

# An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea)

Dominic A. Evangelista, Benjamin Wipfler, Olivier Béthoux, Alexander Donath, Mari Fujita, Manpreet K. Kohli, Frédéric Legendre, Shanlin Liu, Ryuichiro Machida, Bernhard Misof, Ralph S. Peters, Lars Podsiadlowski, Jes Rust, Kai Schuette, Ward Tollenaar, Jessica L. Ware, Torsten Wappler, Xin Zhou, Karen Meusemann, Sabrina Simon

**This README and corresponding data refer to: Evangelista *et al.* (2019), *Proceedings of the Royal Society B*, DOI: 10.1098/rspb.2018.2076**

This README was written on January 9, 2019. If you have any questions regarding the nature of our data, please feel free to contact Sabrina Simon (sabrina.simon[AT]wur.nl)

## Supplementary files and descriptions provided via the Digital Repository Dryad

**NOTE:** In several files and figures, some species names may be different in the Supplementary Online Material due to identification updates and corrections. Valid names are provided in Table S1. It concerns the following names:

<b>Name in analysed files (outdated)</b>	<b>Correct and valid name</b>	<b>Explanation</b>
<i>Ischnoptera</i> sp.	<i>Ischnoptera deropeltiformis</i>	Correct species name updated
<i>Eurycotis floridiana</i> , <i>Eurycotis decipiens</i>	<i>Eurycotis floridana</i>	Typo corrected and wrong species name corrected which was partially used by accident
<i>Cryptocercus</i> sp.	<i>Cryptocercus wrighti</i>	Correct species name updated
<i>Lamproblatta albipalpa</i>	<i>Lamproblatta albipalpus</i>	Typo corrected
<i>Gratidia madagassa</i>	<i>Antongilia madagassa</i>	Correct genus name updated
<i>Timema christinae</i>	<i>Timema cristinae</i>	Typo corrected
<i>Blatella germanica</i>	<i>Blattella germanica</i>	Typo corrected
<i>Princisia vanwaerebecki</i>	<i>Princisia vanwaerebeki</i>	Typo corrected

## **Supplementary Archive 1.**

This archive contains all files relevant to molecular analyses (see Supplementary text S1: Datasets and molecular analyses). [Supplementary\_Archive\_1.zip: 54.17 MB, md5sum: e5cad31efa62d6015f2666ee781396c7]

The archive contains the following subdirectories:

### **S1.1\_OrthologSet**

Files can be used as input for Orthograph. The subdirectory includes a table with all orthologous sequence groups (OGs) as used in this study from the OrthoDB8 database. We provide the table with annotation and without; latter serves as input Orthograph (\*\_orthograph.txt), and the four official gene sets on amino-acid (\*.protein.fas) and nucleotide level (\*.CDS.fas) cleaned and with corresponding headers (FASTA format) for the reciprocal BLAST search (= ready to use for HaMStR-ad and Orthograph). Shortcuts: EDAN: *E. danica*, LFUL: *L. fulva*, RPRO: *R. prolixus*, ZNEV: *Z. nevadensis*, see Supplementary Table S3 and Methods section.

## S1.2\_Datasets

This subdirectory includes two supermatrices (PHYLIP format) and respective partition files (NEXUS format) including selected models (\*partitions.nex, assigned models to each partition are listed at the bottom). The partitioning based on protein domains annotation.

- decisive supermatrix (superalignment) on amino-acid level + corresponding partition file including selected models (“decisive amino-acid dataset”, see Figure 1)
- supermatrix on nucleotide level including only 2<sup>nd</sup> codon positions + corresponding partition file including selected models (“full nucleotide dataset”, see Supplementary Figure S3)

## S1.3\_AU\_Tests

This subdirectory includes supplementary files and results of all AU-test analyses in plain text format.

- `Blattodea_AUtest_command.txt`: command as used in IQTree
- `Blattodea_AUtest_trees.tre`: all trees tested in Newick format (best ML tree inferred plus 13 alternative topologies, see comments and Supplementary Table S7)
- `Blattodea_AUtest.log`; `Blattodea_AUtest.iqtree`: IQTree output files
- `Blattodea_AUtest_information_on_tested_trees.txt`: additional information corresponding to the log file and AUtest result

## S1.4\_Anallacta

This subdirectory includes the alignment (nucleotide level, FASTA format) of cytochrome oxidase I (COI) sequences of i) a cultured specimen of *Anallacta* sp. (SANGER sequences) and ii) extracted from the transcriptome used in this study, see also Supplementary text S1).

## S1.5\_Dating

This subdirectory includes files and subdirectories related to divergence time analyses (plain text files).

- `Blattodea_calibrations.tre`: tree with minimum and maximum prior age calibrations (NEWICK format as required for MCMCTree), see Figure S5 and Supplementary material S2.
- `Blattodea_mcmctree.ct1`: MCMCTree control file (example filenames) for the main analyses after generation of the Hessian matrix. Note that for all runs parameters were kept identical.
- `Blattodea_supermatrix_aa_reduced_95.fas`: reduced amino-acid dataset (FASTA format). See Supplementary texts S1. For resulting divergence dates see Figure 1.
- `Blattodea_priorOnly_noData_1run.tre`: tree inferred from the analysis utilizing priors only used to check if priors were conflicting and to map effective and specified priors in relation to each other. For further details see Supplementary text S1 and Figure S6.
- `Blattodea_FigTree_full_aa_run3.tre`: Chronogram inferred from unreduced amino-acid dataset (replicate run 3, chosen arbitrarily among 4 replicates, for details see Supplementary text S1).
- `Blattodea_FigTree_95_aa_run1.tre`: Chronogram inferred from the reduced amino-acid dataset (replicate run 3, chosen arbitrarily among 4 replicates, for details see Supplementary text S1).
- `Blattodea_FigTree_95_aa_run1_altCalibrations.tre`: Chronogram inferred from the reduced amino-acid dataset with alternative prior calibrations (for details see Supplementary text S1 and Table S9).
- Subdirectory “dated\_trees\_unreduced”: Chronograms (NEWICK format) inferred from the unreduced amino-acid dataset, 4 replicates; the chronogram from run 3 was arbitrarily chosen as the representative for dates reported in our study (also see Supplementary text S1 and Table S9).  
Files: `FigTree_full_run1.tre`; `FigTree_full_run2.tre`; `FigTree_full_run3.tre`; `FigTree_full_run4.tre`
- Subdirectory “dated\_trees\_reduced”: Chronograms (NEWICK format) inferred from the reduced amino-acid dataset, 4 replicates; the chronogram from run 1 was arbitrarily chosen as the representative for dates reported in our study (also see Supplementary text S1 and Table S9).  
Files: `FigTree_95_run1.tre`; `FigTree_95_run2.tre`; `FigTree_95_run3.tre`; `FigTree_95_run4.tre`

## **S1.6\_Blattabacterium**

This subdirectory includes a spreadsheet with the number of transcripts identified as *Blattabacterium* for each taxon (sheet 1) and the sequence name and organismal identity of each candidate transcript (sheet 2), details are provided in Supplementary text S1.

## **Supplementary Archive 2.**

This archive refers to analyses as described in Supplementary text S3: Evolution, Morphology & Behavior. [Supplementary\_Archive\_2.zip: 93.97 KB, md5sum: 84590a5ef1d875f748a8c0e3cad4483d]

## **Ancestral state reconstruction**

The file includes the input character matrix (NEXUS format) and output (Figure 3) of the ancestral state reconstructions for 19 morphological characters listed in the Supplementary text S3. Details on character states are provided in S3.3, see Table S11 for a more readable summary of the input matrix and Table S12 for the ancestral states of selected nodes only. This matrix was used as input for parsimony analysis with Mesquite.

File: Blattodea\_ancStates\_datamatrix.nex

## **Subdirectory “Stochastic character analysis”**

This subdirectory includes supplementary files, subdirectories and results of the stochastic character mapping Bayesian ancestral state inference with Phytools.

- PhytoolsCode.R: R-script used to analyze the input data.
- 95datedTree.run1.tre: divergence dated tree obtained with the reduced dataset.
- morpho\_data\_selected\_PriorProb\_version.csv : Input data matrix with prior probabilities for the two states of each character analyzed.
  
- **Character 15** - subdirectory with results of the analyses for character 15 (wing presence and absence).
  - Info.txt: Output from the Phytools (make.simmap) analysis, summary output, description of the color code used in the corresponding figure, and notes about the results.
  - Wing presence and absence.png: result of the Bayesian ancestral state reconstruction of wing presence and absence in Blattodea. Red coloration indicates posterior probability of wing presence; black coloration indicates posterior probability of wing absence.
- **Character 16** - subdirectory with results of the analyses for character 16 (wings dropped or retained).
  - Info.txt: Output from the Phytools (make.simmap) analysis, summary output, description of the color code used in the corresponding figure, and notes about the results.
  - Wing dropping.png: results of the Bayesian ancestral state reconstruction of wing dropping and retention in Blattodea. Red coloration indicates posterior probability of wing dropping; black coloration indicates posterior probability of wing retention.
- **Character 17** - subdirectory with results of the analyses for character 17 (hindwing folding simple or fanlike).
  - Info.txt: Output from the Phytools (make.simmap) analysis, summary output from a summary of that analysis, a description of the color code used in the corresponding figure, and notes about the results.
  - Hindwing folding.png: A Bayesian ancestral state reconstruction of hindwing folding in Blattodea. Red coloration indicates posterior probability of a non-fan-like fold; black coloration indicates posterior probability of fan-like fold.
- **Character 18** - subdirectory with results of the analyses for character 18 (forewing tegminization).
  - Info.txt: Output from the Phytools (make.simmap) analysis, summary output, a description of the color code used in the corresponding figure, and a notes about the results.
  - Tegminization.png: A Bayesian ancestral state reconstruction of hindwing folding in Blattodea. Red coloration indicates posterior probability of a tegminized forewings; black coloration indicates posterior probability of membranous forewings.