

1                    **Adaptive evolution of anti-viral siRNAi genes in bumblebees**

2    Sophie Helbing<sup>1</sup>, H. Michael G. Lattorff<sup>1,2</sup>

3

4        <sup>1</sup> *Institute of Biology, Molecular Ecology, Martin-Luther University Halle-Wittenberg, Hoher*

5    *Weg 4, 06099 Halle (Saale), Germany*

6    <sup>2</sup> *German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig,*

7    *Deutscher Platz 5e, 04103 Leipzig, Germany*

8

9        **Corresponding author:**

10        Sophie Helbing

11        Email: [sophie.helbing@zoologie.uni-halle.de](mailto:sophie.helbing@zoologie.uni-halle.de)

12

13        **Keywords:** social insects, positive selection, parallel evolution, *Bombus*, cuckoo bumblebee

14

15 **Abstract**

16 The high density of frequently interacting and closely related individuals in social insects  
17 enhance pathogen transmission and establishment within colonies. Group-mediated behavior  
18 supporting immune defenses tend to decrease selection acting on immune genes. Along with  
19 low effective population sizes this will result in relaxed constraint and rapid evolution of  
20 genes of the immune system. Here we show that sociality is the main driver of selection in  
21 antiviral siRNAi genes in social bumblebees compared to their socially parasitic cuckoo  
22 bumblebees that lack a worker caste. RNAi genes show frequent positive selection at the  
23 codon level additionally supported by the occurrence of parallel evolution and their  
24 evolutionary rate is linked to their pathway specific position with genes directly interacting  
25 with viruses showing the highest rates of molecular evolution. We suggest that indeed higher  
26 pathogen load in social insects drive adaptive evolution of immune genes, if not compensated  
27 by behavior.

28

## 29 **Introduction**

30 Social insects, like ants, bees, wasps and termites, represent an extremely successful group.  
31 Although they represent only 2 % of all insect (invertebrate) species, they contribute up to  
32 25% of the total insect biomass (Wilson 1990). This success is mainly attributed to the  
33 division of labor amongst workers, but also to the reproductive division of labor between  
34 castes with queens monopolizing reproduction whereas workers remain functionally sterile  
35 (Lattorff and Moritz 2013). This reproductive division of labor is enhanced by the high degree  
36 of relatedness allowing kin selection (Hamilton 1964) to work efficiently.

37 However, this system might represent also some major drawbacks. Social insects are a prime  
38 target for parasites (Schmid-Hempel 1998), as their transmission is enhanced due to frequent  
39 social interactions and their establishment might be augmented by the high relatedness within  
40 colonies as well as by the high degree of nest homeostasis (Boomsma et al. 2005). Despite the  
41 high parasite pressure, a lack of immune system genes has been recognized within all the  
42 completely sequenced social insect genomes when compared to other, non-social insect  
43 species (Evans et al. 2006; Smith et al. 2011). Several explanations have been put forward in  
44 order to explain this discrepancy. The most prominent one explains the lack of immune genes  
45 by an advanced level of socially mediated physiological adaptations and behavioral activities  
46 directed against intruding parasites, summarized as “social immunity” (Cremer et al. 2007).  
47 Indeed, a huge range of activities has been described that altogether might reduce the  
48 selection coefficient acting on genes of the innate immune system. In combination with  
49 drastically lowered effective population sizes in social insects (Romiguier et al. 2014) due to  
50 the reproductive division of labor, this might result in reduced or even absent purifying  
51 selection on immune genes (Evans et al. 2006). This will result in non-efficient removal of  
52 slightly deleterious mutations (also known as relaxed constraint) that will behave like nearly

53 neutral mutations (Ohta 1987) so that non-synonymous substitutions will increase in  
54 frequency and finally contribute to non-synonymous divergence between species. The latter  
55 one has been frequently observed for social insects implying faster evolution of immune  
56 genes (Harpur & Zayed 2013), especially when compared to solitary insects like *Drosophila*  
57 (Viljakainen et al. 2009).

58 Nevertheless, comparisons between highly eusocial ants or the honeybee (*Apis mellifera*) with  
59 solitary insects like *Drosophila* or *Anopheles* suffer from several confounding factors.  
60 Hymenopteran social insects are haplo-diploid (females are diploid whereas males are  
61 haploid), a genetic constitution that by itself reduces the effective population size (Crozier  
62 1977). Furthermore, these lineages might have had a common ancestor 300 million years ago  
63 and might also show lineage specific effects.

64 We study the effect of sociality on the rates of evolution in a superior system consisting of  
65 social bumblebees and their socially parasitic cuckoo bumblebees, the latter one are non-  
66 social due to the lack of a worker caste. The eusocial lineages are characterized by annual  
67 colonies headed by one single-mated queen. The colony cycle is composed of different stages:  
68 (I.) nest establishment; (II.) production of a sterile worker caste ensuring colony growth via  
69 division of labor and (III.) towards the end of the season production of sexual offspring  
70 (Goulson 2010). The cuckoo bumblebee females enter host nests at a time point during early  
71 development of the social nest and kill the resident queen in order to lay their own eggs that  
72 will be cared for by the resident worker force. As this relationship is very specific with every  
73 cuckoo bumblebee species parasitizing a unique host species or at least only a restricted  
74 number of host species, a certain couple of host/parasite species do show similar life history  
75 characters and use a shared environment, at least for the time of shared nests. Cuckoo  
76 bumblebees will show an even more reduced effective population size than their hosts, as they

77 never may parasitize all available nests (Erlor and Lattorff 2010). This might result in even  
78 stronger non-synonymous divergence under relaxed constraint than it might be observed for  
79 social species. In a large comparison of several host/social parasite couples spread throughout  
80 the order of insects, Bromham and Leys (2005) have shown that indeed social parasites show  
81 higher rates of molecular evolution, at least in seven out of eight couples of species.

82 We are studying the rates of molecular evolution in genes of the anti-viral siRNAi pathway  
83 that is responsible for the degradation of double stranded RNA usually derived from viruses.  
84 This system is not inducible, as this pathway also interferes with the regulation of gene  
85 expression of host genes via microRNAs. This system, when acting in different castes like  
86 queens and workers, which might show antagonistic patterns of gene regulation, might  
87 additionally slow down the rates of evolution in these genes. Nevertheless, if parasites, in this  
88 case viruses, which are known to be widespread in social insects, in which interspecific  
89 transmission might also occur, impose a high selective pressure on them, then this might  
90 become visible by higher rates of evolution within the siRNAi genes in those species that are  
91 social.

92

### 93 **Materials and Methods**

94 Six genes of the siRNAi pathway were partially sequenced and compared across six social  
95 *Bombus* species and six of their respective cuckoo bumblebees (*Bombus*, subgenus *Psithyrus*)  
96 with each species represented by a maximum of four haploid drones. Detailed information  
97 concerning the studied species, their geographic origin and method for DNA extraction has  
98 been published elsewhere (Erlor et al. 2014).

99 In order to estimate intra-specific polymorphism 20 and 19 haploid males of *B. terrestris* and  
100 *B. vestalis*, respectively, as well as 5 diploid females of *B. lapidarius* (workers) and *B.*  
101 *rupestris* (females) were sequenced for one of the siRNAi genes (*r2d2*). All those individuals  
102 were unrelated to each other as revealed by microsatellite genotyping (data not shown).

### 103 *Homology searches & Primer design*

104 The RNAi genes were identified in the draft assembly of the *Bombus terrestris* genome  
105 (assembly: Bter 1.0, AELG01000001:AELG01010672) and the *Bombus impatiens* genome  
106 (assembly: Bimp 1.0, AEQM01000001-AEQM01091524) (International Bumblebee Genome  
107 Sequencing Consortium, 2014) using homology searches based on *Drosophila melanogaster*  
108 protein sequences and *Apis mellifera* DNA sequences (Honey Bee Genome, Assembly 4).  
109 Primers were designed on the consensus sequence resulting from the alignment of the *B.*  
110 *terrestris* and *B. impatiens* sequences using Primer3 (Rozen and Skaletsky 2000). Primers  
111 used in this study are reported in table S2.

### 112 *PCR*

113 PCR-protocols were chosen according to the fragment length of the target gene. Fragments  
114 with expected sizes above 2000 bp were amplified with long-range PCR utilizing TaKaRaLA  
115 Taq Polymerase (MoBiTec, Göttingen, Germany). Smaller fragments were amplified using  
116 PeqLab Gold Taq (Peqlab, Erlangen, Germany). PCR protocols are given in table S3.

### 117 *Sequencing*

118 Sequencing of *r2d2* (interspecific variation, 12 species), *vig* and *maelstrom* (*mael*) was  
119 performed by LGC Genomics (Berlin, Germany) using traditional Sanger sequencing.  
120 Samples were sequenced directly and in both directions. Sequence chromatograms were  
121 inspected and confirmed manually. Forward and reverse sequences were assembled using

122 ContigExpress implemented in Vector NTI Advance 10.2.0 (Invitrogen, Karlsruhe, Germany)  
123 in order to detect PCR or sequencing errors.

124 PCR products of *argonaute2* (*ago2*), *armitage* (*armi*) and *dicer 2* (*dcr2*) were sequenced  
125 using the Ion Torrent PGM (Life Technologies). Therefore each species was tagged with a  
126 species-specific barcode adapter and for each gene all individuals per species were pooled in  
127 equimolar amounts. Finally, all amplicons per species were pooled together in equimolar  
128 amounts. The 100 base-read barcoded libraries were prepared using the Ion Xpress Barcode  
129 Adapters 1-16 Kit following the instructions of the Ion Xpress Plus gDNA and Amplicon  
130 Library Preparation user guide. CLC Bio Genomics Workbench 5.0 (CLC Bio, Aarhus,  
131 Denmark) was used to (I) map the reads to the *Bombus terrestris* reference sequences using  
132 default settings and (II) to detect SNP's in the high-throughput sequencing data using the SNP  
133 detection tool. The latter was performed using default settings except for the minimum  
134 coverage and minimum variant frequency which were adjusted to the library (species /gene)  
135 specific requirements.

136 Sequencing of *r2d2* (intraspecific variation, 4 species) was done using the Ion Torrent PGM  
137 (Life Technologies). All amplicons within a species were pooled in equimolar amounts and  
138 species were barcoded using the Ion Xpress Barcode Adapters 1-16 Kit according to the  
139 instructions of the Ion Xpress Plus gDNA and Amplicon Library Preparation user guidelines.  
140 Pooled sequences were mapped to reference sequences for the respective species derived from  
141 Sanger sequencing. SNPs were detected using the Quality-based variant detection function  
142 implemented in CLC Genomics Workbench v7.0.4 (CLC Bio, Aarhus, Denmark) using the  
143 default settings except for the minimum frequency of variants, which was adjusted to the  
144 number of chromosomes and the average coverage.

145 For all species, except *B. perezii*, Cameron et al. 2007 provided sequences of four nuclear  
146 genes (*EF-1 alpha*, *Arginine kinase-AK2*, *rhodopsin*, *PEPCK*) on GenBank. Erler et al. 2014

147 provided *B. perezii* sequences (*EF-1 alpha*, *Arginine kinase-AK2*, *rhodopsin*, *PEPCK*,  
148 GenBank accession numbers: KC662163-67). Based on these genes the phylogenetic  
149 relationship between the focal species was reconstructed and they were also used to estimate  
150 the rate of evolution on non-immune genes.

### 151 *Evolutionary analyses*

152 Multiple sequence alignment was performed using ClustalW implemented in Mega 5.05  
153 (Tamura et al. 2011). Rates of molecular evolution of the RNAi genes were calculated based  
154 on coding sequences separately for each gene.

155 Within species polymorphism as well as the ratio of the number of non-synonymous  
156 substitutions per non-synonymous sites to the number of synonymous substitutions per  
157 synonymous sites including Jukes-Cantor correction (Jukes and Cantor 1969) was estimated  
158 using DnaSP v5 (Librado and Rozas 2009). In order to compare the evolutionary rates  
159 between social and non social cuckoo bumblebees, sequences were furthermore classified  
160 according to their affiliation to either the social or parasite dataset. Stability analyses by  
161 means of jack-knifing over species were performed to confirm the values for each group.

162

### 163 *Site-specific selection*

164 Codeml, implemented in the PAML package (Yang 2007) providing models allowing  $\omega$  to  
165 vary among sites, was used in order to identify sites under positive selection. Therefore, non-  
166 immune gene based phylogeny reconstruction was performed using PHYLIP version 3.69  
167 (Neighbor-Joining method) (Felsenstein 2005). The comparisons comprised models M7 (beta)  
168 vs. M8 (beta &  $\omega$ ) and were repeated twice changing the initial  $\omega$  ( $\omega < 1$ ;  $\omega > 1$ ). Significance  
169 was assessed using a likelihood ratio test (LRT) comparing M7 and M8 models. We used the



170 Bayes empirical Bayes (BEB) approach (codeml implementation) for the inference of site-  
171 specific selection based on calculated posterior probabilities of  $\omega$  classes for each site.

172

### 173 *Parallel evolution*

174 The detection of signs of parallel evolution may also indicate the occurrence of positive  
175 selection. The reconstruction of the ancestral amino acid sequence, based on the phylogeny  
176 and the amino acid sequences of the present-day species enables the inference of the  
177 evolutionary pathway of amino acid substitutions at each site (Yang et al. 1995). Therefore,  
178 nucleotide sequences were translated and resulting amino acid sequences were aligned using  
179 ClustalW (Thompson et al. 1994) implemented in BioEdit (Hall 1999). The ancestral  
180 character reconstruction using the maximum likelihood (ML) algorithm (substitution model:  
181 Jones-Taylor-Thornton model, JTT), as well as the construction of the phylogenetic tree  
182 (Fig.4), based on the non-immune genes, was conducted with Mega 5.05 (Tamura et al.  
183 2011). Hence, it was possible to infer changes that occurred on each lineage and therefore  
184 enabled the identification of shared amino acid replacements. In order to distinguish whether  
185 the observed change is attributable to chance alone or to parallel evolution, the program  
186 CAPE (Zhang and Kumar 1997) was used. CAPE computes the probability that the observed  
187 parallel substitutions are attributable to random chance alone - based on the amino acid  
188 sequences and the divergence between species according to a reference phylogenetic tree  
189 (Zhang and Kumar 1997). The CAPE computations were performed using the JTT  
190 substitution model. The required tree topology (PHYLIP format) as well as the notation of the  
191 nodes was verified using the program package ANCESTOR (Zhang and Nei 1997).

192

193

## 194 **Results**

195 Six RNAi loci were partially amplified: *argonaute2* (*ago2*), *armitage* (*armi*), *dicer2* (*dcr2*),  
196 *maelstrom* (*mael*), *r2d2*, *vasa intronic gene* (*vig*), and coding sequences were compared  
197 across different *Bombus* species. Length of coding sequences analyzed as well as  
198 polymorphism data are reported in table S1. In summary, the total number of within-species  
199 polymorphism was markedly low and affect almost exclusively host species. For *B. t.*  
200 *xanthopus* we observed two synonymous substitutions in the coding region of *vig*;  
201 furthermore one non-synonymous (*ago2*) and one synonymous substitution (*r2d2*) in *B.*  
202 *lucorum*; always one synonymous polymorphism appeared in *B. terrestris* (*mael*) and *B.*  
203 *lapidarius* (*dcr2*) and one non-synonymous substitution in *B. bohemicus* (*r2d2*). In addition,  
204 InDel events appeared in the coding sequences of *armi* and *r2d2* (Table S1). Due to the low  
205 sample size representing each species this study fails to infer patterns of selection also  
206 considering population data. Nevertheless, at least we can draw on intra-specific  
207 polymorphism data for *r2d2* - sequenced in 4 *Bombus* species: *B. terrestris* (drones, N = 20);  
208 *B. vestalis* (drones, N = 19); *B. lapidarius* (workers, N = 5) and *B. rupestris* (workers, N = 5).  
209 These data confirm the observed lack of diversity within coding regions.

210

### 211 *Evolutionary rates of anti-viral siRNAi genes*

212

213 In order to detect footprints of selection in siRNAi genes, the overall ratio of non-  
214 synonymous- to synonymous substitutions was calculated and compared with the  
215 evolutionary rate of non-immune genes (*Arginine kinase*, *EF-1 alpha*, *PEPCK* and  
216 *Rhodopsin*;). The studied siRNAi genes showed an elevated rate of molecular evolution  
217 compared to the non-immune genes ( $\text{mean}_{\text{RNAi}} K_A/K_S = 0.35$ ;  $\text{mean}_{\text{non-imm}} K_A/K_S = 0.02$ ;  
218 Mann-Whitney-U-Test: N = 10, Z = 2.56; p = 0.01). We examined if higher values of this  
219 ratio are driven by non-synonymous changes and indeed,  $K_A$  is 28 times higher in RNAi

220 genes compared to non-immune genes (t-test;  $K_A$ :  $\text{mean}_{\text{RNAi}}=0.028$ ;  $\text{mean}_{\text{non-imm}}=0.001$ ;  $P <$   
221  $0.05$ ), while  $K_S$  is not significantly different (t-test;  $K_S$ :  $\text{mean}_{\text{RNAi}}=0.101$ ;  $\text{mean}_{\text{non-imm}}=0.081$ ;  
222  $P > 0.05$ ). Additionally, non-immune genes showed consistently low  $K_A/K_S$  values (0.01-  
223 0.03) indicating strong purifying selection, while the rate of adaptive evolution in RNAi genes  
224 varied. The strongest signs of adaptive evolution were observed for *r2d2* ( $K_A/K_S = 0.52$ ) and  
225 *ago2* ( $K_A/K_S = 0.65$ ), both encoding proteins directly interacting with viral components and  
226 are therefore expected to be hotspots of adaptive changes. A slightly elevated rate of adaptive  
227 evolution was observed for *dcr2* ( $K_A/K_S = 0.29$ ) and the helicase *armi* ( $K_A/K_S = 0.43$ ). In  
228 contrast, we also identified genes (*vig*, *mael*) showing the occurrence of purifying selection  
229 indicated by low  $K_A/K_S$  values. Especially *vig*, involved in the RNAi defense mechanism but  
230 also in heterochromatin formation (Caudy et al.2002; Gracheva et al. 2009), showed a low  
231  $K_A/K_S$  value compared to the non-immune genes. Besides that it is the smallest of all genes  
232 analyzed reducing the likelihood to pick up a sufficient number of changes and it is located  
233 within the intron of another gene (*vasa*) exposing it to selective pressures acting on this gene.

234

### 235 *Rates of molecular evolution as a function of life-history*

236

237 Reduced selective constraint is expected to increase non-synonymous divergence in socially  
238 parasitic species. Astonishingly, we observed faster evolution for three out of six genes for the  
239 social species, while only two siRNAi genes showed higher rates of molecular evolution in  
240 the non-social parasite species (Fig. 1). Thus, social host species showed an elevated rate of  
241 adaptive evolution relative to the socially parasitic cuckoo bumblebees for *ago2* ( $K_A/K_S$   
242 (Host): 0.91;  $K_A/K_S$  (Parasites): 0.39; T-Test,  $P = 9 \times 10^{-6}$ ), *mael* ( $K_A/K_S$  (Host): 0.34;  $K_A/K_S$   
243 (Parasites): 0.03; T-Test,  $P = 2 \times 10^{-6}$ ) and *armi* ( $K_A/K_S$  (Host): 0.45;  $K_A/K_S$  (Parasites): 0.16;  
244 T-Test,  $P = 1.84 \times 10^{-4}$ ), while for *dcr2* ( $K_A/K_S$  (Host): 0.27;  $K_A/K_S$  (Parasites): 0.74; T-Test,

245  $P = 1.44 \times 10^{-3}$ ) and *vig* ( $K_A/K_S$  (Host): 0.07;  $K_A/K_S$  (Parasites): 0.16; T-Test,  $P = 1 \times 10^{-6}$ )  
246 parasites'  $K_A/K_S$  ratios significantly exceed the values for social species. Comparing the  
247 strength of selection between social hosts and parasites for *r2d2* showed that the evolutionary  
248 rate is not significantly different ( $K_A/K_S$  (Host): 0.54;  $K_A/K_S$  (Parasites): 0.48; T-Test,  $P =$   
249 0.08) indicating high selective pressures supported by the complete lack of within species  
250 polymorphism (see above).

251

252

### 253 *Codon-based positive selection & parallel evolution*

254 None of the studied RNAi genes had (global)  $K_A/K_S > 1$  – indicative for positive selection.  
255 Nevertheless, we found evidence for positively selected sites within the coding region –  
256 especially for *r2d2* and *ago2*. In the ML-based approach we compared model M7 vs. M8 in  
257 order to detect sites experiencing positive selection. For *ago2*, *r2d2*, *dcr2* and *armi* the LRTs  
258 suggests a significant difference between the compared models, indicating the presence of  
259 sites evolving under positive selection. As parallel evolving sites might also serve as a hint for  
260 positive selection, substitution events shared between the social species and their social  
261 parasites were determined. Our data do not support any general pattern of parallel evolution,  
262 neither between cuckoo bumblebees and their respective hosts nor with respect to  
263 geographical origin. Nevertheless, the occurrence of parallel substitutions was observed for  
264 each locus, except for *vig* (Table 1, Table 2).

265 The comparison between models M7 and M8 for *ago2* yielded a LRT statistic of 16.76,  
266 significant at  $p = 0.001$ , with a proportion of sites (6.3%) with  $\omega = 7.08$  (Table S4). The ML  
267 analysis for *ago2* revealed three positively selected sites with  $P > 95\%$  (BEB: sites 41 and 68  
268 with  $P > 95\%$ ; site 77 with  $P > 99\%$ ). All these sites are located in the nucleic acid binding  
269 domain PAZ (Fig. 2). Furthermore, five parallel changed sites were detected, always in pairs

270 including *B. ruderatus*. Interestingly, two identical substitutions (sites 67 and 186) were  
271 shared between *B. ruderatus* and most of the social parasites except for *B. campestris* and *B.*  
272 *rupestris*, with one of them (site 67) showing an alteration of the chemical property of the  
273 amino acid (Table 1).

274 The comparison of M7 vs. M8 for *r2d2* yielded a test statistic of 9.86, significant at  $p = 0.01$   
275 (Table S4). Here, M8 identified 30 positive selected sites ( $P > 50\%$ ) – three of them with  $P >$   
276 95% (BEB: site 87, prob.  $> 99\%$ ; sites 90 and 181, prob.  $> 95\%$ ; Fig. 3). Seven identical  
277 changes were observed, but only four of them passed the test for parallel evolution. One  
278 charge-altering substitution, which takes place in the *dsrcm* domain, occurred in *B. t.*  
279 *xanthopus* and *B. pascuorum* (T54I,  $\Phi = 0.014$ ). *B. terrestris*, *B. t. xanthopus* and *B.*  
280 *pascuorum* shared the same charge altering amino acid replacement at position 103 (Y103H),  
281 unaccountable to chance alone, ( $\Phi = 0.002$ ) and thus strongly suggesting the occurrence of  
282 parallel evolution. Furthermore, there are two parallel-change sites (T87N; N161S;  $\Phi = 0.000$ )  
283 in the *B. rupestris*- *B. lucorum*/*B. terrestris*/*B. t. xanthopus* comparison. However, it was not  
284 possible to infer the ancestral state at these positions.

285 For *dcr2* (M7 vs. M8: LRT= 15.38, significant at  $p = 0.001$ ; Table S4) the BEB identified two  
286 sites (79, 82) to be positively selected with a probability of  $> 99\%$ . Both sites are  
287 exceptionally variable and we observed potential parallel substitutions. Unfortunately, it was  
288 not possible to infer the ancestral state for both sites. Nevertheless, we calculated the  
289 probability for observing two identical substitutions shared between *B. ruderatus* and *B.*  
290 *rupestris* – as both species even showed identical substitutions at both sites. Testing for  
291 parallel evolution revealed that these changes are highly significant exceeding random chance  
292 expectation.

293

294 **Discussion**

295 Life-history traits associated with sociality, infection risk within groups, are predicted to  
296 affect the rates of molecular evolution. Several studies (Viljakainen et al. 2009; Bulmer and  
297 Crozier 2004, 2006; Harpur and Zayed 2013; Erler et al. 2014) targeted at quantifying host  
298 parasite conflict-effects on the evolution of genes involved in antimicrobial or antifungal  
299 defense mechanisms in social insects. However, antiviral defense mechanisms (e.g. RNAi) are  
300 also part of the innate immune system leading to the inhibition of expression of viral genes  
301 essential for viral replication. Counteracting this, viruses encode essential virulence factors,  
302 called viral suppressors of RNAi (VSR) interfering with key-steps of host immune responses  
303 (Li and Ding 2006). Virus-associated characteristics (high replication and mutation rates) as  
304 well as VSR impose severe requirements on the host immune-system resulting in an  
305 evolutionary arms-race between components of the RNAi-pathway and viruses (Moissiard et  
306 al. 2004; Li et al. 2005; Obbard et al. 2006; Obbard et al. 2009a). For *Drosophila*, high rates  
307 of adaptive evolution of siRNAi genes are indicative for these co-evolutionary interactions  
308 (Obbard et al. 2006; Obbard et al. 2009b, Kolaczkowski et al. 2011).

309 We studied the evolutionary pattern of antiviral RNAi genes across the genus *Bombus*, as it  
310 comprises social- as well as non-social lineages (cuckoo bumblebees). A global comparison  
311 of  $K_A/K_S$  values between siRNAi and non-immune genes revealed that the former evolve  
312 significantly faster indicating permanent exposure to selection pressure imposed by viruses,  
313 especially as insect infecting (RNA) viruses are emerging as serious threats (Singh et al. 2010;  
314 Levitt et al. 2013; Fürst et al. 2014). Antiviral RNAi mechanisms are activated in the presence  
315 of viral double-stranded (ds)-RNA, leading to the generation of small interfering RNAs  
316 (siRNAs) serving as template for RISC-mediated viral mRNA degradation (Meister and  
317 Tuschl 2004; Buchon and Vaury 2006; Ding 2010; Nayak et al. 2013). The studied siRNAi  
318 genes showed huge variance in their evolutionary rates, possibly reflecting their different  
319 functions. In that regard we identified genes potentially under purifying selection, especially

320 *vig* and *mael*, while others showed elevated rates of evolution, even though none of them with  
321  $K_A/K_S > 1$  for the whole gene. The highest rates of adaptive evolution were detected for *r2d2*  
322 and *ago2*. *R2d2* functions in siRNA production, while *ago2* (core element of the RISC)  
323 mediates viral mRNA degradation. Thus, both gene products are required for an effective  
324 RNAi and represent potential targets for VSR (Nayak et al. 2010; van Mierlo et al. 2012).  
325 The overall trend for the rate of adaptive evolution is largely consistent with results reported  
326 for *Drosophila* (Obbard et al. 2009). Nevertheless, *mael* was assigned to five per cent of the  
327 fastest evolving genes, while the rate of adaptive evolution in bumblebees is just slightly  
328 elevated compared to non-immune genes.

329 Although, we could not detect signals of positive selection for full genes, we found candidates  
330 with positively selected sites. For *r2d2* we identified 16 positively selected sites (BEB;  $P >$   
331 75%), interestingly only three fall within the *DSRM* domain. This is consistent with the  
332 outcome of a study by Kolaczkowski et al. (2011), who found a large proportion of adaptive  
333 changes in the region separating both *dsrm* domains, while just a few adaptive protein-coding  
334 changes in the N-terminal *dsrm* domain were detected. The relative abundance of adaptive  
335 changes in the *dsrm* domain might be driven by the removal of deleterious mutations, since  
336 point mutations within this domain abolished dsRNA binding (Liu et al. 2003). By contrast,  
337 the *ago2* PAZ domain showed an elevated proportion of positively selected sites. This leads to  
338 the suggestion that this rapidly changing domain might be a target of parasite-immune evasion  
339 strategies, especially considering that *ago2* is a target of viral suppressors of RNAi (VSR) as  
340 shown for *Drosophila* (Nayak et al. 2010; Barribeau et al. 2014). Nevertheless, sites under  
341 positive selection, identified in a comparison of highly (*Apis mellifera*, *A. florea*) and  
342 primitively eusocial (*B. terrestris*, *B. impatiens*) as well as a non-social (*Megachile rotundata*)  
343 species, either on the branch leading to sociality (social vs. non-social species) or on the  
344 branch to bumblebees (*Bombus* vs. *Apis*) (Barribeau et al. 2014) were not recovered as being

345 positively selected in our study suggesting that these sites have been spread to fixation before  
346 the diversification within the genus *Bombus*.

347 Based on stability analyses, one species – *B. rupestris* - was identified to have a drastic impact  
348 on the estimated values of the social parasite species for *ago2* and *dcr2*: Excluding *B.*  
349 *rupestris* in case of *ago2* will lead to an increase of the parasite diversity, while in case of  
350 *dcr2* this diversity is reduced.

351 Furthermore, five out of six genes showed evidence for parallel changed substitutions across  
352 the genus *Bombus* - also indicative for positive selection. Parallel substitutions were observed  
353 within host species belonging to different clades as well as between social parasites and  
354 clades of the social host. In some cases shared substitutions affect multiple sites in a gene; for  
355 instance we observed two substitutions in close proximity in *dcr2* shared between *B. rupestris*  
356 and *B. ruderatus* (Table 2). We hypothesized positive selection might favor same sites in  
357 cuckoo bumblebees and their respective hosts, as they show similarities in their life-cycles,  
358 share the same environment and are therefore predicted to be exposed to similar pathogens.  
359 We also expected to identify parallel changed sites for species originating from the same  
360 sampling site, reflecting geographically-restricted selective pressure due to the occurrence of  
361 location-specific viruses. However, no evidence was found for parallel substitution patterns  
362 reflecting either the relationship between host- social parasite couples, or location-restricted  
363 adaptation.

364 Sociality is expected to influence the molecular evolution of genes in a number of ways:  
365 mainly through reduction in  $N_e$  due to reproductive monopolization, but also continuous  
366 oogenesis and/or caste biased gene expression. Utilizing a comparative approach  
367 encompassing social and non-social lineages throughout the order of insects (including  
368 host/social parasite couples), Bromham and Leys (2005) investigated sociality-associated  
369 effects on the rate of molecular evolution. Here, evidence for an effect of lower effective



370 population size is rising, as eusocial Hymenoptera showed increased evolutionary rates  
371 compared to their nonsocial relatives, but decreased rates compared to social parasites  
372 experiencing a further reduction in  $N_e$  (Bromham and Leys 2005). However, for the studied  
373 anti-viral siRNAi genes the impact of low  $N_e$  is not that conclusive: for three out of six genes  
374 the rate of molecular evolution is considerably faster in social species, even though assumed  
375 to have a higher  $N_e$  (Erlor and Lattorff 2010). Hence, this implies not primarily an effect of  $N_e$   
376 on substitution rates, but also other sociality-related features may impact the observed rates of  
377 molecular evolution of anti-viral siRNAi genes. Especially the high parasite pressure  
378 associated with group-living of highly related individuals, are expected to overpower the  
379 effect of  $N_e$  in this study, as the antiviral RNAi pathway is part of the innate immune system,  
380 where selection is expected to be particularly strong. The loss of immune genes observed for  
381 honeybees compared to *Drosophila*, is assumed to be due to alternative defense strategies  
382 such as socially mediated physiological adaptations and behavioral activities. Many of these  
383 social defense mechanisms, e.g. removal of infected individuals, collection and secretion of  
384 antimicrobial substances, serve to avoid bacterial or fungal infection (Cremer et al. 2007).  
385 Hence, features associated with social immunity may particularly reduce the selection  
386 coefficient acting on genes involved in antibacterial or antifungal defense mechanisms, but  
387 not on antiviral genes. With regard to their short “generation times” and high mutation rates  
388 viruses *per se* pose a particular challenge to host immune system. And, as honeybee viruses  
389 have been shown to infect bumblebees, there is increasing evidence for cross-species  
390 transmission and establishment (Genersch et al. 2006; Singh et al. 2010; Levitt et al. 2013;  
391 Fürst et al. 2014), which likewise might reinforce a selective pressure. For honeybees  
392 silencing of viral RNA after ingestion of viral dsRNA leads to a reduction in viral load and  
393 bee mortality (Maori et al. 2009; Desai et al. 2012) - indicating the functional importance of  
394 an effective RNAi pathway in managing viral infections. Moreover, within-species

395 comparisons revealed only scarce evidence for polymorphisms, most of them being  
396 synonymous (Table S1), indicating rather selection than relaxed selective constraint may  
397 impact the molecular evolution of the anti-viral siRNAi genes.

398 Also reproductive features might contribute to the different substitution rates. In social  
399 Hymenoptera, the delayed production of sexuals is assumed to increase the chance for  
400 inheriting mutations due to accumulation of DNA copy errors, as oogenesis in Hymenoptera  
401 is continuous, which is assumed to increase mutation rates as the number of germline cell  
402 divisions increases with generation (Büning 1994; Crow 1997; Bromham and Leys 2005). In  
403 contrast, socially parasitic cuckoo bumblebees lack a worker caste and therefore, the female  
404 immediately starts with the production of reproductive offspring. Thus, the effect of  
405 continuous oogenesis on the mutation rate might not be that strong.

406

#### 407 *Conclusion*

408 High infection risk is one of the main challenges social insects have to counteract. In that  
409 regard their downscaled immune gene repertoire appears to be counterintuitive. Here,  
410 behavioral defense mechanisms might provide valuable support reducing infection risk and in  
411 that regard selective pressure on immune genes.

412 For antiviral defense mechanisms, where selection pressure is exceptionally strong and social  
413 immunity is unlikely to work efficiently, sociality associated features (high infection risk;  
414 centralized reproductive privilege causing low  $N_e$ ; consequences of continuous oogenesis and  
415 delayed sexuals production on mutation rate) seems to affect the rate of molecular evolution.  
416 Evidence is given as for three out of six siRNAi genes higher evolutionary rates were  
417 observed for the social species compared to their non-social parasitic cuckoo bumblebees. The

418 latter ones are assumed to show higher rates due to a relaxed selective constraint as a  
419 consequence of a stronger reduction in  $N_e$ .

420

## 421 **Acknowledgements**

422 We are grateful to D. Kleber and P. Leibe for assistance with lab work, especially for NGS  
423 sequencing. We would like to thank S. Erler for providing genomic DNA of bumblebee  
424 species, which were collected by P. Lhomme and P. Rasmont. We also thank B. Fouks, P.  
425 Thoisy, B. Quaye and A. Weiss for providing sequences of intraspecific r2d2 variability. We  
426 thank the Bumblebee Genome Consortium (<http://hymenopteragenome.org/beebase/>) for  
427 providing genomic resources that were used for this study. This work was supported by a  
428 grant from the German Ministry for Education and Research (BMBF) within the program  
429 FUGATO-Plus (FKZ: 0315126 to HMGL).

430 RNAi gene sequences for all species were deposited in GenBank (accession numbers:)

431

## 432 **References**

- 433 Barribeau, SM. et al. A depauperate immune repertoire precedes evolution of sociality in  
434 bees. *Genome Biol* (in press) (2014).
- 435 Boomsma JJ, Schmid-Hempel P, Hughes WOH. 2005. Life Histories and Parasite Pressure  
436 Across the Major Groups of Social Insects. In Fellowes M, Holloway G, Rolff J,  
437 editors, *Insect Evolutionary Ecology*. Wallingford Oxon, UK & Cambridge MA, USA:  
438 CABI Publishing. p. 139-175.
- 439 Bromham L, Leys R. 2005. Sociality and the rate of molecular evolution. *Mol. Biol. Evol.* 22:  
440 1393–1402. doi: 10.1093/molbev/msi133

- 441 Buchon N, Vaury C. 2006. RNAi: a defensive RNA-silencing against viruses and  
442 transposable elements. *Heredity* 96: 195–202. doi:10.1038/sj.hdy.6800789
- 443 Bulmer MS, Crozier RH. 2004. Duplication and diversifying selection among termite  
444 antifungal peptides. *Mol. Biol. Evol.* 21: 2256-2264. doi: 10.1093/molbev/msh236
- 445 Bulmer MS, Crozier RH. 2006. Variation in positive selection in termite GNBPs and Relish.  
446 *Mol. Biol. Evol.* 23: 317-326. doi: 10.1093/molbev/msj037
- 447 Büning J. 1994. *The Insect Ovary: Ultrastructure, Previtellogenic Growth and Evolution.*  
448 Springer.
- 449 Cameron SA, Hines HM, Williams PH. 2007. A comprehensive phylogeny of the bumble  
450 bees (*Bombus*). *Biol. J. Linn. Soc.* 91: 161–188. doi: 10.1111/j.1095-  
451 8312.2007.00784.x
- 452 Caudy AA, Myers M, Hannon GJ, Hammond SM. 2002. Fragile X-related protein and VIG  
453 associate with the RNA interference machinery. *Genes Dev.* 16: 2491–2496.  
454 doi: 10.1101/gad.1025202
- 455 Cremer S, Armitage SA, Schmid-Hempel P. 2007. Social immunity. *Curr Biol.* 17: 693-702.  
456 doi:10.1016/j.cub.2007.06.008
- 457 Crow JF. 1997. The high spontaneous mutation rate: is it a health risk? *Proc. Natl. Acad. Sci.*  
458 USA. 94: 8380-8386.
- 459 Crozier RH. 1977. Evolutionary genetics of the Hymenoptera. *AnnuRev Entomol.* 22:263-  
460 288. doi: 10.1146/annurev.en.22.010177.001403

- 461 Desai S D, Eu Y-J, Whyard S, Currie RW. 2012. Reduction in deformed wing virus infection  
462 in larval and adult honey bees (*Apis mellifera* L.) by double-stranded RNA ingestion.  
463 *Insect Mol. Biol.* 21: 446–455. doi: 10.1111/j.1365-2583.2012.01150.x.
- 464 Ding S-W. 2010. RNA-based antiviral immunity. *Nat. Rev. Immunol.* 10: 632–644. doi:  
465 10.1038/nri2824.
- 466 Erler S, Lattorff HMG. 2010. The degree of parasitism of the bumblebee (*Bombus terrestris*)  
467 by cuckoo bumblebees (*Bombus (Psithyrus) vestalis*). *Insectes Sociaux* 57: 371–377.  
468 doi: 10.1007/s00040-010-0093-2.
- 469 Erler S, Lhomme P, Rasmont P, Lattorff HMG. 2014. Rapid evolution of antimicrobial  
470 peptide genes in an insect host-social parasite system. *Infect. Genet. Evol.* 23: 129–  
471 137. doi: 10.1016/j.meegid.2014.02.002.
- 472 Evans JD, Aronstein K, Chen YP, Hetru C, Imler J-L, Jiang H, Kanost M, Thompson GJ, Zou  
473 Z, Hultmark D. 2006. Immune pathways and defence mechanisms in honey bees *Apis*  
474 *mellifera*. *Insect Mol. Biol.* 15: 645–656. doi: 10.1111/j.1365-2583.2006.00682.x.
- 475 Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the  
476 author. Department of Genome Sciences, University of Washington, Seattle.
- 477 Fürst MA, McMahon DP, Osborne JL, Paxton RJ, Brown MJF. 2014. Disease associations  
478 between honeybees and bumblebees as a threat to wild pollinators. *Nature* 506: 364-  
479 366. doi: 10.1038/nature12977.
- 480 Genersch E, Yue C, Fries I, de Miranda JR. 2006. Detection of Deformed wing virus, a honey  
481 bee viral pathogen, in bumble bees (*Bombus terrestris* and *Bombus pascuorum*) with  
482 wing deformities. *J. Invertebr. Pathol.* 91: 61–63. doi:10.1016/j.jip.2005.10.002.

- 483 Goulson D. 2010. *Bumblebees: Behaviour, Ecology and Conservation*. Oxford University  
484 Press.
- 485 Gracheva E, Dus M, Elgin SCR. 2009. Drosophila RISC component VIG and its homolog  
486 Vig2 impact heterochromatin formation. *PLoS ONE* 4: e6182. DOI:  
487 10.1371/journal.pone.0006182.
- 488 Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis  
489 program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41: 95-98.
- 490 Hamilton WD. 1964. Genetical evolution of social behaviour. *J. Theor. Biol.* 7: 1-52.
- 491 Harpur BA, Zayed A. 2013. Accelerated evolution of innate immunity proteins in social  
492 insects: adaptive evolution or relaxed constraint? *Mol. Biol. Evol.* 30: 1665–1674.  
493 doi: 10.1093/molbev/mst061.
- 494 Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editors.  
495 *Mammalian Protein Metabolism*. New York: Academic Press. p. 21-132.
- 496 Kolaczowski B, Hupalo D N, Kern AD. 2011. Recurrent adaptation in RNA interference  
497 genes across the Drosophila phylogeny. *Mol. Biol. Evol.* 28: 1033–1042.  
498 doi: 10.1093/molbev/msq284.
- 499 Lattorff HMG, Moritz RFA. 2013. Genetic underpinnings of division of labor in the honeybee  
500 (*Apis mellifera*). *Trends Genet.* 29: 641-648. doi: 10.1016/j.tig.2013.08.002.
- 501 Levitt AL, Singh R, Cox-Foster DL, Rajotte E, Hoover K, Ostiguy N, Holmes EC. 2013.  
502 Cross-species transmission of honey bee viruses in associated arthropods. *Virus Res.*  
503 176: 232–240. doi: 10.1016/j.virusres.2013.06.013.

- 504 Li H-W, Ding S-W. 2005. Antiviral silencing in animals. *FEBS Lett.* 579: 5965–5973.  
505 doi: 10.1016/j.febslet.2005.08.034.
- 506 Li F, Ding S-W. 2006. Virus counterdefense: diverse strategies for evading the RNA-  
507 silencing immunity. *Annu. Rev. Microbiol.* 60: 503–531.  
508 doi: 10.1146/annurev.micro.60.080805.142205.
- 509 Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA  
510 polymorphism data. *Bioinformatics* 25: 1451-1452. doi:  
511 10.1093/bioinformatics/btp187.
- 512 Liu Q, Rand TA, Kalidas S, Du F, Kim HE, Smith DP, Wang X. 2003. R2D2, a bridge  
513 between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science*  
514 301: 1921–1925. doi: 10.1126/science.1088710.
- 515 Maori E, Paldi N, Shafir S, Kalev H, Tsur E, Glick E, Sela I. 2009. IAPV, a bee-affecting  
516 virus associated with Colony Collapse Disorder can be silenced by dsRNA ingestion.  
517 *Insect Mol. Biol.* 18: 55–60. doi: 10.1111/j.1365-2583.2009.00847.x.
- 518 Meister G, Tuschl T. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature*  
519 431: 343–349. doi:10.1038/nature02873.
- 520 Moissiard G, Voinnet O. 2004. Viral suppression of RNA silencing in plants. *Mol. Plant*  
521 *Pathol.* 5: 71–82. DOI: 10.1111/j.1364-3703.2004.00207.x.
- 522 Nayak A, Berry B, Tassetto M, Kunitomi M, Acevedo A, Deng C, Krutchinsky A, Gross J,  
523 Antoniewski C, Andino R. 2010. Cricket paralysis virus antagonizes Argonaute 2 to  
524 modulate antiviral defense in *Drosophila*. *Nat. Struct. Mol. Biol.* 17: 547–554. doi:  
525 10.1038/nsmb.1810.

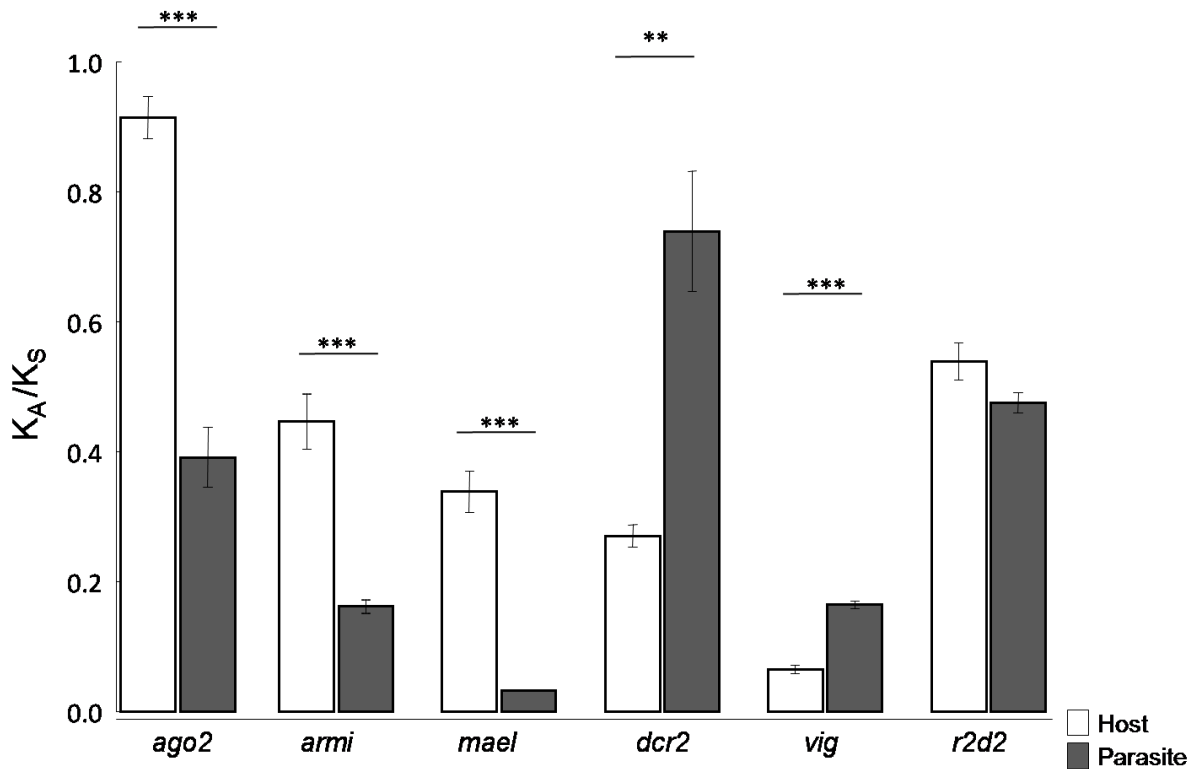
- 526 Nayak A, Tassetto M, Kunitomi M, Andino R. 2013. RNA interference-mediated intrinsic  
527 antiviral immunity in invertebrates. In: Cullen BR, editor. *Intrinsic Immunity*. Current  
528 topics in Microbiology and Immunology. 371: 183-200.
- 529 Obbard DJ, Jiggins FM, Halligan DL, Little TJ. 2006. Natural selection drives extremely  
530 rapid evolution in antiviral RNAi genes. *Curr. Biol.* 16: 580–585.  
531 doi:10.1016/j.cub.2006.01.065.
- 532 Obbard DJ, Gordon K H J, Buck A H, Jiggins FM. 2009a. The evolution of RNAi as a  
533 defence against viruses and transposable elements. *Philos. Trans. R. Soc. Lond. B.*  
534 *Biol. Sci.* 364: 99–115. doi: 10.1098/rstb.2008.0168.
- 535 Obbard DJ, Welch JJ, Kim KW, Jiggins FM. 2009b. Quantifying adaptive evolution in the  
536 *Drosophila* immune system. *PLoS Genetics* 5: e1000698. doi:  
537 10.1371/journal.pgen.1000698
- 538 Ohta T. 1987. Very slightly deleterious mutations and the molecular clock. *J. Mol. Evol.* 26:  
539 1-6.
- 540 Romiguier J, Lourenco J, Gayral P, Faivre N, Weinert LA, Ravel S, Ballenghien M, Cahais V,  
541 Bernard A, Loire E, et al. 2014. Population genomics of eusocial insects: the costs of a  
542 vertebrate-like effective population size. *J. Evol. Biol.* 27: 593–603 (2014). doi:  
543 10.1111/jeb.12331
- 544 Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist  
545 programmers. *Methods Mol Biol.* 132: 365-386.
- 546 Schmid-Hempel P. 1998. *Parasites in Social Insects*. Princeton: Princeton University Press.



- 547 Singh R, Levitt AL, Rajotte EG, Holmes EC, Ostiguy N, vanEngelsdorp D, Lipkin WI,  
548 Depamphilis CW, Toth AL, Cox-Foster DL. 2010. RNA viruses in hymenopteran  
549 pollinators: evidence of inter-Taxa virus transmission via pollen and potential impact  
550 on non-*Apis* hymenopteran species. *PLoS One* 5, e14357. doi:  
551 10.1371/journal.pone.0014357.
- 552 Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, Holt C, Hu H,  
553 Abouheif E, Benton R, et al. 2011. Draft genome of the red harvester ant  
554 *Pogonomyrmex barbatus*. *Proc. Natl. Acad. Sci. USA* 108: 5667–5672. doi:  
555 10.1073/pnas.1007901108.
- 556 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular  
557 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and  
558 maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739. doi:  
559 10.1093/molbev/msr121.
- 560 Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of  
561 progressive multiple sequence alignment through sequence weighting, position  
562 specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.
- 563 van Mierlo JT, Bronkhorst AW, Overheul GJ, Sadanandan SA, Ekström JO, Heestermans M,  
564 Hultmark D, Antoniewski C, van Rij RP. 2012. Convergent evolution of argonaute-2  
565 slicer antagonism in two distinct insect RNA viruses. *PLoS Pathog.* 8: e1002872. doi:  
566 10.1371/journal.ppat.1002872.
- 567 Viljakainen L, Evans JD, Hasselmann M, Rueppell O, Tingek S, Pamilo P. 2009. Rapid  
568 evolution of immune proteins in social insects. *Mol. Biol. Evol.* 26: 1791–1801. doi:  
569 10.1093/molbev/msp086.

- 570 Wilson EO. 1990. Success and dominance in ecosystems: the case of the social insects.  
571 Oldendorf/Luhe.
- 572 Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum  
573 likelihood. *Mol. Biol. Evol.* 24: 1586-1591. doi: 10.1093/molbev/msm088.
- 574 Yang Z, Kumar S, Nei M. 1995. A new Method of inference of ancestral nucleotide and  
575 amino acid sequences. *Genetics* 141: 1641-1650.
- 576 Zhang J, Kumar S. 1997. Detection of convergent and parallel evolution at the amino acid  
577 sequence level. *Mol. Biol. Evol.* 14: 527-536.
- 578 Zhang J, Nei, M. 1997. Accuracies of ancestral amino acid sequences inferred by the  
579 parsimony, likelihood, and distance methods. *J. Mol. Evol.* 44: 139-146.
- 580

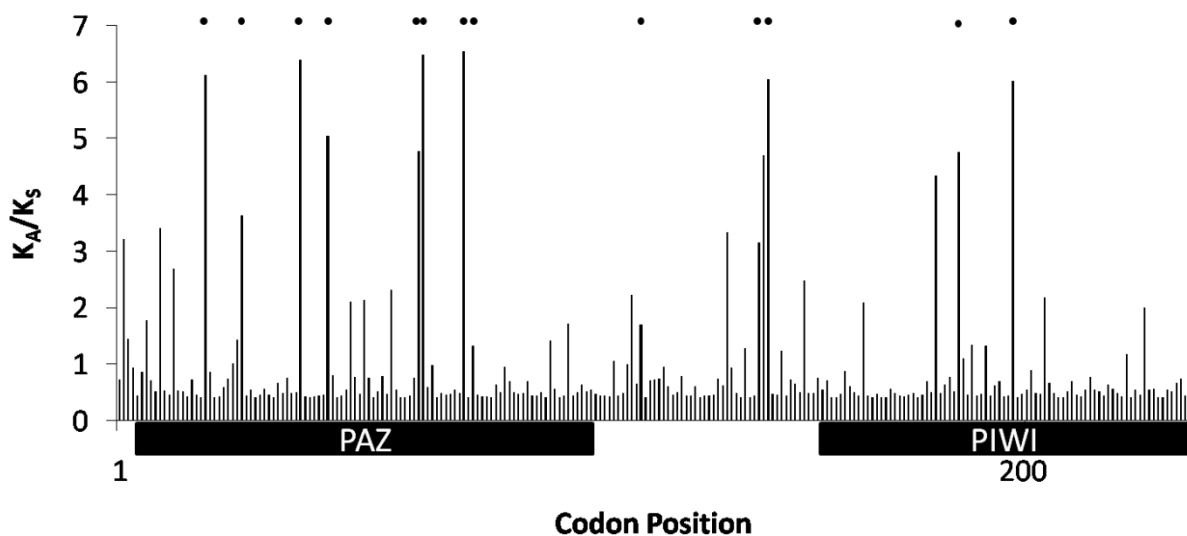
581 **Figures**



582

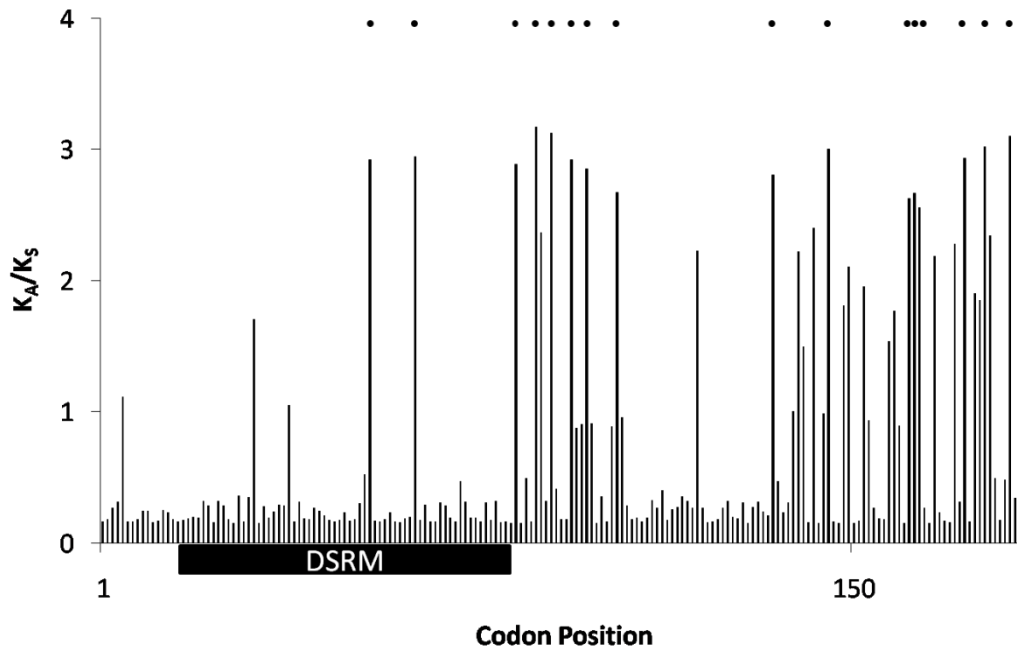
583 **Figure 1. Evolutionary rate of siRNAi genes for social and socially parasitic cuckoo**  
 584 **bumblebees.** Ka/Ks values by means of jack-knifing over species for the social (open bars) and non-  
 585 social (filled bars) datasets. Statistically (T-Test) supported differences: \*\* P-value<0.01; \*\*\* P-value  
 586 <0.0001.

587



588

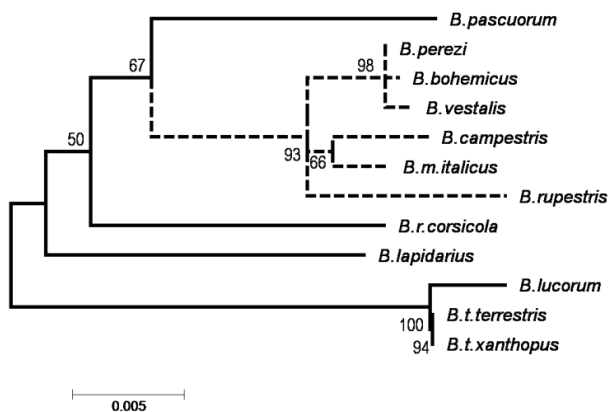
589 **Figure 2. Evidence for positively selected sites in ago2.** Posterior mean  $\omega$  for each codon position  
 590 of *ago2* under model M8. Symbols refer to positively selected sites ( $\omega > 1$ ) with posterior probabilities  
 591 under model M8 (BEB method):  $P > 0.75$  (BEB; sites: 20, 47, 144, 198);  $P > 0.95$  (BEB; sites: 41, 68,  
 592 77) and parallel changed sites: 28<sup>a</sup>, 41, 67, 79<sup>a</sup>, 116<sup>a</sup>, 142, 186. <sup>a</sup> - ancestor could not be inferred. Bars  
 593 below the plot indicate conserved domains.



594

595 **Figure 3. Evidence for positively selected sites in *r2d2*.** Posterior mean  $\omega$  for each codon position  
 596 of *r2d2* (model M8). Symbols refer to positively selected sites ( $\omega > 1$ ) with posterior probabilities under  
 597 model M8 (BEB method):  $P > 0.75$  (BEB; sites: 54, 63, 83, 94, 97, 103, 134, 145, 161, 162, 163, 172,  
 598 176);  $P > 0.95$  (BEB; sites: 87, 90, 181) and parallel changed sites: 54, 87<sup>a</sup>, 103, 119<sup>a</sup>, 134<sup>a</sup>, 149<sup>a</sup>,  
 599 161<sup>a</sup>; <sup>a</sup> - ancestor could not be inferred. Bars below the plot indicate conserved domains.

600



601

602 **Figure 4. *Bombus* phylogenetics.** Phylogenetic relationships of the six host (solid line) and social  
 603 parasite (dashed line) bumblebee species based on the coding sequences of non-immune genes  
 604 (Arginine kinase, EF-1 alpha, PEPCK, Rhodopsin) inferred using the Maximum Likelihood method  
 605 based on the Kimura 2-parameter model method (including test of phylogeny: bootstrap with 500  
 606 replicates).

607

	<b>ago2</b>								<b>armi</b>			
<b>Position</b>	28 <sup>a</sup>	41	47	67	79 <sup>a</sup>	116 <sup>a</sup>	142	186	88	257	648	773
<b>Ancestral</b>		A	V	T			M	R	S	Q	A	S
<i>B.luc</i> (H)	<b>K</b>	<b>S</b>	.	.	K	K	<b>L</b>	.	.	<b>H</b>	.	<b>N</b>
<i>B.tx</i> (H)	<b>K</b>	.	.	.	K	K	<b>L</b>	.	NA			
<i>B.ter</i> (H)	<b>K</b>	.	.	.	K	K	<b>L</b>	.	.	<b>H</b>	<b>T</b>	<b>N</b>
<i>B.lap</i> (H)	R	.	.	.	<b>R</b>	<b>R</b>	.	.	.	.	.	.
<i>B.pas</i> (H)	NA								.	<b>H</b>	.	.
<i>B.max</i> (P)	R	.	.	<b>I</b>	K	K	.	<b>K</b>	NA			
<i>B.bo</i> (P)	R	.	.	<b>I</b>	K	K	.	<b>K</b>	.	.	<b>T</b>	.
<i>B.vest</i> (P)	R	.	.	<b>I</b>	K	K	.	<b>K</b>	.	.	<b>T</b>	.
<i>B.per</i> (P)	R	.	.	<b>I</b>	K	K	.	<b>K</b>	.	.	<b>T</b>	.
<i>B.cam</i> (P)	R	T	<b>A</b>	.	K	K	.	.	<b>G</b>	.	.	.
<i>B.rup</i> (P)	R	.	.	.	K	K	.	.	.	.	.	<b>N</b>
<i>B.rud</i> (H)	<b>K</b>	<b>S</b>	<b>A</b>	<b>I</b>	<b>R</b>	<b>R</b>	<b>L</b>	<b>K</b>	<b>G</b>	.	.	.
$\phi$	*	*	n.s.	*	*	*	*	*	*	n.s.	*	*
	1			2	3	3	1	2				

H - Host, P - social Parasite; NA – not available; n.s. – not significant; <sup>a</sup> - ancestor could not be inferred;  $\phi$  – chance probability of having *n* parallel amino acid substitutions; last row indicate paired analysis (assigned numbers) for multiple shared substitutions; *B.luc* – *B. lucorum*, *B.tx* – *B.t. xanthopus*, *B.ter* – *B. t. terrestris*, *B.lap* – *B. lapidarius*, *B.pas* – *B. pascuorum*, *B.max* – *B. m. italicus*, *B.bo* – *B. bohemicus*, *B.vest* – *B. vestalis*, *B.per* - *B. perezi*, *B.cam* – *B.campestris*, *B.rup* – *B.rupestris*, *B.rud* – *B. r. corsicola*

**Table 2: Parallel changed amino acid sites for *dcr2*, *mael* and *r2d2*.**

Position	<i>dcr2</i>			<i>mael</i>	<i>r2d2</i>						
	56	79 <sup>a</sup>	82 <sup>a</sup>	29	54	103	87 <sup>a</sup>	119 <sup>a</sup>	134 <sup>a</sup>	149 <sup>a</sup>	161 <sup>a</sup>
<b>Ancestral</b>	R			N	T	Y					
<i>B.luc</i> (H)	<b>K</b>	N	H	.	.	.	<b>N</b>	T	L	<b>T</b>	<b>S</b>
<i>B.tx</i> (H)	<b>K</b>	N	H	.	<b>I</b>	<b>H</b>	<b>N</b>	T	L	<b>T</b>	<b>S</b>
<i>B.ter</i> (H)	NA			.	.	<b>H</b>	<b>N</b>	T	L	<b>T</b>	<b>S</b>
<i>B.lap</i> (H)	.	N	L	<b>D</b>	<b>S</b>	.	<b>S</b>	T	I	<b>T</b>	N
<i>B.pas</i> (H)	<b>K</b>	Y	L	<b>S</b>	<b>I</b>	<b>H</b>	<b>G</b>	<b>I</b>	L	I	N
<i>B.max</i> (P)	.	Y	F	.	.	.	T	T	I	I	N
<i>B.bo</i> (P)	.	Y	F	.	.	.	T	T	I	I	N
<i>B.vest</i> (P)	.	Y	F	.	.	.	T	T	I	I	N
<i>B.per</i> (P)	.	Y	F	.	.	.	T	T	I	I	N
<i>B.cam</i> (P)	.	Y	F	<b>D</b>	.	.	T	T	I	I	N
<i>B.rup</i> (P)	.	<b>C</b>	<b>S</b>	<b>D</b>	.	.	<b>N</b>	T	I	I	<b>S</b>
<i>B.rud</i> (H)	.	<b>C</b>	<b>S</b>	<b>S</b>	.	.	T	<b>I</b>	L	<b>T</b>	N
φ	n.s.	*	*	*, *	*	*	*	n.s.	n.s.	n.s.	*
		4	4	5,6			7				7

H - Host, P - social Parasite; NA – not available; n.s. – not significant; <sup>a</sup> - ancestor could not be inferred; φ – chance probability of having *n* parallel amino acid substitutions; last row indicate paired analysis (assigned numbers) for multiple shared substitutions; *B.luc* – *B. lucorum*, *B.tx* – *B.t. xanthopus*, *B.ter* – *B. t. terrestris*, *B.lap* – *B. lapidarius*, *B.pas* – *B. pascuorum*, *B.max* – *B. m. italicus*, *B.bo* – *B. bohemicus*, *B.vest* – *B. vestalis*, *B.per* - *B. perezi*, *B.cam* – *B. campestris*, *B.rup* – *B. rupestris*, *B.rud* – *B. r.corsicola*