No evidence for immune-gene specific signals of selection in termites

1Running title: Selection on termite immune genes

2Karen Meusemann^{1,2}, Judith Korb¹, Maximilian Schughart¹, Fabian Staubach^{1*}

³Evolutionary Biology & Ecology, Biology I (Animal Zoology), University of Freiburg, Freiburg
⁴(Brsg.), Germany

⁵Australian National Insect Collection, CSIRO, Acton, Canberra, ACT, Australia

6* Correspondence:

7Fabian Staubach

8fabian.staubach@biologie.uni-freiburg.de

9

10Life Science Identifiers (as available Zoobank)

11*Ephemera danica*:

12urn:lsid:zoobank.org:act:06633F75-4809-4BB3-BDCB-6270795368D5

13*Coptotermes* sp.

14urn:lsid:zoobank.org:pub:D6724B7F-F27A-47DC-A4FC-12859ECA0C71

15*Blattella germanica*:

16rn:lsid:zoobank.org:pub:1EA126BA-E9D2-4AA6-8202-26BA5B09B8AD

17*Locusta migratoria*

18urn:lsid:zoobank.org:pub:D792A09E-844A-412A-BFCA-5293F8388F8C

19*Periplaneta americana* (*Blatta americana*):

20urn:lsid:zoobank.org:act:95113A55-4C6D-4DC7-A0E5-620BACADFFE5

21*Apis mellifera*:

22urn:lsid:zoobank.org:act:9082C709-6347-4768-A0DC-27DC44400CB2

23*Bombyx mori* (*Phalæna* (*Bombyx*) *mori*)

24urn:lsid:zoobank.org:act:215466E3-E77F-46E9-8097-372837D7A375

25*Drosophila melanogaster*:

26urn:lsid:zoobank.org:act:5B39F0AA-270D-4AA8-B9A3-C36A3A265910

27

28Keywords: immunity, social insects, termites, selection, comparative genomics

29Abstract

30It has been hypothesized that selection pressure from pathogens plays an important role in shaping
 31social evolution. Social behaviour, in particular brood care, is associated with pathogen pressure in
 32wood-dwelling “lower” termites. Yet, generally pathogen pressure is low in wood-dwelling termite
 33species that never leave the nest except for the mating flight. In comparison, pathogen pressure is
 34higher in species that leave the nest to forage, and thus constantly encounter a diversity of microbes
 35from their environment. We hypothesized that such differences in pathogen pressure are also
 36reflected by differences in the intensity of natural selection on immune genes. We tested this

hypothesis in a phylogenetic framework, analyzing rates of non-synonymous and synonymous substitutions on single-copy immune genes. Therefore, we leveraged recent genomic and transcriptomic data from eight termite species, representing wood-dwelling and foraging species as well as 14 additional species spanning the winged insects (Pterygota). Our results provide no evidence for a role of pathogen pressure in selection intensity on single-copy immune genes. Instead, we found evidence for a genome-wide pattern of relaxed selection in termites.

Introduction

Like in other organisms, pathogens seem to be important drivers of evolution in social insects. On the one hand, social insects are a ‘desirable’ target for pathogens as an insect colony represents a large source of many potential hosts, all with a similar genetic background. Thus, if pathogens manage to enter a colony they can exploit many individuals. However, social insects are also well protected as they evolved ‘social immunity’ (Cremer et al., 2007), a repertoire of defensive mechanisms that work at the colony level. For instance, behavioral task division limits contact to potentially infected individuals and can lead to their eviction. Also molecular mechanisms of social immunity exist, for example indirect immunization of colony members (e.g., Cremer et al., 2007; Masri and Cremer, 2014) or impregnation of the nest walls with fungicidal compounds (Rosengaus et al., 2011). Thus, social immunity can be considered a selected emergent property of insect colonies where the whole is more than the sum of the individuals. Social immunity is an evolved social trait that aligns with the complexity of social organization. So the question might arise, can pathogens even be a driver of complex social organization and does this vary across different social species with different ecologies?

In termites, there is evidence that implies that pathogen load may have been an important driver of brood care when comparing species of early branching lineages of the termite phylogeny (Korb et al., 2012). These species share a similar life type of nesting in a single piece of wood that serves as food and shelter (one-pieces nesting or wood-dwelling termites, see Abe, 1987; Korb, 2007). ‘Workers’ are developmentally totipotent immatures that can become winged as well as wingless reproductives. Brood care by workers is less developed in these wood-dwelling species where all individuals sit inside their food. Interactions are generally not altruistic but cooperative as all individuals function equally as donors and recipients (Korb, 2007). Yet, the degree of brood care seems to be malleable depending on the pathogen load. It is negligible / absent in several drywood termites (Kalotermitidae) that nest in sound dry wood with a low pathogen load (Rosengaus et al., 2003). However, the dampwood termite (Archotermopsidae), *Zootermopsis nevadensis*, which nests in rotten, decaying wood with many fungi and other potential pathogens, show more intensive brood care mainly consisting of allogrooming (Rosengaus et al., 1999; Korb et al., 2012). However, even in the drywood termites, parasites can affect social behavior. Infestations of *Cryptotermes secundus* colonies with mites alter ‘reproductive decision making’ as workers of infested colonies (even if they are not themselves infected) develop into winged sexuals that leave the nest during nuptial flights (Korb and Fuchs, 2006). In the dampwood termite, *Z. nevadensis*, pathogens present a permanent threat which is reflected in huge constitutive investments in immune defense at different levels, the individual as well as the colony level, potentially with a socially acquired immunity (Rosengaus et al., 1999; Traniello et al., 2002; Bulmer et al., 2009; Rosengaus et al., 2011). This is also reflected in *Z. nevadensis*’ genome where six copies of Gram-negative binding proteins (GNBPs) were found, more than in any other insect species (Terrapon et al., 2014). Four of these genes were also found in the genome of the fungus growing termite *Macrotermes natalensis* (Termitidae) and they seem to be termite specific (Poulsen et al., 2014; Korb et al., 2015). *M. natalensis* is not a wood-dwelling species but a foraging species with intensive brood care and complex social organization. It builds the nest

83(mound) separate from the foraging grounds so that individuals leave the nest to forage and bring
84back food to the central nest (i.e., central place foraging). Thus, it also is constantly exposed to
85pathogens, but in contrast to wood-dwelling dampwood termites, it encounters pathogens mainly
86during foraging. For other central-place foraging termites, such as several Australian *Nasutitermes*
87species (Termitidae) which have a complex social organization., GNBPs had previously been shown
88to be under positive selection (Bulmer and Crozier, 2004; Bulmer and Crozier, 2006; Bulmer et al.,
892010).

90Based on these results, the hypothesis was proposed that selective pressure on immune defense genes
91(IGs) differs across termites depending on their life type (Korb et al., 2015). We tested the hypothesis
92that wood-dwelling termites, which do not leave the nest to forage outside show relaxed selection on
93immune genes and fewer signs of positive selection compared to soil-foraging species. Additionally,
94we tested whether within the wood-dwelling life type, the dampwood termite *Z. nevadensis* had
95stronger signs of selection than other wood-dwellers that nest in sound wood.

96In order to test for differences in selective forces acting on IGs between wood-dwelling and foraging
97species, we analyzed a set of 81 previously identified single-copy IGs (see Materials and Methods) in
98eight termite species (published data see Misof et al., 2014; Terrapon et al., 2014; Poulson et al.,
992014; Harrison et al., 2018; Evangelista et al., 2019; data sources are provided in Supplementary
100Table S1). Four of the species we investigated are wood-dwelling species: *C. secundus*, *Incisitermes*
101*marginipennis*, *Prorhinotermes simplex*, and *Z. nevadensis*. The remaining four species are foraging
102species: *Mastotermes darwiniensis*, *Reticulitermes santonensis* (i.e. *R. flavipes*), *Coptotermes* sp. and
103*M. natalensis*. These were analyzed in a phylogenetic framework of 22 species (one mayfly, twelve
104polyneopteran insects including above listed termites and their closest relatives Cryptocercidae, two
105paraneopteran and six holometabolous insects) spanning winged insects to increase the statistical
106power of lineage specific tests for selection.

107Results

108Phylogenetic relationships

109The species tree was inferred from a supermatrix including 1178 single-copy orthologs (SCOs) and
110spanning an alignment length of 555,906 amino acid positions (partition coverage 100%, site-
111completeness score $C_a = 74.61\%$, see Supplementary Materials and Supplementary Figure S1).
112Termites were monophyletic with *Cryptocercus* as sister group, consistent with earlier work (e.g., Lo,
1132000; Klass, 2006; Inward et al., 2007; Legendre et al., 2008). Phylogenetic relationships within
114termites are largely consistent with Evangelista et al. (2019) and are statistically maximally supported
115(Figure 1). Consistent with earlier work (e.g., Legendre et al., 2008), neither wood-dwellers nor
116foragers constitute monophyletic groups, confirming that several independent switches in life type
117were included in our analyses. More details on phylogenetic analyses are provided in the
118Supplementary Materials.

119Patterns of selection on termite immune genes

120Between 13 and 78 single-copy orthologs of IGs per species were included in the analyses (Table 1).
121We found no evidence for positive selection on the IGs (see Supplementary Table S2).

122Next, we tested the hypothesis that selection on the IGs of wood-dwelling species is relaxed
123compared to foragers. We found 62 cases of significantly relaxed selection across all termite species
124analyzed (Table 1, Supplementary Table S3). There was no evidence for a difference in the number of

125IGs under relaxed selection between the life types (generalized linear mixed effects model with
126binomial error distribution: $df = 7$, $z = -0.191$, $P = 0.85$). There was also no evidence for differences
127between species (generalized linear model assuming binomial error distribution: $df = 7$, $z = -1.74 -$
1280.729, $P = 0.082$, ranges of z and P are for the different species). Because changes in the selection
129intensity on IGs could be obscured by genome wide differences in selective constraint, it is important
130to test these hypotheses against the genomic background. To this end, we generated sets of
131background genes (BGs) that consisted of genes matching the GC-content and sequence length of
132each IG for each species (see Materials and Methods). The number of IGs under relaxed selection did
133not differ significantly from that expected from the analysis of the BGs for any of the species (see
13495% confidence intervals (CIs) for BG sets in Table 1). In order to take genome-wide effects of
135selective constraint into account, when comparing selection intensity between wood-dwellers and
136foragers, we (i) calculated the ratio of genes under significantly relaxed selection between wood-
137dwellers and foragers for IGs and (ii) compared this ratio to that expected from BGs (see Materials
138and Methods). If selection is relaxed specifically in the IGs of wood-dwellers relative to their
139genomic background, we expect the ratio of the number of significant genes under relaxed selection
140between wood-dwellers and foragers to be larger for the IGs than for the BGs. However, the ratio of
141the number of IGs under relaxed selection between wood-dwellers and foragers did not differ
142significantly from the expectation derived from BGs (Figure 2A), supporting the view that patterns of
143relaxed selection on IGs follow genome wide trends in selective constraint.

144Because we did not find any evidence for an increase in the number of IGs under significantly
145relaxed selection in wood-dwellers, we reasoned that a putative signal of relaxation of selective
146constraint might be more diffuse and only become visible as a general trend over all IGs investigated.
147In order to capture such more general trends, we assessed potential differences in k , a measure for the
148intensity of selection, for all IGs between life types. K did not differ significantly between species
149(Kruskal-Wallis test: $df = 7$, $X^2 = 3.95$, $n = 322$, $P = 0.79$, for n per species see Table 1) nor was it
150lower for wood-dwellers (Mann-Whitney U test, one-sided: $U = 11690$, $n = 322$, $P = 0.41$). This
151indicated similar selection intensity on IGs for wood-dwellers and foragers. For the comparison of k
152between life types it is, as above, important to take the selective constraint on the genomic
153background into account. K for the IGs did not differ significantly from k for sets of BGs for any of
154the species investigated (Table 1). Following the same rationale as above, we used the ratio of
155medians of k between wood-dwellers and foragers as a test statistic. This ratio can be interpreted as
156the relative intensity of selection between wood-dwellers and foragers. We found that the relative
157intensity of selection between wood-dwellers and foragers on IGs matched that of the BG sets
158(Figure 2B), again suggesting that the IGs follow genome-wide trends of selective constraint.

159Finally, we hypothesized that our results might be affected by the particular selection pressures that
160act on *Z. nevadensis*, which is a wood-dwelling species, but lives in relatively microbe-rich damp
161wood (Rosengaus, 2003). Hence, selection would be expected to be stronger on *Z. nevadensis* IGs
162than on IGs in the other wood-dwellers, resulting in a smaller fraction of genes under significantly
163relaxed selection in *Z. nevadensis*. Contrary to this expectation, we found significantly more IGs
164under relaxed selection in *Z. nevadensis* (generalized linear model assuming binomial error
165distribution: $df = 3$, $z = 2.53$, $P = 0.01$). However, when we excluded *Z. nevadensis* from the
166comparison between life types, there was also no difference in the number of IGs under relaxed
167selection (generalized linear mixed effects model with binomial error distribution: $n = 8$, $z = -0.94$, P
168= 0.35), showing that our results are robust. The overall intensity of selection on IGs (k) did not differ
169significantly between *Z. nevadensis* and the other wood-dwellers (Mann-Whitney U test, $U = 5471$, n
170= 215, $P = 0.77$). Similarly, it could be argued that the assumption of relaxed selection only holds for
171the dry wood-dwellers *C. secundus* and *I. marginipennis*, assuming that only dry wood is a truly

172pathogen poor substrate. We could not find a difference in the number of genes under relaxed
173selection between the dry wood-dwellers and the other species (generalized linear model with
174binomial error distribution: $df = 7$, $z = 0.73$, $P = 0.464$) nor for the intensity of selection over all IGs
175as measured by k (Mann-Whitney U test: $U = 10538$, $n = 322$, $P = 0.37$).

176To our surprise, we observed that median k for IGs was smaller than one for seven of the eight
177investigated termite species (Table 1). This difference is significant (Mann-Whitney U test: $U =$
17820958, $n = 322$, $P < 0.01$), indicating an overall relaxation of selection. This signal was not IG
179specific: k for the sets of BGs was also significantly smaller than 1 for all species ($P = 4 \times 10^{-18}$ -
1804.8 $\times 10^{-5}$), suggesting a genome-wide relaxation of selection on the termite lineages compared to the
181background branches of the phylogeny.

182Discussion

183In this study we combined recent genomic and transcriptomic resources to test whether termite
184ecology, in particular exposure to pathogens, might affect the evolution of immune genes.
185Surprisingly, and in contrast to studies from *Drosophila* (Clark et al., 2007; Sackton et al. 2007; Hill
186et al., 2019) we could not find evidence for positive selection on the immune genes for eight termite
187species. We extended our analyses, employing recent tests to explicitly assess relaxed selection
188(Wertheim et al., 2015), revealing 62 cases of significantly relaxed selection in IGs.

189We expected that the intensity of selection would differ between termites of the wood-dwelling and
190foraging life types, as foraging termites are assumed to experience higher selection pressure from
191pathogens due to higher exposure. Contrary to our expectation, we did not detect an effect of life type
192on signs of selection for immune genes. Neither did wood-dwelling species differ from the soil
193foraging species, nor did the dampwood termite *Z. nevadensis* differ from the other wood-dwelling
194termites. *Z. nevadensis* that has a higher microbial load compared to the other wood-dwellers
195(Rosengaus, 2003), even had an elevated number of IGs, for which selection was significantly
196relaxed. However, the number of IGs under relaxed selection in *Z. nevadensis* did not differ from the
197expectation derived from BGs, suggesting that this is not an IG specific signal. Taken together, this
198suggests that individual immune response remains an important trait in termites even if social
199immunity is effective or pathogen exposure is low to moderate. A possible explanation for similar
200selection pressure on immune genes between wood-dwelling and foraging species is that both harbor
201complex gut microbial communities that are essential for termite survival (see Waidele et al., 2017).
202It seems reasonable to assume that the immune system is likely to play a role in modulating these
203communities in wood-dwellers and foragers alike resulting in constant selection pressure.

204We applied state of the art methods (Pond et al., 2005; Murrell et al., 2015; Wertheim et al., 2015)
205and carefully controlled our study by analyzing background genes that matched IGs in GC-content
206and sequence length (see Materials and Methods). We also combined, to our knowledge, the largest
207data set of termite genomic and transcriptomic resources for our study that has been used in that
208context so far. Nonetheless, our study might lack statistical power: First, we have relatively few
209species for life type comparisons (4 vs 4). Second, for five of these species (*M. darwiniensis*, *I.*
210*marginipennis*, *R. santonensis*, *P. simulans* and *Coptotermes* sp.) only transcriptome data are
211available. Naturally, the number of genes under significantly relaxed selection is lower for the species
212for which only transcriptome data are available, as fewer genes could be annotated and investigated
213in the transcriptome data (Table 1, 2 and Supplementary Table S4). However, potential biases in data
214availability were taken into account in the procedure to sample sets of matching background genes.

When taking the genomic background and the sampling procedure into account (Figure 2), our data provide no evidence that selection intensity on single-copy ortholog IGs differs between life types. Do our results imply that selection pressure on immune genes does not differ between termite life types? No, we cannot conclude this as our study excluded immune genes that are not single-copy orthologs. Antimicrobial peptides (AMPs) or Gram-negative binding proteins (GNBPs) that have several copies within the genome of some species were not addressed in our study. Different copies can be caste-specifically expressed in termites, as shown for *Z. nevadensis* (Terrapon et al., 2014), and for *Reticulitermes speratus* (Mitaka et al., 2017), and they may be under positive selection as indicated, for instance, by studies on GNBPs for several foraging *Nasutitermes* species (Bulmer and Crozier, 2006; Bulmer et al., 2010). Note that the original hypothesis for a difference in immune defense between wood-dwellers and foraging termites considered specifically GNBPs and AMPs (Korb et al., 2015). Thus, more studies are warranted that test selection on such multi-copy genes. However, orthology is difficult to infer for multi-copy genes making the restriction to SCOs a standard procedure (e.g., Pauli et al., 2016; Dowling et al., 2016; Mitterboeck et al., 2017; Ran et al., 2018; Brandt et al., 2019; Hill et al., 2019). Thus, we restricted our analysis to single-copy orthologs because orthology is a basic assumption of the current methods that identify selection in a powerful phylogenetic framework (e.g. codeML implemented in PAML, Yang, 2007; HyPhy applying BUSTED, see Murrell et al., 2015, and RELAX, see Wertheim et al., 2015).

Applying state of the art methods on the most comprehensive termite data set for which genome or transcriptome data are currently available, we found no evidence for differences in selection on immune genes that correlate with termite ecology, specifically pathogen exposure. Our results suggest that the putative evolutionary response to differences in pathogen exposure can not be found in single-copy immune genes. Interestingly, we detected a signal of overall relaxation of selective constraint in termites. This could be related to smaller effective population sizes due to their social organization opening up an interesting new line of research (Rubinstein et al., 2019).

240

241 **Materials and Methods**

242 **Identifying orthologous sequence groups of protein-coding single-copy genes**

As basis to identify single-copy orthologs (SCOs), we designed an ortholog set from official gene sets (OGS) from available full (draft) genomes of four species: *Cryptotermes secundus*, *Macrotermes natalensis*, *Zootermopsis nevadensis*, and *Blattella germanica* (Supplementary Table 1). The set of single-copy orthologs was inferred with the software OrthoFinder v.1.1.4 (Emms and Kelly, 2015) using default settings. As input, the OGS of respective species were downloaded from public databases as amino acid and nucleotide sequences. The OGS of *C. secundus* was kindly provided by the *C. secundus* consortium (via J. Korb) before it was published (Harrison et al., 2018). We only kept the longest isoform per orthologous group (OG). All OGs that included the amino-acid Selenocysteine (U) were removed since they might be putative pseudo-genes and to avoid difficulties in downstream analyses as many software packages are not able to handle Selenocysteine. This was done using the package Biobundle (script isoformCleaner with boost 1.61.0 environment (Kemena, 2017, available from <https://github.com/CarstenK/BioBundle>). SCOs inferred with OrthoFinder, were summarized with custom-made Python scripts (kindly provided by A. Faddeeva and L. Wissler, available upon request). This resulted in a set of 5382 SCOs across the four selected species.

257Taxon sampling

258We included the four reference species that were used to create the ortholog set as well as genome
259and transcriptome data of 18 additional species in our analyses. Eight of the included species are
260termites: *Coptotermes* sp., *Incisitermes marginipennis*, *Mastotermes darwiniensis*, *Prorhinotermes*
261*simplex*, and *Reticulitermes santonensis* (with published transcriptome data); *C. secundus*, *Z.*
262*nevadensis* and *M. natalensis* (with published OGS), see Evangelista et al. (2019). Other included
263species were a representative of Cryptocercidae as it is supposed to be the sister group of termites
264(e.g., Lo et al., 2000; Inwards et al., 2007), two other non-social cockroach species, and
265representatives from other polyneopteran, paraneopteran and holometabolous insects, and a mayfly as
266outgroup (Adams et al., 2000; Bonasio et al., 2010; Elsik et al., 2014; Evangelista et al., 2019;
267Harrison et al., 2018; International Aphid Genomics Consortium, 2010; Misof et al., 2014; Mesquita
268et al., 2015; Pauli et al., 2016; Poulsen et al., 2014; Richards et al., 2008; Sinkins, 2007; Terrapon et
269al., 2014; Wang et al., 2014; Xia et al., 2004; the full species list is provided in Supplementary Table
270S1). Access to transcriptome data (see Figure 1) was kindly granted by 1KITE before they were
271published, access to the OGS of the locust and the mayfly was granted by the i5K community.

272Assignment of putative orthologous transcripts to the SCOs

273The ortholog set was used as input for the assignment of putative SCOs (provided as Supplementary
274files on DRYAD). Inference and assignment of putative orthologs from genome and transcriptome
275data of the 18 species that were not included for generating the ortholog set was performed with
276Orthograph v.0.6.1 (Petersen et al., 2017). Orthograph is recommended to infer orthologs from
277transcriptome data for which no official gene set are available (see Petersen et al., 2017). Orthograph
278analyses resulted in 5366 SCOs that were identified in at least one species that was not used as
279reference species to create the ortholog set.

280Multiple sequence alignments, species tree inference and testing for selection

281Individual SCOs were aligned at the amino acid level with MAFFT v7.310 using the L-INS-i
282algorithm (Kato and Standley, 2013).

283Species tree inference

284For inferring the species tree, we only kept those SCOs that were present in all 22 species. This
285resulted in 1178 SCOs. Ambiguously aligned sections on the amino acid level were identified with
286Aliscore v2.2 (Misof and Misof, 2009; Kück et al., 2010) (settings: -r with all pairwise sequence
287comparisons, -e for gap-rich alignments, otherwise defaults) and masked with Alicut v2.3 (Kück,
2882011). Masked amino-acid multiple sequence alignments (MSAs) were concatenated into a
289supermatrix with FASconCAT-G v1.02 (Kück and Longo, 2014). We inferred phylogenetic
290relationships using a maximum-likelihood (ML) approach with IQtree v1.5.4 (Nguyen et al., 2014;
291Chernomor et al., 2016). Statistical support was determined from 200 non-parametric, slow and
292thorough bootstrap replicates. We ensured bootstrap convergence with *a posteriori* bootstrap criteria
293(Pattengale et al., 2010) as implemented in RAxML (Stamatakis, 2014), v.8.2.11. The best ML tree,
294out of 50 inferred trees, which all showed an identical topology, was rooted with *Ephemera danica*
295using SeaView v.4.5.4 (Gouy et al., 2010; note that multiple tree viewers are not reliable, see Czech
296et al., 2017); trees were graphically edited with Inkscape (v.0.91) (www.inkscape.org). More details
297on the procedure of phylogenetic inference are provided in the Supplementary Materials.

298Inferring natural selection

299 Alignment processing and clean-up

300 Methods to identify selection are sensitive to misalignments (Privman et al., 2012; Markova-Raina
301 and Petrov, 2011). Therefore we performed extensive alignment cleanup. First, we identified and
302 deleted badly aligned or gap-rich sequences on amino acid level with MaxAlign v1.1 (Gouveia-
303 Oliveira et al., 2007). This procedure resulted in five SCOs with only one sequence which were
304 excluded from further analyses. We subsequently compiled corresponding nucleotide (i.e. codon)
305 MSAs with PAL2NAL (Suyama et al., 2006, v14.1, see Misof et al., 2014) using the 5361 amino-
306 acid MSAs as blue-print. The nucleotide MSAs were then used for all following analyses. Second,
307 we deleted all SCOs with less than four sequences (223 SCOs) leaving 5138 SCOs. Third, we
308 identified ambiguously aligned sections on amino acid level with Aliscore v2.2 (Misof and Misof,
309 2009; Kück et al., 2010) with the same settings as described for the species tree inference. Suggested
310 sections were removed from the amino acid and correspondingly from the nucleotide MSAs with
311 Alicut v2.3 (Kück, 2011). Subsequent analyses were performed on the masked nucleotide MSAs.
312 First, we classified 5138 SCOs into 86 immune single-copy genes (IGs) and into the remaining 5052
313 SCOs based on Terrapon et al. (2014, see Table S25 for *Z. nevadensis*) and Korb et al. (2015, for *Z.*
314 *nevadensis* and *M. natalensis*), see Supplementary Table S5. The 5052 non-immune SCOs were used
315 to generate gene sets from the genomic background, i.e. background genes (BGs) that had similar
316 GC-content and sequence length (see below) as the examined IGs. Note that from the 86 IGs
317 (Supplementary Table S5) five IGs were excluded because there was no SCO fulfilling the criteria to
318 serve as BG and these were not listed by Terrapon et al. (2014) or not reported by Korb et al. (2015).
319 This left 81 IGs for analyses (for detailed information see Supplementary Materials).

320 To further reduce potential false positives that may originate from misalignments, we trimmed
321 trailing ends of each multiple sequence alignment, i.e. each multiple sequence alignment started and
322 ended with unambiguous nucleotides for all species. Because visual inspection of the trimmed MSAs
323 still revealed putative misaligned nucleotides, we applied the GUIDe tree based AlignMent
324 Confidence approach (GUIDANCE) Guidance2 (Landan and Graur, 2008; Sela et al., 2015) version
325 2.02 using MAFFT as implemented alignment method on the trimmed MSAs (options: codon as
326 sequence type, sequence cutoff=0 and the default column cutoff=0.93).

327 Inferring positive selection and selection intensity

328 To test for evidence of positive selection we used BUSTED (Murrell et al., 2015) as implemented in
329 the software package HyPhy (Pond et al., 2005). BUSTED uses a branch-site test for positive
330 selection on entire genes in a foreground branch relative to the background branches in a phylogeny.
331 A significant *P*-value means that at least one codon in the foreground branch has experienced an
332 episode of positive selection. The high sensitivity of the method compared to tests from alternative
333 packages (see e.g., Enard et al., 2016; Ebel et al., 2017; Venkat et al., 2018; Hill et al., 2019) and the
334 option to define the foreground branches according to our research question made it perfectly suited
335 for our study.

336 For inferring potential relaxation of selection, we used RELAX (Wertheim et al., 2015) as
337 implemented in the software package HyPhy (Pond et al., 2005). RELAX has been designed to
338 identify changes in the intensity of selection on a given protein-coding gene in a codon-based
339 phylogenetic framework (see Wertheim et al., 2015). The basic expectation of RELAX is that under
340 relaxed selection, the ω of sites under purifying and positive selection will move closer to neutrality.
341 The change of ω for the selected sites relative to the background branches is quantified with the
342 selection intensity parameter k , where

$$f(\omega, k) = \omega^k.$$

If parameter k is significantly larger than one, selection has been intensified along the test branches.
If k is significantly smaller than one, selection has been relaxed.

We used BUSTED and RELAX as implemented in the software package HyPhy, version 2.4.0-alpha.2 (access: April 2019). We determined k for each of the termite species relative to all other species in the tree. We chose this species-wise analysis setup because we wanted to take species level differences in pathogen exposure into consideration.

Comparison of IG selection parameter to the genomic background

To test whether or not signals of selection were specific to IGs or the consequence of genome-wide trends, we generated sets of background genes (BGs). To this end, we searched the 5052 non-immune SCOs for genes that closely matched the GC-content ($\pm 5\%$) and the sequence length ($\pm 5\%$). Following this procedure, we generated lists of matching BGs for each IG and each species. From these lists, we randomly sampled 100 gene sets such that there was a matching BG for each IG that was analyzed in the respective species (e.g., for *Z. nevadensis*, 78 IGs were analyzed, thus each of the 100 sets of BGs contained therefore 78 BGs, see also Table 2. Lists of analyses BG that are similar in GC-content and sequence length of the IGs are provided for each species as Supplementary files on DRYAD). We performed BUSTED and RELAX analyses for each termite species on all IGs and RELAX on the species' respective BG set with default settings.

All analyses were performed on Linux Desktop PCs at the University of Freiburg, Germany and on the Linux HPC CSIRO Cluster Pearcey, Australia. Analyses results of all IGs are summarized in Supplementary Table S2, S3; results of BGs are provided species-wise on DRYAD.

Statistical analyses

In order to assess potential differences in selection intensity on the IGs between species and life types, we summarized the RELAX results by counting the number of genes under significantly relaxed selection: genes with $k < 1$ and $P < 0.05$ were considered. Potential differences were tested for statistical significance with generalized linear models with binomial error distribution using the functions `glm` and `glmer` from the `lme4` R package (Bates et al., 2015, version 1.1-21 with R (2018), version 3.4.4). The number of significant genes divided by the total number of genes analyzed was used as response variable. Species or life type were used as potential predictors. Varying sampling depths between species, as represented by the number of IGs analyzed per species, were taken into account as weights in the model. When comparing life types, species were treated as a random effect. See Supplementary file (RanalysiscriptfortermiteIGs.R on DRYAD) for a detailed R analysis script with all models, commands and functions used.

We also analyzed parameter k to search for more diffuse trends in selection intensity that are distributed over the IGs so that individual IGs do not reach significance. According to its definition, k should map linearly on a logarithmic scale. However, we found six strong outliers on the logarithmic scale that were more than standard deviations away from the mean ($\log(k) < -9$, see Supplementary Table S3) that could make the analysis in a linear framework error prone. Visual inspection of the alignments underlying these extremely small values of k did not reveal any obvious misalignments that would justify their exclusion. Therefore, potential differences in k were assessed with non-parametric tests (Mann-Whitney U test, Kruskal-Wallis test) that are robust to outliers.

Genome-wide trends in selection intensity can potentially obscure IG specific patterns or generate false positives. For example, changes in population size can affect the efficiency of both purifying and positive selection (Ohta, 1973) on a genome wide scale. Population sizes might differ between species and life types in our study depending on reproductive rates and degrees of sociality. Therefore, it is essential to put the results for IGs into the context of the genomic background. To this end, we generated expected values for the number of significant genes and for k based on 100 sets of BGs (see above) per termite species, representing the genomic background. The median of k and 95% CIs from the BG-based distributions for each species were calculated with R (version 3.4.4), using the median and quantile functions with standard settings. In order to compare differences between life types while taking the genomic background into account, we calculated (i) the ratio of the number of genes that were significantly relaxed between wood dwellers and foragers

$$R_{relax} = \frac{\text{number of genes}_{wood-dwellers} | P < 0.05 \cap k < 1}{\text{number of genes}_{foragers} | P < 0.05 \cap k < 1}$$

and (ii) the ratio of median parameter k (i.e., relative selection intensity):

$$R_k = \frac{(\tilde{k} \text{ genes}_{wood-dwellers})}{(\tilde{k} \text{ genes}_{foragers})}$$

These ratios were calculated for the IGs and the BGs. Then the ratio of the IGs was compared to their expectation from the BGs. Significant shifts in selection intensity that are specific to IGs should lead to shifts of R_{relax} and R_k only for IGs. Thus, if there were IG specific patterns of relaxed selection, the ratios R_{relax} and R_k for the IGs should represent extremes of the distribution of sets of BGs.

Data Availability Statement

Supplementary data (see Supplementary Materials) used in this study can be found at the DRYAD digital repository, DOI XXX. Will be available upon manuscript acceptance.

Funding

This project was supported by the German Science Foundation (DFG; KO1895/20-1, KO1895/20-2, STA 1154/2-1 Projektnummer 270882710).

Author Contributions

JK and FS conceived the study; KM, MS, FS performed all analyses; FS, KM and JK wrote the manuscript. All authors edited and approved the Manuscript.

416 Conflict of Interest Statement

417 The authors declare that the research was conducted in the absence of any commercial or financial
418 relationships that could be construed as a potential conflict of interest.

419

420 Acknowledgments

421 We acknowledge Bernhard Misof, Panos Provataris, Coby Schal, Xavier Belles, Stephen Richards
422 and the i5K community the usage of the official gene set of *Ephemera danica*, and *Blattella*
423 *germanica*, and Xianhui Wang and Le Kang for access and usage of the official gene set of *Locusta*
424 *migratoria*. We also thank the 1KITE Dictyoptera group who granted us access to transcriptome
425 assemblies before they were published. We thank David Enard (University of Arizona, USA), Ondrej
426 Hlinka (CSIRO, Australia), Ryan Velazquez and Sergej Pond for their useful input and help with
427 analyses of the molecular sequence data using Guidance2 and HyPhy. KM thanks Thomas Pauli
428 (University of Freiburg, Germany) for fruitful discussions and Hans Pohl (University of Jena,
429 Germany), who kindly allowed the usage of several pictograms in Figure 1 (adopted from Misof et
430 al., 2014).

431

432 Supplementary Materials

433 The Supplementary Materials for this article can be found online at:

434

435 References

- Abe, T. (1987). "Evolution of life types in termites" in Evolution and coadaptation in biotic communities, eds. S. Kawano S, J. H. Connell, and T. Hidaka (University of Tokyo press, Tokyo), 125–148.
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science*. 287, 2185–2195. doi: 10.1126/science.287.5461.2185
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi:10.18637/jss.v067.i01
- Bonasio, R., Zhang, G., Ye, C., Mutti, N.S., Fang, X., Qin, N., et al. (2010). Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* 329, 1068–1071. doi: 10.1126/science.1192428
- Brandt, A., Bast, J., Scheu, S., Meusemann, K., Donath, A., Schütte, K., et al. (2019). No signal of deleterious mutation accumulation in conserved gene sequences of extant asexual hexapods. *Sci. Rep.* 9, 5338. doi: 10.1038/s41598-019-41821-x
- Bulmer, M. S., and Crozier, R. H. (2004). Duplication and diversifying selection among termite antifungal peptides. *Mol. Biol. Evol.* 21, 2256–2264. doi: 10.1093/molbev/msh236

- Bulmer, M. S., and Crozier, R. H. (2006). Variation in positive selection in termite GNBPs and Relish. *Mol. Biol. Evol.* 23, 317–326. doi: 10.1093/molbev/msj037
- Bulmer, M. S., Bachelet, I., Raman, R., Rosengaus, R. B., and Sasisekharan, R. (2009). Targeting an antimicrobial effector function in insect immunity as a pest control strategy. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12652–12657. doi: 10.1073/pnas.0904063106
- Bulmer, M. S., Lay, F., and Hamilton, C. (2010). Adaptive evolution in subterranean termite antifungal peptides. *Insect Mol. Biol.* 19, 669–674. doi: 10.1111/j.1365-2583.2010.01023.x
- Chernomor, O., von Haeseler, A., and Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* 65, 997–1008. doi: 10.1093/sysbio/syw037
- Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C., et al. (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450, 203–218. doi: 10.1038/nature06341
- Cremer, S., Armitage, S. A. O., and Schmid-Hempel, P. (2007). Social immunity. *Curr. Biol.* 17, R693–702. doi: 10.1016/j.cub.2007.06.008
- Ebel, E. R., Telis, N., Venkataram, S., Petrov, D. A., and Enard, D. (2017). High rate of adaptation of mammalian proteins that interact with *Plasmodium* and related parasites. *PLoS Genet.* 13:e1007023. doi: 10.1371/journal.pgen.1007023
- Enard, D., Cai, L., Gwennap, C., and Petrov, D. A. (2016). Viruses are a dominant driver of protein adaptation in mammals. *ELife* 5:e12469. doi: 10.7554/eLife.12469
- Czech, L., Huerta-Cepas, J., and Stamatakis, A. (2017). A critical review on the use of support values in tree viewers and bioinformatics toolkits. *Mol. Biol. Evol.* 34, 1535–1542. doi: 10.1093/molbev/msx055
- Dowling, D., Pauli, T., Donath, A., Meusemann, K., Podsiadlowski, L., Petersen, M., et al. (2016). Phylogenetic origin and diversification of RNAi pathway genes in insects. *Genome Biol. Evol.* 8, 3784–3793. doi: 10.1093/gbe/evw281
- Elsik, G. C., Worley, K. C., Bennet, A. K., Beye, M., Camara, F. Childers, C. P., et al. (2014). Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genom.* 30:86. doi: 10.1186/1471-2164-15-86
- Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157. doi: 10.1186/s13059-015-0721-2
- Evangelista, D. A., Wipfler, B., Béthoux, O., Donath, A., Fujita, M., Kohli, M., et al. (2019). An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). *Proc. Biol. Sci.* 286, 20182076. doi: 10.1098/rspb.2018.2076
- Gouveia-Oliveira, R., Sackett, P. W., and Pedersen, A. G. (2007). MaxAlign: maximizing usable data in an alignment. *BMC Bioinformatics* 8:312. doi: 10.1186/1471-2105-8-312

- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259
- Harrison, M. C., Jongepier, E., Robertson, H. M., Arning, N., Bitard-Feildel, T., Chao, H., et al. (2018). Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat. Ecol. Evol.* 2:557–566. doi: 10.1038/s41559-017-0459-1
- Hill, T., Koseva, B. S., and Unckless, R. L. (2019). The genome of *Drosophila innubila* reveals lineage-specific patterns of selection in immune genes. *Mol. Biol. Evol.* 36, 1405–1417. doi: 10.1093/molbev/msz059
- International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8:e1000313, doi: 10.1371/journal.pbio.1000313
- Inward, D., Beccaloni, G. and Eggleton, P. (2007). Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* 3, 331–335. doi: 10.1098/rsbl.2007.0102
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kemena, C. (2017). *BioBundle*. Available from <https://github.com/CarstenK/BioBundle>
- Klass, K.-D., and Meier, R. (2006). A phylogenetic analysis of Dictyoptera (Insecta) based on morphological characters. *Entomol. Abh. (DRES.)* 63, 3–50.
- Korb, J., and Fuchs, A. (2006) Termites and mites - adaptive behavioural responses to infestation? *Behaviour*, 143, 891–907.
- Korb, J. (2007). Termites. *Curr. Biol.* 17, R995–999. Doi: 10.1016/j.cub.2007.10.033
- Korb, J., Buschmann, M., Schafberg, S., Liebig, J., and Bagneres, A. G. (2012). Brood care and social evolution in termites. *Proc. Biol. Sci.* 279, 2662–2671. doi: 10.1098/rspb.2011.2639
- Korb, J., Poulsen, M., Hu, H., Li, C., Boomsma, J.,J., Zhang, G., et al. (2015). A genomic comparison of two termites with different social complexity. *Front. Genet.* 6:e9. doi: 10.3389/fgene.2015.00009
- Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B. M., Waegle, J. W., et al. (2010). Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front. Zool.* 7:10. doi: 10.1186/1742-9994-7-10
- Kück, P. (2011). *Alicut* v. 2.3. Bonn, ZFMK. Available from <https://github.com/PatrickKueck/AliCUT>
- Kück, P., and Longo, G.C. (2014). FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front. Zool.* 11:81. doi: 10.1186/s12983-014-0081-x

- Legendre, F., Whiting, M. F., Bordereau, C., Canello, E. M., Evans, T.A., and Grandcolas, P. (2008). The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for the evolution of the worker and pseudergate castes, and foraging behaviors. *Mol. Phylogenet. Evol.* 48, 615–627. doi: 10.1016/j.ympev.2008.04.01
- Landan, G., and Graur, D. (2008). Local reliability measures from sets of co-optimal multiple sequence alignments. *Pac. Symp. Biocomput.* 13:15–24.
- Lo, N., Tokuda, G., Watanabe, H., Rose, H., Slaytor, M., Maekawa, K., et al. (2000). Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Curr. Biol.* 10, 801–804. doi: 10.1016/S0960-9822(00)00561-3
- Markova-Raina, P., and Petrov, D. (2011). High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Res.* 21, 863–874. doi: 10.1101/gr.115949.110
- Masri, L., and Cremer, S. (2014). Individual and social immunisation insects. *Trends Immunol.* 35, 471–482. doi: 10.1016/j.it.2014.08.005
- Mesquita, R. D., Vionette-Amaral, R. J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F. A., Minx, P., et al. (2015). Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc. Natl. Acad. Sci. U. S. A.* 112, 14936–14941. doi: 10.1073/pnas.1506226112
- Mitaka, Y., Kobayashi, K., and Matsuura, K. (2017). Caste-, sex-, and age-dependent expression of immune-related genes in a Japanese subterranean termite, *Reticulitermes speratus*. *PLoS One* 12:e0175417. doi: 10.1371/journal.pone.0175417
- Mitterboeck, T. F., Liu, S., Adamowicz, S. J., Fu, J., Zhang, R., Song, W., et al. (2017). Positive and relaxed selection associated with flight evolution and loss in insect transcriptomes. *Gigascience* 10, 1–14. doi: 10.1093/gigascience/gix073
- Misof, B., and Misof, K. (2009). A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Syst. Biol.* 58, 21–34. doi: 10.1093/sysbio/syp006
- Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767. doi: 10.1126/science.1257570
- Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., et al. (2015). Gene-wide identification of episodic selection. *Mol. Biol. Evol.* 32, 1365–1371. doi: 10.1093/molbev/msv035
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Otha, T. (1973). Slightly Deleterious Mutant Substitutions in Evolution. *Nature* 246, 96–98.

- Pattengale, N.D. , Alipour, M., Bininda-Emonds, O.R., Moret, B.M., and Stamatakis, A. (2010). How many bootstrap replicates are necessary? *J. Comput. Biol.* 17, 337–354. doi: 10.1089/cmb.2009.0179
- Pauli, T., Vedder, L., Dowling, D., Petersen, M., Meusemann, K., Donath, A., et al. (2016). Transcriptomic data from panarthropods shed new light on the evolution of insulator binding proteins in insects. *BMC Genomics* 17:861. doi: 10.1186/s12864-016-3205-1
- Petersen, M., Meusemann, K., Donath, A., Dowling, D., Liu, S., Peters, R. S., et al. (2017). Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinf.* 18:111. doi: 10.1186/s12859-017-1529-8
- Privman, E., Penn, O., and Pupko, T. (2012). Improving the performance of positive selection inference by filtering unreliable alignment regions. *Mol. Biol. Evol.* 29, 1–5. doi: 10.1093/molbev/msr177
- Pond, S. L. K., Frost, S. D. W. and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079
- Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S., et al. (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14500–14505.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ran, J.-H., Shen, T.-T., Wang, M.-M., and Wang, X.-Q. (2018). Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Biol. Sci.* 285, 20181012. doi: 10.1098/rspb.2018.1012
- Tribolium Genome Sequencing Consortium, Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S. J., Denell, R., et al. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452, 949–955. doi: 10.1038/nature06784
- Rosengaus, R. B., Traniello, J. F., Chen, T., and Brown, J. J. (1999). Immunity in a social insect. *Naturwissenschaften*, 86, 588–591.
- Rosengaus, R. B., Moustakas, J. E., Calleri, D. V, and Traniello, J. F. A. (2003). Nesting ecology and cuticular microbial loads in dampwood (*Zootermopsis angusticollis*) and drywood termites (*Incisitermes minor*, *I. schwarzi*, *Cryptotermes cavifrons*). *J. Insect Sci.* 3, 31.
- Rosengaus, R. B., Traniello, J. F. A., and Bulmer, M. S. (2011). “Ecology, behavior and evolution of disease resistance in termites” in *Biology of termites: a modern synthesis*, eds. D. E. Bignell, Y. Roisin, and N. Lo (Springer Dordrecht, Heidelberg, London, New York), 165–192.
- Rubenstein, D. R., Ågren, J. A., Carbone, L., Elde, N. C. Hoekstra, H. E., Kapheim, K. M., et al. (2019). Coevolution of Genome Architecture and Social Behavior. *Trends Ecol. Evol.* In press. pii: S0169–5347(19)30133-8. doi: 10.1016/j.tree.2019.04.011

- Sackton, T. B., Lazzaro, B. P., Schlenke T. A., Evans J. D., Hultmark, D., and Clark, A. G. (2007). Dynamic evolution in the innate immune system in *Drosophila*. *Nature Genet.*, 39, 1461–1468. doi: 10.1038/ng.2007.60
- Sela, I., Ashkenazy, H., Katoh, K. and Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* 43, W7–W14. doi: 10.1093/nar/gkq443
- Sinkins, S. (2007). Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316, 1718–1723. doi: 10.1126/science.1138878
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315
- Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., et al. (2014). Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* 5:3636. doi: 10.1038/ncomms4636
- Traniello, J. F. A., Rosengaus, R. B., and Savoie, K. (2002). The development of immunity in a social insect: evidence for the group facilitation of disease resistance. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6838–6842. doi: 10.1073/pnas.102176599
- Venkat, A., Hahn, M. W., and Thornton J. W. (2018). Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nat. Ecol. Evol.* 2:1280–1288. doi: 10.1038/s41559-018-0584-5
- Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., et al. (2014). The locust genome provides insight into swarm formation and long-distance flight. *Nat. Commun.* 5:2957. doi: 10.1038/ncomms3957
- Waidele, L., Korb, J., Voolstra, C. R., Künzel, S., Dedeine, F., and Staubach, F. (2017). Differential ecological specificity of protist and bacterial microbiomes across a set of termite species. *Front. Microbiol.* 8:2518. doi: 10.3389/fmicb.2017.02518
- Wertheim, J. O., Murrell, B., Smith, M. D., Pond, S. L- K., and Scheffler, K. (2015). RELAX: Detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* 32, 820–832. doi: 10.1093/molbev/msu400
- Xia Q1, Zhou Z, Lu C, Cheng D, Dai F, Li B, et al. (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*. 306, 1937–1940. doi: 10.1126/science.1102210
- Yang, Z. (2007): PAML 4: Phylogenetic analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24, 1586–1591.

436Tables

Table 1. Results of RELAX analyses of changes in selection intensity on termite immune genes (IGs). Genes that are under significantly relaxed selection were counted ($k < 1$, $P < 0.05$) for columns containing #. BG: background single-copy ortholog. * indicates species with annotated genomes.

species	life type	# relaxed IGs	# total IGs	# relaxed/total IGs	median # relaxed BGs [95% CI]	median IGs [95% CI]	k median BGs [95% CI]	k
<i>Macrotermes natalensis</i> *	foraging	14	53	0.264	12 [6.475-19]	0.72 [0.16-43.10]	0.74 [0.61-0.87]	
<i>Coptotermes</i> sp.	foraging	2	16	0.125	3 [1-6]	0.84 [0.13-31.80]	0.68 [0.44-1.13]	
<i>Reticulitermes santonensis</i>	foraging	1	13	0.077	1 [0-4]	1.35 [0.23-48.15]	0.73 [0.42-1.75]	
<i>Mastotermes darwiniensis</i>	foraging	4	25	0.160	4 [1-7]	0.69 [0.13-26.31]	0.8 [0.63-1.13]	
<i>Prorhinotermes simplex</i>	wood-dwelling	4	47	0.085	6 [2.475-11]	0.76 [0.26-34.35]	0.81 [0.69-1]	
<i>Incisitermes marginipennis</i>	wood-dwelling	0	18	0.000	1 [0-3]	0.81 [0-49.79]	0.92 [0.53-1.28]	
<i>Cryptotermes secundus</i> *	wood-dwelling	15	72	0.208	14 [10-22]	0.73 [0-45.85]	0.77 [0.69-0.89]	
<i>Zootermopsis nevadensis</i> *	wood-dwelling	22	78	0.282	21 [14-27.475]	0.74 [0.33-43.90]	0.73 [0.66-0.81]	

437

Table 2. Number of analyzed IGs with HyPhy-BUSTED and RELAX and respective BG analyses with HyPhy-RELAX for each of the included termite species.

species	life type	# analyzed IGs	# BG ¹ analyses
<i>Macrotermes natalensis</i>	foraging	53	5300
<i>Coptotermes</i> sp.	foraging	16	1700
<i>Reticulitermes santonensis</i>	foraging	13	1300
<i>Mastotermes darwiniensis</i>	foraging	25	2400
<i>Prorhinotermes simplex</i>	wood-dwelling	47	4700
<i>Incisitermes marginipennis</i>	wood-dwelling	18	1800
<i>Cryptotermes secundus</i>	wood-dwelling	72	7300
<i>Zootermopsis nevadensis</i>	wood-dwelling	78	7800

438 We refrained to run BUSTED on the background genes (BGs) because there was no evidence for positive selection on the IGs. Note
439 that when running analyses with RELAX, few BG analyses failed due to numeric instability or convergence errors (from the error
440 logfile from HyPhy RELAX: “relax.K evaluated to a NaN; this can cause all kinds of odd behavior downstream”): one analysis failed
441 for *C. secundus*, one for *P. simplex*, two for *M. natalensis* and three for *Z. nevadensis*.
442

Figure 2. Comparison of selection on immune genes to sets of genes that represent the genomic background.

Relaxed selection on IGs does not differ from the genomic background. Red bar = immune genes (IGs); black dots = 100 sets of background genes (BGs). (A) The ratio of the number of genes under significantly relaxed selection between wood-dwellers and foragers (R_{relax}). (B) The relative intensity of selection (R_k) as measured by the ratio of median k between wood-dwellers and foragers.