldonado LA, Ward AC, Goodfellow M, & Bull AT (2003) New primers for the class Actinobacteria: application to marine and terrestrial environments. *Environ Microbiol* 5(10):828-841.