



**HAL**  
open science

## A genome scan of diversifying selection in *Ophiocordyceps zombie-ant* fungi suggests a role for enterotoxins in co-evolution and host specificity

Noppol Kobmoo, Duangdao Wichadakul, Nuntanat Arnarnart, Ricardo  
Rodriguez de La Vega, Janet J Luangsa-Ard, Tatiana Giraud

► **To cite this version:**

Noppol Kobmoo, Duangdao Wichadakul, Nuntanat Arnarnart, Ricardo Rodriguez de La Vega, Janet J Luangsa-Ard, et al.. A genome scan of diversifying selection in *Ophiocordyceps zombie-ant* fungi suggests a role for enterotoxins in co-evolution and host specificity. *Molecular Ecology*, 2018, 27 (18), pp.3582 - 3598. 10.1111/mec.14813 . hal-01921814

**HAL Id: hal-01921814**

**<https://hal.science/hal-01921814>**

Submitted on 22 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **A genome scan of diversifying selection in *Ophiocordyceps* zombie-ant fungi**  
2 **suggests a role for enterotoxins in coevolution and host specificity**

3  
4 **Running title: Diversifying selection in zombie-ant fungi**

5  
6 Noppol KOBMOO<sup>1,2</sup>, Duangdao WICHADAKUL<sup>3,4</sup>, Nuntanat ARNAMNART<sup>2</sup>, Ricardo C.  
7 RODRÍGUEZ DE LA VEGA<sup>1</sup>, Janet-Jennifer LUANGSA-ARD<sup>2</sup>, Tatiana GIRAUD<sup>1</sup>.

8 <sup>1</sup> Ecologie Systématique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay 91400 Orsay,  
9 France

10 <sup>2</sup> National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Development  
11 Agency (NSTDA), 113 Thailand Science Park, Khlong Neung, Khong Luang, 12120 Pathum Thani, Thailand

12 <sup>3</sup> Chulalongkorn University Big Data Analytics and IoT Center (CUBIC), Department of Computer Engineering,  
13 Faculty of Engineering, Chulalongkorn University, Payathai Rd., Wangmai, Pathumwan 10330, Bangkok Thailand

14 <sup>4</sup> Center of Excellence in Systems Biology, Faculty of Medicine, Chulalongkorn University, Payathai Rd., Wangmai,  
15 Pathumwan 10330 Bangkok, Thailand

16  
17 Corresponding author: Noppol Kobmoo,

18 E-mail: [noppol.kob@biotec.or.th](mailto:noppol.kob@biotec.or.th)

19  
20 Key words: *Ophiocordyceps*, host specificity, adaptation, positive selection, enterotoxins

21

## 22 **Abstract**

23 Identification of the genes underlying adaptation sheds light on the biological functions targeted  
24 by natural selection. Searches for footprints of positive selection, in the form of rapid amino-acid  
25 substitutions, and the identification of species-specific genes have proved to be powerful  
26 approaches to identifying the genes involved in host specialization in plant-pathogenic fungi. We  
27 used an evolutionary comparative genomic approach to identify genes underlying host adaptation  
28 in the ant-infecting genus *Ophiocordyceps*, which manipulates ant behavior. A comparison of the  
29 predicted genes in the genomes of species from three species complexes – *O. unilateralis*, *O.*  
30 *australis* and *O. subramanianii* – revealed an enrichment in pathogenesis-associated functions,  
31 including heat-labile enterotoxins, among species-specific genes. Furthermore, these genes were  
32 overrepresented among those displaying significant footprints of positive selection. Other  
33 categories of genes suspected to be important for virulence and pathogenicity in entomopathogenic  
34 fungi (e.g., chitinases, lipases, proteases, core secondary metabolism genes) were much less  
35 represented, although a few candidate genes were found to evolve under positive selection. An  
36 analysis including orthologs from other entomopathogenic fungi in a broader context showed that  
37 positive selection on enterotoxins was specific to the ant-infecting genus *Ophiocordyceps*.  
38 Together with previous studies reporting the overexpression of an enterotoxin during behavioral  
39 manipulation in diseased ants, our findings suggest that heat-labile enterotoxins are important  
40 effectors in host adaptation and coevolution in the *Ophiocordyceps* entomopathogenic fungi.

41

## 42 **Introduction**

43           The identification of genes underlying adaptation is a major goal in evolutionary biology,  
44 as it can shed light on the biological functions targeted by natural selection and the genetic  
45 mechanisms generating new, adaptive variants. Innovation may be generated during evolution by  
46 gene duplication followed by rapid amino-acid substitutions in one of the copies (Ohno 1970,  
47 Zhang et al. 2002). Gene losses can also be adaptive (Juárez-Vázquez et al. 2017), particularly in  
48 pathogens, as the absence of a molecule recognized by the host may enable the pathogen to colonize  
49 its host without triggering a response from the host immune system (Albalat and Cañestro 2016;  
50 Ghanbarnia et al. 2015, Rouxel and Balesdent 2017). Gene duplications and losses result in the  
51 presence of species-specific genes, which are often over-represented among the genes involved in  
52 adaptation (Gladieux et al. 2014, Lespinet et al. 2002; Zhou et al. 2015). Adaptation may also occur  
53 through positive selection, with rapid amino-acid substitutions, typically detected as higher rates  
54 of non-synonymous substitutions ( $d_N$ ) than of synonymous substitutions ( $d_S$ ) among orthologous  
55 genes of closely related species (Kimura 1983; Ina 1996). Comparisons of  $d_N/d_S$  ratios to neutral  
56 expectations therefore also constitute a powerful approach to identifying genes under recurrent  
57 positive selection (Yang and Nielsen 1998; Yang et al. 2000, Yang et al. 2005).

58           Pathogens are particularly interesting models for investigations of the genomic mechanisms  
59 of adaptation, as they are locked in an arms race with their hosts, leading to continuous, rapid  
60 evolution (Anderson et al. 2010; Kurtz et al. 2016). Identification of the genes underlying host-  
61 specific adaptations in pathogens improves our fundamental understanding of natural selection and  
62 evolution, but it also has more applied implications, shedding light on major epidemics and disease  
63 emergence in plants and animals (Möller and Stukenbrock 2017).

64 Fungi are the principal pathogens of plants (Anderson et al. 2004), and they also represent  
65 threats to the health of many animals (Sexton and Howlett 2006; Fisher et al. 2012). Many studies  
66 have searched for genes under positive selection as a means of identifying genes and functions  
67 involved in the species-specific adaptation of fungal pathogens of plants (Aguileta et al. 2009;  
68 Möller and Stukenbrock 2017). For example, in the *Microbotryum* and *Botrytis* fungal plant  
69 pathogens, many such genes have been identified through comparative transcriptomics studies as  
70 being under recurrent positive selection and they were involved in biological processes important  
71 for the recognition and cell signaling between the host and the pathogen (Aguileta et al. 2010,  
72 2012). More recently, next-generation sequencing made it possible to perform genome-wide scans  
73 in plant -pathogenic fungi, resulting in the identification of an array of effectors under positive  
74 selection (Badouin et al. 2017; Poppe et al. 2015; Schirmann et al. 2018; Stukenbrock et al. 2011;  
75 Wicker et al. 2013) and species- or lineages-specific genes underlying adaptations (Baroncelli et  
76 al. 2016, Hartmann et al. in press2018).

77 By contrast, far fewer such studies have been performed on entomopathogenic fungi (Wang  
78 and Wang 2017), despite the importance of identifying genes underlying host-specific adaptation  
79 for the use of these fungi as biological control agents against insect pests in agriculture (Wang and  
80 Feng 2014). Furthermore, an understanding of host-specificity and evolution in these insect  
81 pathogens is of fundamental interest in its own right, particularly for fungi able to manipulate the  
82 behavior of the insect host for their own benefit, as in the “zombie ant” phenomenon. Most genomic  
83 studies on entomopathogenic fungi have focused on species with agricultural applications such as  
84 *Beauveria bassiana* and *Metarhizium anisopliae* (Gao et al. 2011, Hu et al. 2014, Pattemore et al.  
85 2014). However, these species have broad host ranges, and may not, therefore, be the best models  
86 in which to study the genomics of host specialization and coevolution. Specialist *Metarhizium*

87 strains and species generally have fewer genes, with notably fewer genes encoding host-killing  
88 toxins (Wang et al. 2009), but more genes evolving under positive selection than generalists (Hu  
89 et al. 2014). However, it remains unclear how adaptation has shaped the genomes of closely related  
90 fungal entomopathogens specializing on different hosts.

91 We therefore tried to identify genes involved in host specificity in three complexes of  
92 closely related species from the genus *Ophiocordyceps* (Hypocreales, Ascomycota): *O. unilateralis*  
93 *sensu lato*, *O. subramanianii s.l.* and *O. australis s. l.* One of the key features of these pathogens  
94 is their ability to manipulate their hosts to promote their own dispersal. Infected ants, often  
95 described as “zombie ants”, leave their nests and develop erratic behavior, wandering alone into  
96 vegetation and then biting into a leaf located at a precise height and orientation optimal for  
97 subsequent fungal dispersal just before they die. Fungal spores produced from the diseased ant are  
98 thus dispersed farther, from a height (Hughes et al. 2011; Pontoppidan et al. 2009; De Bekker et  
99 al. 2017; Hughes et al. 2016). *Ophiocordyceps unilateralis s.l.* is a highly diverse complex of  
100 pathogenic cryptic species specific to formicine ants. It is distributed worldwide and many species  
101 occur together in sympatry while displaying strong host specificity (Araújo et al. 2017, Evans et  
102 al. 2011; Kobmoo et al. 2012, 2015). Ants develop erratic behavior only when infected with their  
103 specific pathogen species (De Bekker et al. 2014, Sakolrak et al. 2018). The taxonomy and  
104 phylogeny of the other ant-manipulating *Ophiocordyceps* species complexes have been studied in  
105 less detail, but host specificity is also considered to be the rule for these other taxa (Araujo et al.  
106 2017).

107 We conducted a comparative genomic study of ant-infecting *Ophiocordyceps* species, with  
108 the aim of identifying genes underlying host specificity by searching for species-specific genes and  
109 genes evolving under positive selection. We sequenced the genomes of two closely related species

110 of the *O. unilateralis* complex from Thailand: *O. camponoti-leonardi* and *O. camponoti-saundersi*,  
111 specific to the ants *Camponotus leonardi* and *C. saundersi*, respectively. We also improved the  
112 available genome assembly of another species of this complex, *O. polyrhachis-furcata*, specific to  
113 *Polyrhachis furcata* (Wichadakul et al. 2015), and used the published genomes of other ant-  
114 infecting *Ophiocordyceps* species (De Bekker et al. 2017): one genome of each of two species of  
115 *O. unilateralis s.l.*, *O. kimflemingiae* from the US infecting *C. castaneus* (Araujo et al. 2017; de  
116 Bekker et al. 2015) and *O. camponoti-rufipedis* from Brazil specific to *C. rufipes* (Araujo et al.  
117 2017; Evans et al. 2011); one genome of *O. subramanianii s.l.* from a ponerine ant in Ghana; one  
118 genome of each of two strains of *O. australis s.l.* found on different ponerine ant species, from  
119 Ghana and Brazil, probably belonging to different cryptic species (De Bekker et al. 2017).

120 The entomopathogenic fungi of the order Hypocreales are known to infect their host by  
121 penetrating the cuticle (Boomsma et al. 2014). This process requires an array of proteinases, lipases  
122 and chitinases. The acquisition of nutrients from the host requires proteases and glycoside  
123 hydrolases, including trehalases in particular, as trehalose is a major carbon source present in the  
124 insect hemolymph (Thompson 2003). Secondary metabolites, including toxins, help to combat the  
125 host immune system and eventually kill the insect (Ortiz-Urquiza et al. 2010; Schrank and  
126 Vainstein 2010). *Ophiocordyceps polyrhachis-furcata* has a more extensive family of genes  
127 encoding putative heat-labile enterotoxins than other specialist entomopathogenic fungi  
128 (Wichadakul et al. 2015), and some of these genes are expressed during host-specific behavioral  
129 manipulation. Heat-labile enterotoxins may, therefore, act as neuromodulators (de Bekker et al.  
130 2015). We hypothesized that enterotoxin-coding genes would be under recurrent positive selection  
131 in ant-manipulating *Ophiocordyceps* fungi, as they have probably been involved in coevolution  
132 with the host and in host-specific adaptation. Small proteins secreted by fungal pathogens are often

133 involved in interactions with the host (Barrett and Heil 2012; Rafiqi et al. 2012). We therefore  
134 conducted genome scans for positive selection and focused on the heat-labile enterotoxin gene  
135 family and small secreted proteins. We conducted formal tests for positive selection (statistical  
136 comparisons of models of evolution with and without diversifying selection). As such tests detect  
137 only highly recurrent and rapid positive selection, we also investigated the 5% of genes with the  
138 highest dN/dS values. High dN/dS values, even if below 1, may be indicative of positive selection  
139 at a few sites in the protein, although they may also result from relaxed selection. In several classes  
140 of genes thought to be important for virulence and pathogenicity in entomopathogenic fungi (e.g.,  
141 chitinases, lipases, proteases, small secreted proteins), only a few genes showed signs of selection  
142 or species specificity. By contrast, we found that heat-labile enterotoxins were overrepresented  
143 among both the species-specific genes and the genes with significant footprints of positive  
144 selection. An analysis including enterotoxin-encoding genes from other entomopathogenic fungi  
145 (Hypocreales), that do not manipulate host behavior, showed that positive selection was specific to  
146 the ant-infecting genus *Ophiocordyceps*. These findings suggest that heat-labile enterotoxins are  
147 important effectors involved in host adaptation and coevolution in entomopathogenic  
148 *Ophiocordyceps* fungi.

149

## 150 **Materials and methods**

### 151 *Sampling and sequencing*

152 In 2015, we collected a sample of *O. camponoti-leonardi* (strain NK511ss-8) from  
153 Kalayaniwattana district, in Chiang Mai province in Thailand, and a sample of *O. camponoti-*  
154 *saundersi* (strain NK405ss-6) from the Phu Kiew National Park, in Chaiyaphum province. We used

155 the reference genome of *O. polyrhachis-furcata* (strain BCC54312) (Wichadakul et al. 2015); we  
156 aimed to improve the existing reference genome, but the original strain BCC54312 could not be  
157 grown from the culture collection. We therefore collected three additional samples of this species  
158 (strains NK275ss-12, NK142ss and NK294ss-20), in 2013 and 2014, from the same site as the  
159 reference strain, in Khao Yai National Park, Nakhon Rachasima province. The collected samples  
160 were isolated and grown as described by Wongsas et al. (2005). We complied with the Nagoya  
161 protocols on access and benefit-sharing, by obtaining authorization from the Department of  
162 National Parks, Wildlife and Plant Conservation (DNP) at the Ministry of Natural Resources and  
163 Environment of Thailand for all strain collections. After two to three months of growth on Grace  
164 Insect Cell Medium (Sigma-Aldrich), the mycelia and spores were harvested and DNA was  
165 extracted with the NucleoSpin® Soil kit (Macherey-Nagel). The long incubation period is due to  
166 the fact that *O. unilateralis* species in Thailand are very fastidious to grow, requiring few steps of  
167 enlarging the culture scale to a sufficient level for DNA extraction. Genomic libraries were  
168 constructed (150 bp paired-end reads) for sequencing with an Illumina HiSeq3000 machine at the  
169 GenoToul platform (Toulouse, France).

170

### 171 ***Read pretreatment, de novo assembly and improvement of the reference genome***

172 The raw reads were trimmed to remove adapters and low-quality bases from their ends ( $q >$   
173 20). Duplicate reads were removed using Picard Tools MarkDuplicate. The reference genomes for  
174 *O. camponoti-leonardi* and *O. camponoti-saundersi* were assembled *de novo* with SPAdes  
175 (Bankevich et al. 2012), which progressively integrates k-mers of increasing size. The k-mer sizes  
176 used were 21, 33, 55, 77, 99, 119, and 127 for NK405ss-6, and 21, 33, 55, 77, 99 and 115 for

177 NK511ss-8. The appropriate maximum k-mer sizes were estimated with Kmergenie (Chikhi and  
178 Medvedev 2013).

179 The reads obtained for the new *O. polyrhachis-furcata* samples were used to fill gaps in the  
180 existing reference genome of this species with GapFiller (Boetzer et al. 2012), which mapped the  
181 reads onto the reference sequence over the regions flanking the gaps and identified a consensus  
182 between reads overlapping the gaps. In total, 175 of 3915 gaps were closed (identifying around 1.6  
183 Mb from a total gap length of 2.4 Mb in the reference genome).

184

### 185 ***Gene prediction and functional annotation***

186 Gene prediction was based exclusively on scaffolds of more than 1 kb in length, and  
187 involved a two-round approach based on MAKER (Cantarel et al. 2008). Gene sets were initially  
188 predicted with CEGMA (Parra et al. 2007) and GeneMark-ES (Lomsadze et al. 2005), and were  
189 then used as inputs into MAKER for the first round of prediction. The predicted proteins and  
190 transcripts identified in previous studies on *O. polyrhachis-furcata* (Wichadakul et al. 2015) were  
191 also used as a training set for MAKER. The predicted gene set from this first round was then fed  
192 into SNAP (Korf 2004) and Augustus (Keller et al. 2011). The output of these two tools was then  
193 fed back into MAKER for a second round of prediction.

194 The predicted proteins were annotated with InterProScan 5 (Jones et al. 2014), which also  
195 associated the protein domains detected with sequences in the Pfam (Finn et al. 2016) and KEGG  
196 (Ogata et al. 1999; Kanehisa et al. 2016) databases and with Gene Ontology (GO) terms. Small  
197 secreted proteins (SSPs) were identified as proteins of less than 300 amino acids with signal  
198 peptides but no transmembrane signature detected by SignalP (Petersen et al. 2011). Proteolytic  
199 enzymes were also annotated, by Blast analysis of the predicted proteins against the MEROPS

200 database (Rawlings et al. 2016) with a threshold e-value of  $1e^{-20}$ . Enzymes with activity against  
201 carbohydrates were annotated with dbCAN (Yin et al. 2012); only genes with Pfam domains  
202 consistent with those of the CAZYme database were retained. InterProScan recognized glycoside  
203 hydrolase domains for some genes that were not detected by the CAZYme analysis toolkit; these  
204 genes were also retained. The core genes of secondary metabolic gene clusters (SMGCs) were  
205 predicted with SMURF (Khaldi et al. 2010), based on Pfam and Tigrfam domains and on gene  
206 positions on scaffolds. SMGCs were predicted with the fungal version of antiSMASH (Weber et  
207 al. 2015). SMGC homology across species was inferred with BiG-SCAPE (Navarros-Munõz J.,  
208 <https://git.wageningenur.nl/medema-group/BiG-SCAPE/wikis/home>), which classified SMGCs  
209 into families on the basis of Jaccard similarity indices between clusters. RepeatMasker was used  
210 to predict repetitive elements for the three species from Thailand.

211

### 212 ***Orthology and phylogenomics***

213 In addition to the predicted proteins from the *de novo* assembled and improved genomes of  
214 *O. unilateralis* species from Thailand, we also included in our analyses the predicted proteins of  
215 other ant-infecting *Ophiocordyceps* fungi specific to different ant species and originating from  
216 different geographic areas (De Bekker et al. 2017). We used the available genomes from two  
217 additional *O. unilateralis s.l.* species (*O. kimflemingiae* from the USA and *O. camponoti-rufipedis*  
218 from Brazil), from two cryptic species of *O. australis s.l.*, from Ghana and Brazil (De Bekker et al.  
219 2017), and from *O. subramanianii s.l.*, also from Ghana. The predicted proteins corresponding to  
220 all these genomes were subjected to Blast comparisons with each other, with a significance  
221 threshold e-value of  $1e^{-5}$ . The Blast results were used as input for orthAogue (Ekseth et al. 2014),  
222 a tool for the rapid inference of orthologous groups with the Markov clustering algorithm (MCL,

223 Dongen 2000). This algorithm recovers species-specific paralogous groups, with genes from a  
224 given species considered to be more closely related to each other than to any other gene in any  
225 other species. The functional annotations obtained for *O. polyrhachis-furcata* were transferred to  
226 the other species for gene copies in the same orthologous group. Species-specific paralogous genes  
227 were annotated as described above. We analyzed GO term enrichment among species- or complex-  
228 specific paralogs, with the topGO package in R (Alexa and Rahnenfuhrer 2016). Fisher's exact  
229 tests were used to compare gene counts between paralogous species-specific or complex-specific  
230 groups and the whole gene set for the species or complex, respectively.

231 Sequences within all orthologous groups were aligned with MACSE (Ranwez et al. 2011)  
232 for further analyses. A phylogenetic tree with bootstrap support was constructed according to the  
233 GTRCAT model under RaxML-HPC v8.1.5 (Stamatakis 2014), exclusively with the nucleotide  
234 sequences for all the one-to-one orthologous groups in which each species was represented.

235

### 236 ***Detection of positive selection***

237 Pairwise ratios of non-synonymous-to-synonymous substitutions (dN/dS) were calculated  
238 between species (Yang and Nielsen 2000) with the yn00 program implemented in PAML v.4.8a  
239 (Yang 2007). The mean pairwise dN/dS ratios were calculated and used to assess variation across  
240 single-copy orthologous groups with at least 4 species represented. Groups with dS < 0.01,  
241 potentially resulting in inaccurate dN/dS estimates, and groups with excessively high dN/dS (>10)  
242 ratios, were discarded. The functions overrepresented among the 5% of genes with the highest  
243 dN/dS ratios were inferred by an analysis of enrichment in GO terms. A mean dN/dS>1 for a given  
244 gene indicates positive selection, whereas high dN/dS values below 1 can be due to positive  
245 selection on a small number of sites within the protein or to relaxed selection.

246 We also formally tested for positive selection by performing site-model likelihood ratio  
247 tests (LRTs) with the CodeML program implemented in PAML v.4.8a (Yang 2007), excluding  
248 gaps and ambiguous sites and using trees inferred under GTRCAT model from respective  
249 orthologous groups. CodeML estimates the parameter omega ( $\omega = dN/dS$ ) by maximum likelihood  
250 methods, allowing variation between sites. While the pairwise measures above only approximate  
251 synonymous and non-synonymous rates, likelihood ratio tests (LRTs) statistically compare two  
252 models of evolution, one in which  $\omega < 1$  (null model) at all sites and another in which  $\omega > 1$  at  
253 some sites (alternative hypothesis of positive selection); LRTs thus indicate whether a model with  
254 positive selection is more likely than a model without positive selection. We compared the M7  
255 (beta distribution of  $\omega$ ) and M8 (beta distribution of  $\omega$  with a proportion of sites with  $\omega > 1$ ; Nielsen  
256 & Yang 1998, Yang et al. 2000) models, and the M8a (similar to M8 but with a category of sites  
257 evolving with  $\omega = 1$ ) and M8 (Swanson et al. 2003) models in LRTs. Only genes with a *p*-value  
258 below 0.05 after false-discovery rate (fdr) correction were considered significant. The M7 vs. M8  
259 test is known to lack robustness when the probability mass is located around  $\omega=1$ , in which case  
260 this test gives a high proportion of false positives; under these conditions, the M8a vs. M8 test is  
261 preferred (Swanson et al. 2003). We ensured the robustness of our results by considering only  
262 genes in which significant evolution under positive selection was detected in both tests. We  
263 checked for enrichment in particular GO terms among the genes evolving under positive selection.

264 We also investigated whether genes encoding heat-labile enterotoxins evolved under  
265 positive selection specifically in ant-infecting *Ophiocordyceps* and not in other Hypocrealean  
266 fungi. We therefore downloaded predicted gene sequences from other Hypocrealean fungi that  
267 were annotated as putative heat-labile enterotoxins from the Ensembl Genome database (Herrero  
268 et al. 2016). Putative heat-labile enterotoxin genes were retrieved for 14 entomopathogenic fungi

269 (one strain per species) (Additional Table S1): *Metarhizium anisopliae* ARSEF23 (24 genes), *M.*  
270 *acridum* CQMa 102 (3 genes) (Pattemore et al. 2014); *M. album* ARSEF1941 (12 genes), *M.*  
271 *brunneum* ARSEF3297 (32 genes), *M. guizhouense* ARSEF977 (32 genes), *M. majus* ARSEF297  
272 (32 genes) (Hu et al. 2014); *M. rileyi* RCEF4871 (3 genes), *Isaria fumosorosea* ARSEF2679 (5  
273 genes), *Aschersonia aleyrodinis* RCEF2490 (14 genes), *Cordyceps confragosa* RCEF1005 (6 genes),  
274 *C. brongniartii* RCEF3172 (30 genes) (Shang et al. 2016), *Cordyceps militaris* CM01 (one gene,  
275 Zheng et al. 2011), *Beauveria bassiana* ARSEF2860 (6 genes, Xiao et al. 2012); and  
276 *Ophiocordyceps sinensis* Co18 (13 genes, Xia et al. 2017). We also included putative heat-labile  
277 enterotoxin sequences from two nematode-killing fungi: *Purpureocillium lilacinum* PLBJ-1 (two  
278 genes, Wang et al. 2016) and *Pochonia chlamydosporia* 170 (four genes). Orthologs between these  
279 sequences and the putative enterotoxins of *O. unilatealis* species studied here were identified. The  
280 occurrence of clade-specific positive selection in *O. unilateralis* was assessed with branch model  
281 LRTs in PAML (Yang 1998; Yang and Nielsen 1998) and with the BUSTED test, an alignment-  
282 wide test of episodic positive selection (Murrell et al. 2015). Both these tests are log-likelihood  
283 ratio tests comparing a model in which positive selection is allowed in the foreground branches  
284 (i.e. the clade of interest) to the null model in which positive selection is not allowed. The branch  
285 model (Yang and Nielsen 1998), as implemented in PALM, detects positive selection by allowing  
286 a candidate clade to have a dN/dS ratio higher than those of the other branches (background  
287 branches) without taking into account variation between sites or allowing variation between  
288 branches of the same category. By contrast, BUSTED is a stochastic test using information from  
289 all sites and branches; it is therefore considered to have greater statistical power (Murrell et al.  
290 2015).

291

## 292 **Results**

### 293 *General genome features*

294       The three reference genomes of closely related species sequenced here differed  
295 considerably in size, *O. camponoti-saundersi* (OCS) being the largest (49.26 Mb), followed by *O.*  
296 *polyrhachis-furcata* (OPF) (43.25 Mb) and *O. camponoti-leonardi* (OCL) (37.91 Mb). These  
297 differences probably partly reflect methodological differences as the OPF genome is an improved  
298 version of a genome sequenced with a different technology (454 pyrosequencing combined with  
299 Illumina mate-pair sequencing, Wichadakul et al. 2015). OCL and OCS were sequenced and  
300 assembled with the same methodology, so the observed differences probably reflect genuine  
301 differences in genome size. OCS also had more scaffolds (1700) than OCL (531). OPF had fewer  
302 scaffolds and larger contigs, due to the use of variable-size mate-pair libraries (Wichadakul et al.  
303 2015) (Table 1). These genomes are markedly larger than those reported for *O. kimflemingiae*  
304 (OKi: 23.91 Mb), *O. camponoti-rufipedis* (OCR: 21.91 Mb), *O. australis s.l.* from Brazil (OAB:  
305 23.32 Mb) and from Ghana (OAG: 22.19 Mb), and *O. subramanianii s.l.* (OSS: 32.31 Mb), but all  
306 these previously published genomes were more fragmented than our assemblies (Table 1).

307       Despite the differences in genome size, the numbers of predicted genes were of a same  
308 order of magnitude across species (Table 1), although the number of predicted genes was  
309 nevertheless largest for OSS. For the three species from Thailand, OPF had the largest number of  
310 predicted genes, probably because the protein and transcript training set used for prediction came  
311 from this species. The number of SSPs was similar between the three Thai species. The number of  
312 genes with assigned Pfam domains or InterPro classification and the complete predicted gene sets

313 obtained by core eukaryotic genes mapping (CEGMA) were also very similar in the three species  
314 (~95 %: Table 1), but smaller than those for species from the New World (~99 %).

315

### 316 ***Orthology and phylogenomics***

317 The genomes used in this study were sequenced from individuals belonging to one of the  
318 three species complexes: *O. unilateralis s.l.*, *O. australis s.l.* and *O. subramanianii s.l.* Most of the  
319 genes were common to all three complexes (Fig. 1a): 8,554 orthologous groups were retrieved,  
320 5,718 of which were common to all complexes. For orthologous groups present in only one of the  
321 three complexes (Supporting information: Table S2), pathogenesis (GO:0009405) was the function  
322 displaying the most significant enrichment in all complexes (Bonferroni-corrected  $p$ -values:  $2e^{-10}$   
323 for *O. unilateralis s.l.*, 0.016 for *O. australis s.l.*,  $3.4e^{-5}$  for *O. subramanianii s.l.*), mostly due to  
324 the presence of genes encoding putative heat-labile enterotoxins in these complex-specific genes.  
325 Complex-specific genes were also found to be enriched in inter-species interactions and multi-  
326 organism process functions.

327 Within each species complex, most of the genes were common to several species (Fig. 1b and  
328 1c). The function pathogenesis was found to be overrepresented among species-specific genes  
329 (Supporting information: Tables S3 - S4), due to the presence of genes encoding heat-labile  
330 enterotoxins, and SSPs (Tables 2 and 3). In particular, we detected an overrepresentation of SSPs  
331 among the genes unique to *O. kimflemingiae* ( $p$ -value = 0.003) relative to *O. unilateralis s.l.*  
332 complex, and among the genes unique to *O. australis* from Brazil relative to *O. australis s.l.*  
333 complex ( $p$ -value = 0.001). None of these species-specific SSPs had a predicted function,  
334 suggesting an expansion of rapidly evolving families of effectors (Kim et al. 2016).

335           There were 4,651 orthologous groups common to all eight genomes. We used a subset of  
336 4,014 single-copy orthologous groups common to all species to construct a phylogenetic tree (Fig.  
337 2). This tree recovered the expected relationships between the sibling species from Thailand, with  
338 *O. polyrhachis-furcata* being the most closely related to *O. camponoti-leonardi* (Kobmoo et al.  
339 2012, 2015); the species from the Americas, *O. kimflemingiae* and *O. camponoti-rufipedis*,  
340 clustered together but were separate from those from Thailand, corresponding to the separation  
341 between the Old and New Worlds observed in a previous study (Evans et al. 2018). The two *O.*  
342 *australis s.l.* species were grouped together and formed, with *O. subramanianii*, an outgroup to the  
343 *O. unilateralis* complex.

344

#### 345 ***Variation of dN/dS across genomes and putative functions***

346           The median pairwise dN/dS ratio was 0.081, indicating that most single-copy orthologs  
347 evolved under strong purifying selection (Fig. 3a). No orthologous group had dN/dS > 1  
348 (Supporting information: Table S5). We investigated the putative functions of the 5% of genes with  
349 the highest dN/dS values (297 genes) (Supporting information: Table S5), even if these ratios were  
350 below 1, as this could be indicative of positive selection at a small number of sites in the protein,  
351 although relaxed selection cannot be excluded for dN/dS <1. Fisher's exact tests for enrichment in  
352 GO term annotation among the 5% of genes with the highest dN/dS values relative to all single-  
353 copy orthologous groups (i.e., dN/dS > 0.253) indicated that the functions pathogenesis,  
354 interspecies interaction between organisms and multi-organism process (respective Bonferroni-  
355 corrected *p*-value = 3.4e-17, 2.2e-18 and 5.5e-180.005) were the most significantly  
356 overrepresented. The genes annotated with the pathogenesis (GO:0009405) GO term and with a  
357 dN/dS ratio in the top 5% were all putative heat-labile enterotoxin alpha chains. Twenty

358 orthologous groups annotated as heat-labile enterotoxins were found in one-to-one orthologous  
359 groups common to at least four species, thirteen of which had dN/dS values in the top 5%.  
360 Enterotoxins had a significantly higher mean dN/dS than the genome as a whole (*t*-test: Bonferroni-  
361 corrected  $p$ -value =  $5.19e^{-08}$ ) (Fig. 3b). The GO terms relating to glycoside hydrolases (GH  
362 families: chitinases GH18, trehalases GH37), proteases and lipases, essential to the primary  
363 functions of entomopathogenic fungi, such as nutrient acquisition and host cell penetration, were  
364 not overrepresented among the 5% of genes with the highest dN/dS values. The mean dN/dS values  
365 for these gene families were not significantly higher than that for the whole genome (*t*-tests;  
366 chitinases: bonferroni-corrected  $p$ -value=0.926, lipases: bonferroni-corrected  $p$ -value=1.00,  
367 proteases: bonferroni-corrected  $p$ -value = 1.00, trehalases: bonferroni-corrected  $p$ -value=1.00)  
368 (Fig. 3b). None of the core genes in secondary metabolic gene clusters was found among the 5%  
369 of genes with the highest dN/dS values, and their mean dN/dS value was not significantly higher  
370 than that of other genes (*t*-test: bonferroni-corrected  $p$ -value=0.885) (Fig. 3b). The genes encoding  
371 putative SSPs had a significantly higher dN/dS ratio than non-SSP genes (*t*-test:  $p$ -value= $2.2e^{-16}$ )  
372 (Fig. 3b). This suggest that a higher proportion of genes may evolve under positive selection among  
373 SSP-encoding genes than among other genes.

374

### 375 ***Formal tests for positive selection***

376 LRTs (likelihood ratio tests comparing models with or without positive selection) were  
377 performed as formal genome-wide tests for positive selection. Of the 5,950 single-copy  
378 orthologous groups common to at least four species, 759 and 244 (12.76% and 4.10 %) were found  
379 to be evolving under positive selection in comparisons of the models M7 vs. M8 and M8a vs. M8,  
380 respectively (Supporting information: Table S6), with 144 genes found to be evolving under

381 positive selection in both tests. The functions of pathogenesis and toxin activity were significantly  
382 overrepresented among the annotations for these 144 genes (Table 4). This enrichment in the  
383 pathogenesis and toxin activity functions was due entirely to enterotoxin genes. Five to seven (25%  
384 - 35%) putative heat-labile enterotoxin-coding genes were identified as evolving under positive  
385 selection, depending on the evolution model tested (Fig. 4 and Supporting information: Table S7).  
386 These genes were also predicted to have a function in the extracellular region (GO:0005576), and  
387 most were predicted to have signal peptides. Enrichment was also observed for functions relating  
388 to kinases (e.g. ATP binding, kinase activity, transferase activities: Table 4.); significant results  
389 were obtained for 15 to 31 (8.1% - 16.8 %) kinases in LRTs and eight kinases gave significant  
390 results in both tests (Fig. 4 and Table S7). Kinases catalyze the transfer of a phosphate group from  
391 a high-energy molecule (ATP) to a substrate and are involved in various cellular processes. The  
392 proportion of kinases evolving under positive selection was lower than that of heat-labile  
393 enterotoxins (Fig. 4), but the numbers of kinases and heat-labile enterotoxins evolving under  
394 positive selection were similar, and these two functions were overrepresented among the genes  
395 evolving under positive selection. Most of these kinases were annotated as protein kinases,  
396 histidine kinases and phosphatidylinositol 3 and 4-kinases. These families of kinases are well  
397 known to be involved in cell signaling, essential for pathogen growth and survival and, thus, for  
398 pathogenesis and virulence (Lee et al. 2016).

399         The functions relating to hydrolytic enzymes important for pathogenesis (glycoside  
400 hydrolases, lipases, proteases) were not overrepresented among the genes evolving under positive  
401 selection. Indeed, the proportions of genes in these families found to be under positive selection  
402 were markedly smaller than those for heat-labile enterotoxins (Fig. 4). Nevertheless, several of the  
403 genes from these gene families were found to evolve under positive selection in model tests

404 (Supporting information: Table S7) and can be considered good candidates for involvement in  
405 coevolution and host specificity. One to three of the 11 chitinases (GH18) presented significant  
406 footprints of positive selection depending on the evolution model considered, and significant *P*-  
407 values were obtained in all tests for one of these enzymes. Chitinases are involved in the  
408 degradation of the insect cuticle, a major component of the insect exoskeleton, and in the  
409 degradation and remodeling of fungal cell walls (Adams 2004; Langner and Göhre 2016). Other  
410 GH families converging to various functions such as cellulase, glucanase, glucosidase,  
411 galactosidase (e.g., GH5, GH16, GH47, GH76) also included a few genes displaying significant  
412 tests of positive selection (Supporting information: Table S7). Neither of the two trehalases  
413 (GH37), which are thought to play important roles in nutrient acquisition within the host body,  
414 displayed significant signs of positive selection.

415         Zero to four of the nine subtilisin-like (MEROPS family S08 and S53) and trypsin-like  
416 proteases (MEROPS family S01), which are considered to act as cuticle-degrading proteases,  
417 presented significant footprints of positive selection, depending on the evolution model considered.  
418 Zero to one of 16 putative aspartyl proteases (MEROPS family A01) was found to evolve under  
419 positive selection following different models. However, none of these proteases yielded significant  
420 results in both tests (Fig. 4). Two to six of the 39 putative lipases yielded significant *P*-values in  
421 positive selection tests, and only one yielded significant *P*-values in both tests (Fig. 4; Supporting  
422 information: Table S7).

423         One to four of the seven core genes of secondary metabolites displayed significant  
424 signatures of positive selection, depending on the evolution model considered (Supporting  
425 information: Table S7). The only gene to yield significant *P*-values in both tests (orthologous group  
426 ORTHAg2248, Supporting information Table S7) encoded a polyketide synthase (PKS)-like

427 protein with a beta-ketoacyl synthase domain. Beta-ketoacyl synthase is involved in fatty-acid  
428 biosynthesis and has been shown to be involved in the production of polyketide antibiotics in fungi  
429 (Beck et al. 1990). The gene encoding this enzyme is part of a secondary metabolic gene cluster  
430 that is highly syntenic across the species of the *O. unilateralis* complex, but located in different  
431 clusters in *O. australis* and in *O. subramaniani* (Fig. 5).

432 We also investigated whether the genes previously identified as encoding possible  
433 “neuromodulators” (de Bekker et al. 2017), based on their overexpression during the manipulation  
434 of ant behavior, showed signs of positive selection. In total, 12 to 41 of these genes yielded  
435 significant results in tests for positive selection (Supporting information: Table S8). Five genes  
436 yielded significant results in both tests. These genes encoded a short-chain dehydrogenase, a DNA  
437 mismatch repair protein (MutC), a DNA replication factor, an ATPase and a protein with no  
438 functional annotation. Seven other genes yielded results only in the M8a vs. M8 test, which is more  
439 robust than the M7 vs. M8 test. These seven genes included oxidoreductases clearly involved in  
440 metabolic reactions: a protein with a ferric reductase transmembrane-like domain, a flavodoxin  
441 oxidoreductase, and an oxidoreductase binding to a molybdopterin cofactor with a cytochrome b5-  
442 like heme/steroid-binding domain. The first of these genes was shown to be associated with iron  
443 uptake in yeast (Roman et al. 1993), whereas the product of the second mediates iron-free electron  
444 transfer. The third of these genes may encode a nitrate reductase or sulfite oxidase, both of which  
445 are involved in nitrogen assimilation. These putative neuromodulators thus seem to be involved in  
446 host resource utilization. The neurological disorder displayed by zombie ants infected with  
447 *Ophiocordyceps* may result from the pathogen outcompeting the host for iron and nitrogen.

448

449 ***Positive selection of heat-labile enterotoxin genes specific to the ant-manipulating O.***  
450 ***unilateralis species complex***

451 The above results and those of previous studies (de Bekker et al. 2015; Wichadakul et al.  
452 2015) suggest that heat-labile enterotoxin genes are candidate genes for host-specific adaptation.  
453 We therefore investigated whether the positive selection detected above was specific to the ant-  
454 infecting *Ophiocordyceps* species or general to Hypocrealean entomopathogenic and nematode-  
455 killing fungi. Thirty-six orthologous groups of heat-labile enterotoxin genes were inferred for a  
456 group of 16 Hypocrealean entomopathogenic and nematode-killing fungi in addition to our eight  
457 focal species (Supporting information: Table S1); 22 of these orthologous groups included at least  
458 one sequence from the ant-infecting *Ophiocordyceps*, and 10 (42%) of these groups included only  
459 sequences from the ant-infecting *Ophiocordyceps* species. We further analyzed the only group  
460 (ORTHA<sub>gEnt13</sub>) common to at least four of the ant-infecting *Ophiocordyceps* species considered  
461 and sequences recovered from other species from Hypocreales, for which both site-model LRTs  
462 for positive selection were significant. This group included five sequences each from an *O.*  
463 *unilateralis* species. In a maximum likelihood tree, all the *O. unilateralis* sequences were located  
464 on the same branch (Fig. 6). The PAML branch-model LRTs indicated that this gene was evolving  
465 under positive selection specifically in the *O. unilateralis* clade ( $p$ -values < 0.001). The branch at  
466 the base and the internal branches of the *O. unilateralis* clade therefore had significantly higher  
467 dN/dS ratios than the other branches (Fig. 6). The BUSTED test, which is similar to PAML branch  
468 tests but considered more powerful, also gave a significant result ( $p$ -value = 6.16e-14).

469

470 **Discussion**

471 ***Enterotoxin genes as major candidate genes underlying host adaptation***

472 Comparative genomic studies of closely related species of fungal pathogens have shown  
473 that, in general, genes involved in adaptation, particularly those involved in virulence and  
474 pathogenicity, are species-specific, highly divergent and/or under diversifying selection, as a result  
475 of the arms race between host and pathogen or specialization on new hosts (Ghanbarnia et al. 2015;  
476 Huang et al. 2014; Plissonneau et al. 2017; Stukenbrock et al. 2011). We therefore used an  
477 evolutionary comparative genomic approach for identifying genes underlying host adaptation in  
478 ant-infecting *Ophiocordyceps* from three species complexes (*O. unilateralis s.l.*, *O. australis s.l.*  
479 and *O. subramaninii s.l.*). Genome comparisons showed that species complex-specific genes were  
480 enriched in genes associated with the function pathogenesis which included genes encoding heat-  
481 labile enterotoxins. The species-specific genes were also enriched in this function. However, most  
482 species-specific genes lacked functional annotation, perhaps due to their rapid evolution as part of  
483 the arms race between pathogen and host, resulting in homology no longer being detectable. Most  
484 of the small secreted proteins (SSPs), in particular, lacked predicted functions, but these proteins  
485 were particularly abundant among the species-specific genes. SSPs may act as effectors in  
486 pathogenicity, but the validation of their function requires further studies.

487 Heat-labile enterotoxin genes were also overrepresented in the orthologous groups with the  
488 highest rates of amino-acid differences between species, suggesting the occurrence of diversifying  
489 selection, which was confirmed by formal tests comparing models with and without positive  
490 selection. Furthermore, in the cases in which orthologs of enterotoxin genes were found in other  
491 entomopathogenic fungi, we inferred that positive selection was specific to the ant-infecting  
492 *Ophiocordyceps* clade. These findings support the view that heat-labile enterotoxins are effectors  
493 involved in host adaptation, as previously suggested based on observations of enterotoxin  
494 overexpression during manipulation of the behavior of the diseased ants (de Bekker et al. 2015)

495 and of the species-specific nature of behavioral manipulation (de Bekker et al. 2014, Sakolrak et  
496 al. 2018). The proximal mechanisms via which enterotoxins act during infection and the  
497 manipulation of host behavior remain unclear, but it has been suggested that these molecules  
498 interfere with the chemical communication of social insects; bacterial enterotoxins have been  
499 shown to affect pheromone production in boll weevils (Wiygul and Sikorowski 1986, 1991).  
500 Alterations to chemical communication may contribute to the modification of behavior in infected  
501 ant hosts.

502

### 503 *Minor role of the cuticle in exerting selective pressure leading to diversifying selection*

504 Hypocrealean entomopathogenic fungi are known to infect their insect hosts by penetrating  
505 the cuticle from the outside (Boomsma et al. 2014). An array of hydrolytic enzymes, including  
506 chitinases, lipases and proteases, is required to break through the insect cuticle (Ortiz-Urquiza and  
507 Keyhani 2013). Chitins are major constituents not only of insect cuticles, but also of fungal cell  
508 walls (Lagner and Göhre 2016) while lipids are a major component of the epicuticle waxy layer  
509 (Jarrold et al. 2007; Pedrini et al. 2013). Proteases are important for the penetration of the cuticle  
510 by fungi, and have been shown to be virulence factors for the infection of insect hosts (Shah et al.  
511 2005). Subtilisin proteases have been shown to play a particularly important role in regulating  
512 insect host specificity through the differential expressions of specific isoforms (Bye and Charnley  
513 2008; Mondal et al. 2016). We therefore hypothesized that the genes encoding chitinases, proteases  
514 and lipases might have evolved under diversifying selection. However, we found footprints of  
515 positive selection for only a few of putative genes encoding these enzymes in the ant-infecting  
516 *Ophiocordyceps* species. This challenges the widely accepted view that the insect cuticle, as a  
517 major barrier to infections, exerts a strong selective pressure on entomopathogenic fungi, leading

518 to different host ranges (Boomsma et al. 2014; Ortiz-Urquiza and Keyhani 2013; Wang and St.  
519 Leger 2005; Wang et al. 2011). Nevertheless, as the fungi in the three ant-infecting complexes  
520 considered here are all pathogens of formicine and ponerine ants, our findings do not rule out  
521 diversifying selection occurring across larger phylogenetic scales. These enzymes may be highly  
522 conserved among pathogens of formicine and ponerine ants, providing a common arsenal for  
523 attacking taxonomically related ants. There may also be constraints in the host or the fungus  
524 preventing rapid coevolution through changes to these molecules.

525

### 526 *Utilization of host resources*

527       Once inside the host, the pathogen requires other hydrolases for carbon assimilation.  
528 Efficient nutrient uptake from the host allows optimal proliferation of the fungus within its host,  
529 and ultimately leads to insect death (Luo et al. 2014). It has, therefore, been suggested that host  
530 resource utilization is crucial for host specificity (Gillespie et al. 2000). Trehalases, in particular,  
531 probably play an important role in this respect. Indeed, the fly pathogen *Entomophthora muscae*  
532 (Entomorphthorales) carries more trehalase-encoding genes in its genome than its close relative,  
533 the generalist *Conidiobolus coronatus*, which is a non-obligate pathogen (De Fine Licht et al.  
534 2017). We identified two trehalases with no positive selection signature as conserved across all  
535 species. Other glycoside hydrolases and lipases may be involved in breaking down primary carbon  
536 sources (Ortiz-Urquiza and Keyhani 2013; Schrank and Vainstein 2010). However, the evidence  
537 for positive selection is less robust for these enzymes. Thus, diversifying selection in ant-  
538 pathogenic *Ophiocordyceps* fungi probably acts less strongly on the function of carbon assimilation  
539 than on enterotoxins. Again, there may be constraints preventing the rapid evolution of host cuticle  
540 or fungal hydrolase and lipase functions.

541 Nitrogen also plays a key role in the proliferation of entomopathogenic fungi (Luo et al.  
542 2014). However, our results suggest that initial nutrient acquisition via proteinases is not under  
543 strong diversifying selection. Genes evolving under positive selection were not enriched in  
544 functions related to the assimilation of nitrogen or amino-acid synthesis.

545 In addition to carbon and nitrogen, iron uptake is also crucial for pathogen success (Bairwa  
546 et al. 2017, Haas 2012, Sutak et al. 2008). The candidate neuromodulator genes found to be under  
547 positive selection included iron-related oxidoreductases. In particular, one of the proteins identified  
548 had a ferric-reductase transmembrane domain, and another was a flavodoxin oxidoreductase.  
549 Proteins with ferric-reductase transmembrane domains have been shown to be crucial for ferric  
550 iron uptake in yeast (Roman et al. 1993), whereas flavodoxin is an iron-free electron-transfer  
551 protein facilitating a range of metabolic reactions in the absence of iron. Specialist  
552 entomopathogens kill their hosts more slowly than generalists (Boomsma et al. 2014). In such a  
553 context, ant-specific *Ophiocordyceps* might be expected to have developed strategies for hijacking  
554 resources from the host. The efficient acquisition of iron and an ability to divert its use may be the  
555 key to outcompeting the host during infection.

556

### 557 ***The role of kinases and signal transduction***

558 Kinase enzymes are widely recognized as participating in various cellular processes, crucial  
559 to growth and survival (Lee et al. 2016). The genes under positive selection in the ant-infecting  
560 *Ophiocordyceps* were enriched in kinase-related functions. Most were clearly related to signal  
561 transduction, which plays a crucial role in interactions between hosts and pathogens (Bahia et al.  
562 2018). Pathogens sense and respond to environmental stimuli, including the expression of virulence  
563 factor regulatory systems, in the hostile conditions of the host immune system. As extremely

564 specialized pathogens, ant-infecting *Ophiocordyceps* fungi must fine-tune their responses in the  
565 host body.

566

### 567 ***Importance of lipid metabolism***

568 Many entomopathogenic fungi are also thought to deploy a plethora of metabolites and  
569 toxins within the bodies of their hosts (Schrank and Vainstein 2010; Singh et al. 2016). The nature  
570 of these molecules probably differs between groups of insect-pathogenic fungi and remains to be  
571 precisely determined, but the principal molecules include polyketides (PKs) and non-ribosomal  
572 peptides (NRPs) (Gallo et al. 2013). We detected significant footprints of positive selection in some  
573 of the core genes of secondary metabolites. The most notable case concerned a PKS-like function  
574 involved in lipid biosynthesis. Lipids have been shown to be involved in pathophysiological  
575 processes in pathogenic fungi, but the role of the lipid signaling network in host-specific  
576 pathogenicity remains to be determined (Singh and Poeta 2011). Kinases are also known to  
577 participate in lipid signaling pathways, and the kinases with significant footprints of positive  
578 selection identified included phosphatidylinositol 3 and 4-kinase. The phosphorylated form of  
579 phosphatidylinositol plays an important role in lipid and cell signaling (Funaki et al. 2000). Lipid  
580 metabolism thus seemed to be subject to diversifying selection, although to a lesser extent than  
581 heat-labile enterotoxins.

582

### 583 **Conclusions**

584 We focused on three ant-infecting species complexes from genus *Ophiocordyceps*,  
585 including closely related species. Complex- and species-specific genes were found to be enriched  
586 in genes for heat-labile enterotoxins, and this gene family was found to be evolving under positive

587 selection to a greater extent than other candidate gene families. Our results thus suggest that the  
588 specific adaptation and coevolution of specialist species in the ant-infecting *Ophiocordyceps* fungi  
589 to their hosts is dependent on selection occurring within the body of the host rather than during  
590 cuticle penetration. By contrast, we detected little positive selection on lipases, proteases or  
591 chitinases, although we did identify a few interesting candidate genes from these groups.  
592 Comparative genomic studies of entomopathogenic fungi remain scarce, and the few studies that  
593 have been performed have focused exclusively on species of agricultural or medical interest. The  
594 findings of this study improve our understanding of the mechanisms of fungal adaptation to insect  
595 hosts, and future studies on fungal pathogens associated with other groups of insects should provide  
596 more general insight into the adaptation of entomopathogenic fungi and a more documented  
597 comparison with the mechanisms of adaptation in fungal pathogens of plants. The insect innate  
598 immune response seems to be much more specific than that in plants, suggesting a certain level of  
599 acquired immune response (Cooper and Eleftherianos 2017). Fungal pathogens of insects would  
600 be expected to display extensive expansions and contractions of gene families, as observed in plant  
601 pathogens, but the target functions may be different. Additional insight gleaned from  
602 entomopathogenic fungi would help to improve our general understanding of the mechanisms of  
603 adaptive evolution in eukaryotes.

604

## 605 **Acknowledgments**

606 This work was supported by the Marie Skłodowska Curie Action N° 655278 and Thailand  
607 Research Fund (TRF) Young Scientist Grant (TRG5780162) to NK. We would like to thank  
608 Alodie Snirc for advice concerning DNA extraction, Antoine Branca for suggestions about

609 bioinformatic protocols, Rayan Chikhi for training in genome assembly, Jérôme Collemare and  
610 Jorge C. Navarro-Munoz for their guidance on using AntiSMASH and BigSCAPE, and Suchada  
611 Mongkholsamrit and Kanoksri Tasanathai for the organization of sampling trips. We also would  
612 like to sincerely thank Clarissa De Bekker and David P. Hughes for kindly sharing their data on  
613 the candidate neuromodulators.

## 614 615 **References**

- 616  
617 Adams, D. J. (2004). Fungal cell wall chitinases and glucanases. *Microbiology*, 150(7), 2029–2035.  
618 doi:10.1099/mic.0.26980-0
- 619 Aguilera, G., Lengelle, J., Chiapello, H., Giraud, T., Viaud, M., Fournier, E. *et al.* (2012). Genes under  
620 positive selection in a model plant pathogenic fungus, *Botrytis*. *Infection, Genetics and Evolution*,  
621 12(5), 987–996. doi:10.1016/j.meegid.2012.02.012
- 622 Aguilera, G., Lengelle, J., Marthey, S., Chiapello, H., Rodolphe, F., Gendraud, A. *et al.* (2010). Finding  
623 candidate genes under positive selection in Non-model species: Examples of genes involved in host  
624 specialization in pathogens. *Molecular Ecology*, 19(2), 292–306. doi:10.1111/j.1365-  
625 294X.2009.04454.x
- 626 Aguilera, G., Refrégier, G., Yockteng, R., Fournier, E., & Giraud, T. (2009). Rapidly evolving genes in  
627 pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and  
628 protists. *Infection, Genetics and Evolution*, 9(4), 656–670. doi:10.1016/j.meegid.2009.03.010
- 629 Albalat, R., & Cañestro, C. (2016). Evolution by gene loss. *Nature Reviews Genetics*, 17(7), 379–391.  
630 doi:10.1038/nrg.2016.39
- 631 Alexa A and Rahnenfuhrer J (2016). *topGO: Enrichment Analysis for Gene Ontology*. R package version  
632 2.28.0.
- 633 Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., & Singh, K. B. (2010). Plants  
634 versus pathogens: An evolutionary arms race. *Functional Plant Biology*, 37(6), 499–512.  
635 doi:10.1071/FP09304
- 636 Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004).  
637 Emerging infectious diseases of plants: Pathogen pollution, climate change and agrotechnology  
638 drivers. *Trends in Ecology and Evolution*, 19(10), 535–544. doi:10.1016/j.tree.2004.07.021
- 639 Araújo, J. P. M., Evans, H. C., Kepler, R., Hughes, D. P. (2017). Zombie-ant fungi across continents: 15  
640 new species and new combinations with *Ophiocordyceps*. I. Myrmecophilous hirsutelloid species.  
641 *Studies in Mycology*, 90, 119-160.
- 642 Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguilera, G., Snirc, A., Le Prieur, S., Jeziorski, C.,  
643 Branca, A., Giraud, T. (2017). Widespread selective sweeps throughout the genome of model plant  
644 pathogenic fungi and identification of effector candidates. *Molecular Ecology*, 26, 2041-2062. doi:

- 645 10.1111/mec.13976
- 646 Bahia, D., Satoskar, A., Dussurget O. (2018). Cell signalling in host-pathogen interactions: The host point  
647 of view. *Frontiers in Immunology*, 9, 1-4. doi: 10.3389/fimmu.2018.00221
- 648 Bairwa, G., Hee Jung W., Kronstad J. (2017). Iron acquisition in fungal pathogens of human. *Metallomics:  
649 integrated biometal science*, 9(3), 215-227. doi:10.1039/c6mt00301j
- 650 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S. *et al.* (2012). SPAdes:  
651 A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of  
652 Computational Biology*, 19(5), 455–477. doi:10.1089/cmb.2012.0021
- 653 Baroncelli, R., Amby, D. B., Zapparata, A., Sarrocco, S., Vannacci, G., Le Floch, G. *et al.* (2016). Gene  
654 family expansions and contractions are associated with host range in plant pathogens of the genus  
655 *Colletotrichum*. *BMC Genomics*, 17(555), 1–17. doi:10.1186/s12864-016-2917-6
- 656 Barrett, L. G., & Heil, M. (2012). Unifying concepts and mechanisms in the specificity of plant-enemy  
657 interactions. *Trends in Plant Science*, 17(5), 282–292. doi:10.1016/j.tplants.2012.02.009
- 658 Beck, J., Ripka, S., Siegner, A., Schiltz, E., Schweizer, E. (1990). The multifunctional 6-methylsalicylic  
659 acid synthase gene of *Penicillium patulum*. Its gene structure relative to that of other polyketide  
660 synthases. *European Journal of Biochemistry*, 192(2), 487-498.
- 661 Boetzer, M., Pirovano, W., Zerbino, D., Birney, E., Simpson, J., Wong, K. *et al.* (2012). Toward almost  
662 closed genomes with GapFiller. *Genome Biology*, 13(6), R56. doi:10.1186/gb-2012-13-6-r56
- 663 Boomsma, J. J., Jensen, A. B., Meyling, N. V., & Eilenberg, J. (2014). Evolutionary Interaction Networks  
664 of Insect Pathogenic Fungi. *Annu. Rev. Entomol*, 59, 467–85. doi:10.1146/annurev-ento-011613-  
665 162054
- 666 Bye, N. J., & Charnley, A. K. (2008). Regulation of cuticle-degrading subtilisin proteases from the  
667 entomopathogenic fungi, *Lecanicillium* spp: Implications for host specificity. *Archives of  
668 Microbiology*, 189(1), 81–92. doi:10.1007/s00203-007-0296-8
- 669 Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B. *et al.* (2008). MAKER: An easy-to-  
670 use annotation pipeline designed for emerging model organism genomes. *Genome Research*, 18, 188–  
671 196. doi:10.1101/gr.6743907
- 672 Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly.  
673 *Bioinformatics*, 30(1), 31–37. doi:10.1093/bioinformatics/btt310
- 674 Cooper, D., & Eleftherianos, I. (2017). Memory and specificity in the insect immune system: Current  
675 perspectives and future challenges. *Frontiers in Immunology*, 8(539), 1-6.  
676 doi:10.3389/fimmu.2017.00539
- 677 Dongen, S. V. *Graph Clustering by Flow Simulation*. PhD thesis, University of Utrecht, May 2000.
- 678 de Bekker, C., Ohm, R. A., Evans, H. C., Brachmann, A., Hughes, D. P. (2017). Ant-infecting  
679 *Ophiocordyceps* genomes reveal a high diversity of potential behavioral manipulation genes and a  
680 possible major role for enterotoxins. *Scientific Reports*, 7, 12508. doi:10.1038/s41598-017-12863-w
- 681 de Bekker, C., Ohm, R. A., Loreto, R. G., Sebastian, A., Albert, I., Mellow, M. *et al.* (2015). Gene  
682 expression during zombie ant biting behavior reflects the complexity underlying fungal parasitic  
683 behavioral manipulation. *BMC Genomics*, 16(1), 1–23. doi:10.1186/s12864-015-1812-x

- 684 de Bekker, C., Quevillon, L. E., Smith, P. B., Fleming, K. R., Ghosh, D., Patterson, A. D., & Hughes, D. P.  
685 (2014). Species-specific ant brain manipulation by a specialized fungal parasite. *BMC Evolutionary*  
686 *Biology*, 14(166), doi:10.1186/s12862-014-0166-3. doi:10.1186/s12862-014-0166-3
- 687 De Fine Licht, H. H., Jensen, A. B., & Eilenberg, J. (2017). Comparative transcriptomics reveal host-specific  
688 nucleotide variation in entomophthoralean fungi. *Molecular Ecology*, 26(7), 2092–2110.  
689 doi:10.1111/mec.13863
- 690 Ekseth, O. K., Kuiper, M., & Mironov, V. (2014). OrthAgogue: An agile tool for the rapid prediction of  
691 orthology relations. *Bioinformatics*, 30(5), 734–736. doi:10.1093/bioinformatics/btt582
- 692 Evans, H. C., Elliot, S. L., & Hughes, D. P. (2011). Hidden diversity behind the zombie-ant fungus  
693 ophiocordyceps unilateralis: Four new species described from carpenter ants in Minas Gerais, Brazil.  
694 *PLoS ONE*. doi:10.1371/journal.pone.0017024
- 695 Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L. *et al.* (2016). The Pfam  
696 protein families database: Towards a more sustainable future. *Nucleic Acids Research*, 44(D1), D279–  
697 D285. doi:10.1093/nar/gkv1344
- 698 Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., & Gurr, S. J.  
699 (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, 484(7393), 186–194.  
700 doi:10.1038/nature10947
- 701 Funaki, M., Katagiri, H., Inukai, K., Kikuchi, M., Asano, T. Structure and function of phosphatidylinositol-  
702 3,4 kinase. *Cellular Signalling*, 12(3), 135-142. doi:10.1016/S0898-6568(99)00086-8
- 703 Gallo, A., Ferrara, M., & Perrone, G. (2013). Phylogenetic study of polyketide synthases and nonribosomal  
704 peptide synthetases involved in the biosynthesis of mycotoxins. *Toxins*, 5(4), 717–742.  
705 doi:10.3390/toxins5040717
- 706 Gao, Q., Jin, K., Ying, S. H., Zhang, Y., Xiao, G., Shang, Y. *et al.* (2011). Genome sequencing and  
707 comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M.*  
708 *acridum*. *PLoS Genetics*, 7(1), e1001264. doi:10.1371/journal.pgen.1001264
- 709 Ghanbarnia, K., Fudal, I., Larkan, N. J., Links, M. G., Balesdent, M. H., Profotova, B. *et al.* (2015). Rapid  
710 identification of the *Leptosphaeria maculans* avirulence gene AvrLm2 using an intraspecific  
711 comparative genomics approach. *Molecular Plant Pathology*, 16(7), 699-709. doi:10.1111/mpp.12228
- 712 Gillespie, J. P., Bailey, A. M., Cobb, B., & Vilcinskis, A. (2000). Fungi as elicitors of insect immune  
713 responses. *Archives of Insect Biochemistry and Physiology*, 44(2), 49–68. doi:10.1002/1520-  
714 6327(200006)44:2<49::AID-ARCH1>3.0.CO;2-F
- 715 Gladieux, P., Ropars, J., Badouin, H., Branca, A., Aguilera, G., De Vienne, D. M. *et al.* (2014). Fungal  
716 evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes.  
717 *Molecular Ecology*, 23(4), 753-773. doi:10.1111/mec.12631
- 718 Haas, H. (2012). Iron - a key nexus in the virulence of *Aspergillus fumigatus*. *Frontiers in Microbiology*,  
719 3(28), 1–10. doi:10.3389/fmicb.2012.00028
- 720 Hartmann, F. E., Rodríguez de la Vega, R. C., Brandenburg, J. T., Carpentier, F., Giraud, T. (2018) Species-  
721 specificity and phylogeographic structure of gene presence-absence polymorphism in castrating  
722 anther-smut fungi. *GBE*, *in press*.
- 723 Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M. *et al.* (2016). Ensembl

- 724 comparative genomics resources. *Database: The Journal of Biological Databases and Curation*,  
725 bav096, 1–17. doi:10.1093/database/bav096
- 726 Hu, X., Xiao, G., Zheng, P., Shang, Y., Su, Y., Zhang, X. *et al.* (2014). Trajectory and genomic determinants  
727 of fungal-pathogen speciation and host adaptation. *Proceedings of the National Academy of Sciences*  
728 *of the United States of America*, 111(47), 1–6. doi:10.1073/pnas.1412662111
- 729 Huang, J., Si, W., Deng, Q., Li, P., & Yang, S. (2014). Rapid evolution of avirulence genes in rice blast  
730 fungus *Magnaporthe oryzae*. *BMC Genetics*, 15(45), 1–10. doi:10.1186/1471-2156-15-45
- 731 Hughes, D. P., Andersen, S. B., Hywel-Jones, N. L., Himaman, W., Billen, J., & Boomsma, J. J. (2011).  
732 Behavioral mechanisms and morphological symptoms of zombie ants dying from fungal infection.  
733 *BMC Ecology*, 11(1), 13. doi:10.1186/1472-6785-11-13
- 734 Hughes, D. P., Araújo, J. P. M., Loreto, R. G., Quevillon, L., de Bekker, C., Evans, H. C. (2016). From so  
735 simple a beginning: the evolution of behavioral manipulation by fungi. *Advances in Genetics*, 94, 437-  
736 469. doi:[10.1016/bs.adgen.2016.01.004](https://doi.org/10.1016/bs.adgen.2016.01.004)
- 737 Ina, Y. (1996). Pattern of synonymous and nonsynonymous substitutions: An indicator of mechanisms of  
738 molecular evolution. *Journal of Genetics*, 75(1), 91. doi:10.1007/BF02931754
- 739 Jarrold, S. L., Moore, D., Potter, U., & Charnley, A. K. (2007). The contribution of surface waxes to pre-  
740 penetration growth of an entomopathogenic fungus on host cuticle. *Mycological Research*, 111(2),  
741 240–249. doi:10.1016/j.mycres.2006.10.007
- 742 Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C. *et al.* (2014). InterProScan 5: Genome-  
743 scale protein function classification. *Bioinformatics*, 30(9), 1236–1240.  
744 doi:10.1093/bioinformatics/btu031
- 745 Juárez-Vázquez, A. L., Edirisinghe, J. N., Verduzco-Castro, E. A., Michalska, K., Wu, C., Noda-García, L.  
746 *et al.* (2017). Evolution of substrate specificity in a retained enzyme driven by gene loss. *eLife*, 6,  
747 e22679. doi:10.7554/eLife.22679
- 748 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference  
749 resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), D457–D462.  
750 doi:10.1093/nar/gkv1070
- 751 Keller, O., Kollmar, M., Stanke, M., & Waack, S. (2011). A novel hybrid gene prediction method employing  
752 protein multiple sequence alignments. *Bioinformatics*, 27(6), 757–763.  
753 doi:10.1093/bioinformatics/btr010
- 754 Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., Fedorova, N. D. (2010).  
755 SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology*,  
756 47(9), 736-741. doi:10.1016/j.fgb.2010.06.003.
- 757 Kim, K.T., Jeon, J., Choi, J., Cheong, K., Song, H., Choi, G. *et al.* (2016) Kingdom-wide analysis of fungal  
758 small secreted proteins (SSPs) reveals their potential role in host association. *Frontiers in Plant*  
759 *Science*, 7, 186. doi: [10.3389/fpls.2016.00186](https://doi.org/10.3389/fpls.2016.00186).
- 760 Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press.
- 761 Kobmoo, N., Mongkolsamrit, S., Tasanathai, K., Thanakitpipattana, D., & Luangsa-Ard, J. J. (2012).  
762 Molecular phylogenies reveal host-specific divergence of *Ophiocordyceps unilateralis sensu lato*  
763 following its host ants. *Molecular Ecology*, 21(12), 3022-3031. doi:10.1111/j.1365-

- 764 294X.2012.05574.x
- 765 Kobmoo, N., Mongkolsamrit, S., Wutikhun, T., Tasanathai, K., Khonsanit, A., Thanakitpipattana, D., &  
766 Luangsa-Ard, J. J. (2015). New species of *Ophiocordyceps unilateralis*, an ubiquitous pathogen of ants  
767 from Thailand. *Fungal Biology*, *119*(1), 44-52. doi:10.1016/j.funbio.2014.10.008
- 768 Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 59. doi:10.1186/1471-2105-5-59
- 769 Kurtz, J., Schulenburg, H., & Reusch, T. B. (2016). Host–parasite coevolution—Rapid reciprocal adaptation  
770 and its genetic basis. *Zoology*, *119*, 241–243. doi:10.1016/j.zool.2016.06.011
- 771 Langner, T., & Göhre, V. (2016). Fungal chitinases: function, regulation, and potential roles in  
772 plant/pathogen interactions. *Current Genetics*, *62*(2), 243–254. doi:10.1007/s00294-015-0530-x
- 773 Lee K.-T., So Y.-S., Yang D. et al. (2016). Systematic functional analysis of kinases in the fungal pathogen  
774 *Cryptococcus neoformans*. *Nature Communications*, *7*, 12766. doi: 10.1038/ncomms12766
- 775 Lespinet, O., Wolf, Y. I., Koonin, E. V., & Aravind, L. (2002). The role of lineage-specific gene family  
776 expansion in the evolution of eukaryotes. *Genome Research*, *12*(7), 1048-1059.  
777 doi:10.1101/gr.174302
- 778 Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in  
779 novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research*, *33*(20), 6494–6506.  
780 doi:10.1093/nar/gki937
- 781 Luo, Z., Qin, Y., Pei, Y., Keyhani, N. O. (2014). Ablation of the *creA* regulator results in amino acid toxicity,  
782 temperature sensitivity, pleiotropic effects on cellular development and loss of virulence in the  
783 filamentous fungus *Beauveria bassiana*. *Environmental Microbiology*, *16*(4), 1122-1136.  
784 doi:10.1111/1462-2920.12352
- 785 Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens.  
786 *Nature Reviews Microbiology*. doi:10.1038/nrmicro.2017.76
- 787 Mondal, S., Baksi, S., Koris, A., & Vatai, G. (2016). Journey of enzymes in entomopathogenic fungi. *Pacific  
788 Science Review A: Natural Science and Engineering*, *18*(2), 85–99. doi:10.1016/j.psra.2016.10.001
- 789 Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A. et al. (2015). Gene-wide  
790 identification of episodic selection. *Molecular Biology and Evolution*, *32*(5), 1365–1371.  
791 doi:10.1093/molbev/msv035
- 792 Nielsen, R., & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and  
793 applications to the HIV-1 envelope gene. *Genetics*, *148*, 929–936.
- 794 Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto encyclopedia  
795 of genes and genomes. *Nucleic Acids Research*, *28*(1), 27-30. doi:10.1093/nar/27.1.29
- 796 Ohno, S. (1970). *Evolution by Gene Duplication*. Springer-Verlag.
- 797 Ortiz-Urquiza, A., & Keyhani, N. O. (2013). Action on the surface: Entomopathogenic fungi versus the  
798 insect cuticle. *Insects*, *4*(3), 357-374. doi:10.3390/insects4030357
- 799 Ortiz-Urquiza, A., Riveiro-Miranda, L., Santiago-Álvarez, C., & Quesada-Moraga, E. (2010). Insect-toxic  
800 secreted proteins and virulence of the entomopathogenic fungus *Beauveria bassiana*. *Journal of  
801 Invertebrate Pathology*, *105*(3), 270–8. doi:10.1016/j.jip.2010.07.003

- 802 Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: A pipeline to accurately annotate core genes in  
803 eukaryotic genomes. *Bioinformatics*, 23(9), 1061–1067. doi:10.1093/bioinformatics/btm071
- 804 Pattemore, J. A., Hane, J. K., Williams, A. H., Wilson, B. Al, Stodart, B. J., & Ash, G. J. (2014). The genome  
805 sequence of the biocontrol fungus *Metarhizium anisopliae* and comparative genomics of *Metarhizium*  
806 species, *BMC Genomics*, 15(660), 1–15. doi:10.1186/1471-2164-15-660
- 807 Pedrini, N., Ortiz-Urquiza, A., Huarte-Bonnet, C., Zhang, S., & Keyhani, N. O. (2013). Targeting of insect  
808 epicuticular lipids by the entomopathogenic fungus *Beauveria bassiana*: Hydrocarbon oxidation within  
809 the context of a host-pathogen interaction. *Frontiers in Microbiology*, 4(24), 1–18.  
810 doi:10.3389/fmicb.2013.00024
- 811 Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides  
812 from transmembrane regions. *Nature Methods*, 8(10), 785–786. doi:10.1038/nmeth.1701
- 813 Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F. E., Croll, D. (2017). Using  
814 Population and Comparative Genomics to Understand the Genetic Basis of Effector-Driven Fungal  
815 Pathogen Evolution. *Frontiers in Plant Science*, 8, 119. doi:[10.3389/fpls.2017.00119](https://doi.org/10.3389/fpls.2017.00119)
- 816 Pontoppidan, M. B., Himaman, W., Hywel-Jones, N. L., Boomsma, J. J., & Hughes, D. P. (2009).  
817 Graveyards on the move: The spatio-temporal distribution of dead ophiocordyceps-infected ants. *PLoS*  
818 *ONE*, 4(3), e4835. doi:10.1371/journal.pone.0004835
- 819 Poppe, S., Dorsheimer, L., Happel, P., Stukenbrock, E. H. (2015). Rapidly Evolving Genes Are Key Players  
820 in Host Specialization and Virulence of the Fungal Wheat Pathogen *Zymoseptoria tritici*  
821 (*Mycosphaerella graminicola*). *PLoS Pathogens*, 11(7), e1005055.
- 822 Rafiqi, M., Ellis, J. G., Ludowici, V. A., Hardham, A. R., & Dodds, P. N. (2012). Challenges and progress  
823 towards understanding the role of effectors in plant-fungal interactions. *Current Opinion in Plant*  
824 *Biology*, 15(4), 477-482. doi:10.1016/j.pbi.2012.05.003
- 825 Roman D., Dancis A., Anderson G., Klausner R. (1993). The fission yeast ferric reductase gene *frp1+* is  
826 required for ferric uptake and encodes a protein that is homologous to the gp91-phox subunit of the  
827 human NADPH phagocyte oxidoreductase. *Molecular and Cellular Biology*, 13(7), 4342-4350. doi:  
828 10.1128/MCB.13.7.4342
- 829 Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. P. (2011). MACSE: Multiple alignment of coding  
830 SEquences accounting for frameshifts and stop codons. *PLoS ONE*, 6(9), e22594.  
831 doi:10.1371/journal.pone.0022594
- 832 Rawlings, N. D., Barrett, A. J., Finn, R. (2016). Twenty years of the MEROPS database of proteolytic  
833 enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 44(D1), D343-D350. doi:  
834 10.1093/nar/gkv1118
- 835 Rouxel, T., & Balesdent, M. H. (2017). Life, death and rebirth of avirulence effectors in a fungal pathogen  
836 of Brassica crops, *Leptosphaeria maculans*. *New Phytologist*, 214(2), 526-532.  
837 doi:10.1111/nph.14411
- 838 Sakolrak, B., Blatrix, R., Sangwanit, U., Arnammart, N., Kobmoo, N. (2018). Experimental infection of the  
839 ant *Polyrhachis furcata* with *Ophiocordyceps* reveals specificity of behavioral manipulation. *Fungal*  
840 *Ecology*, 33, 122-124. doi: 10.1016/j.funeco.2018.03.001
- 841 Schirrmann, M. K., Zoller, S., Croll, D., Stukenbrock, E. H., Leuchtman, A., Fior, S. (2018). Genomewide

- 842 signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Molecular Ecology*,  
843 2018 (00), 1-17. doi:[10.1111/mec.14585](https://doi.org/10.1111/mec.14585)
- 844 Schrank, A., & Vainstein, M. H. (2010). *Metarhizium anisopliae* enzymes and toxins. *Toxicon*, 56(7), 1267-  
845 1274. doi:[10.1016/j.toxicon.2010.03.008](https://doi.org/10.1016/j.toxicon.2010.03.008)
- 846 Sexton, A. C., & Howlett, B. J. (2006). Parallels in fungal pathogenesis on plant and animal hosts.  
847 *Eukaryotic Cell*, 5(12), 1941-1949. doi:[10.1128/EC.00277-06](https://doi.org/10.1128/EC.00277-06)
- 848 Shah, F. A., Wang, C. S., & Butt, T. M. (2005). Nutrition influences growth and virulence of the insect-  
849 pathogenic fungus *Metarhizium anisopliae*. *FEMS Microbiology Letters*, 251(2), 259–266.  
850 doi:[10.1016/j.femsle.2005.08.010](https://doi.org/10.1016/j.femsle.2005.08.010)
- 851 Shang, Y., Xiao, G., Zheng, P., Cen, K., Zhan, S., & Wang, C. (2016). Divergent and Convergent Evolution  
852 of Fungal Pathogenicity. *Genome Biology and Evolution*, 8(5), 1374–1387. doi:[10.1093/gbe/evw082](https://doi.org/10.1093/gbe/evw082)
- 853 Singh, A., & Poeta, M. D. (2011). Lipid signalling in pathogenic fungi. *Cellular Microbiology*, 13(2), 177-  
854 185. doi:[10.1111/j.1462-5822.2010.01550.x](https://doi.org/10.1111/j.1462-5822.2010.01550.x)
- 855 Singh, D., Son, S. Y., & Lee, C. H. (2016). Perplexing metabolomes in fungal-insect trophic interactions:  
856 A Terra incognita of mycobiococontrol mechanisms. *Frontiers in Microbiology*, 7, 1678.  
857 doi:[10.3389/fmicb.2016.01678](https://doi.org/10.3389/fmicb.2016.01678)
- 858 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large  
859 phylogenies. *Bioinformatics*, 30(9), 1312–1313. doi:[10.1093/bioinformatics/bt1446](https://doi.org/10.1093/bioinformatics/bt1446)
- 860 Stukenbrock, E. H., Bataillon, T., Dutheil, J. Y., Hansen, T. T., Li, R., Zala, M. *et al.* (2011). The making  
861 of a new pathogen: Insights from comparative population genomics of the domesticated wheat  
862 pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Research*, 21(12), 2157–  
863 2166. doi:[10.1101/gr.118851.110](https://doi.org/10.1101/gr.118851.110)
- 864 Sutak R., Lesuisse E., Tacherzy J., Richardson D. (2008) Crusade for iron: iron uptake in unicellular  
865 eukaryotes and its significance for virulence. *Trends in Microbiology*, 16(6), 261-268. doi:  
866 [10.1016/j.tim.2008.03.005](https://doi.org/10.1016/j.tim.2008.03.005)
- 867 Swanson, W. J., Nielsen, R., & Yang, Q. (2003). Pervasive adaptive evolution in mammalian fertilization  
868 proteins. *Molecular Biology and Evolution*, 20(1), 18–20.  
869 doi:[10.1093/oxfordjournals.molbev.a004233](https://doi.org/10.1093/oxfordjournals.molbev.a004233)
- 870 Thompson, S. N. (2003). Trehalose - The Insect “Blood” Sugar. *Advances in Insect Physiology*, 31, 205–  
871 285. doi:[10.1016/S0065-2806\(03\)31004-5](https://doi.org/10.1016/S0065-2806(03)31004-5)
- 872 Wang, C., & Feng, M. G. (2014). Advances in fundamental and applied studies in China of fungal biocontrol  
873 agents for use against arthropod pests. *Biological Control*, 68(1), 129–135.  
874 doi:[10.1016/j.biocontrol.2013.06.017](https://doi.org/10.1016/j.biocontrol.2013.06.017)
- 875 Wang, C., & St. Leger, R. J. (2005). Developmental and Transcriptional Responses to Host and Nonhost  
876 Cuticles by the Specific Locust Pathogen *Metarhizium anisopliae* var. *acridum*. *Eukaryotic Cell*, 4(5),  
877 937–947. doi:[10.1128/EC.4.5.937-947.2005](https://doi.org/10.1128/EC.4.5.937-947.2005)
- 878 Wang, C., & Wang, S. (2017). Insect Pathogenic Fungi: Genomics, Molecular Interactions, and Genetic  
879 Improvements. *Annu. Rev. Entomol*, 62, 73–90. doi:[10.1146/annurev-ento-031616-035509](https://doi.org/10.1146/annurev-ento-031616-035509)
- 880 Wang, G., Liu, Z., Lin, R., Li, E., Mao, Z., Ling, J. *et al.* (2016). Biosynthesis of Antibiotic Leucinostatins

- 881 in Bio-control Fungus *Purpureocillium lilacinum* and Their Inhibition on *Phytophthora* Revealed by  
882 Genome Mining. *PLoS Pathogens*, 12(7), e1005685. doi:10.1371/journal.ppat.1005685
- 883 Wang S, Fang W, Wang C, St. Leger RJ (2011) Insertion of an Esterase Gene into a Specific Locust  
884 Pathogen (*Metarhizium acridum*) Enables It to Infect Caterpillars. *PLoS Pathog* 7(6): e1002097.  
885 doi:10.1371/journal.ppat.1002097
- 886 Wang, S., Leclercq, A., Pava-Ripoll, M., Fang, W., & St. Leger, R. J. (2009). Comparative genomics using  
887 microarrays reveals divergence and loss of virulence-associated genes in host-specific strains of the  
888 insect pathogen *metarhizium anisopliae*. *Eukaryotic Cell*, 8(6), 888–898. doi:10.1128/EC.00058-09
- 889 Wichadakul, D., Kobmoo, N., Ingsriswang, S., Tangphatsornruang, S., Chantasingh, D., Luangsa-Ard, J. J.,  
890 & Eurwilaichitr, L. (2015). Insights from the genome of *Ophiocordyceps polyrhachis-furcata* to  
891 pathogenicity and host specificity in insect fungi. *BMC Genomics*, 16, 881. doi:10.1186/s12864-015-  
892 2101-4
- 893 Wicker, T., Oberhaensli, S., Parlange, F., Buchmann, J. P., Shatalina, M., Roffler, S. *et al.* (2013). The  
894 wheat powdery mildew genome shows the unique evolution of an obligate biotroph. *Nature Genetics*,  
895 45(9), 1092–1096. doi:10.1038/ng.2704
- 896 Wiygul, G., & Sikorowski, P. P. (1986). The effect of staphylococcal enterotoxin B on pheromone  
897 production in fat bodies isolated from male boll weevils. *Journal of Invertebrate Pathology*, 47(1),  
898 116–119.
- 899 Wiygul, G., & Sikorowski, P. P. (1991). The effect of a heat-stable enterotoxin isolated from *Escherichia*  
900 *coli* on pheromone production in fat bodies isolated from male boll weevils. *Entomologia*  
901 *Experimentalis et Applicata*, 60(3), 305–308. doi:10.1111/j.1570-7458.1991.tb01551.x
- 902 Wongsa, P., Tasanatai, K., Watts, P., & Hywel-Jones, N. (2005). Isolation and in vitro cultivation of the  
903 insect pathogenic fungus *Cordyceps unilateralis*. *Mycological Research*, 109(8), 936–940.  
904 doi:10.1017/S0953756205003321
- 905 Xia, E.-H., Yang, D.-R., Jiang, J.-J., Zhang, Q.-J., Liu, Y., Liu, Y.-L. *et al.* (2017). The caterpillar fungus,  
906 *Ophiocordyceps sinensis*, genome provides insights into highland adaptation of fungal pathogenicity.  
907 *Scientific Reports*, 7, 1806. doi:10.1038/s41598-017-01869-z
- 908 Xiao, G., Ying, S., Zheng, P., Wang, Z.-L., Zhang, S., Xie, X.-Q. *et al.* (2012). Genomic perspectives on  
909 the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Scientific Reports*, 2, 483.
- 910 Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme  
911 evolution. *Mol Biol Evol*, 15(5), 568–573. doi:10.1093/oxfordjournals.molbev.a025957
- 912 Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and*  
913 *Evolution*, 24(8), 1586–1591. doi:10.1093/molbev/msm088
- 914 Yang, Z., & Nielsen, R. (1998). Synonymous and Nonsynonymous Rate Variation in Nuclear Genes of  
915 Mammals. *Journal of Molecular Evolution*, 46, 409–418.
- 916 Yang, Z., & Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under  
917 realistic evolutionary models. *Molecular Biology and Evolution*, 17(1), 32–43.  
918 doi:10.1093/oxfordjournals.molbev.a026236
- 919 Yang, Z., Nielsen, R., Goldman, N., & Krabbe Pedersen, A.-M. (2000). Codon-Substitution Models for  
920 Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics*, 155, 431–449.

- 921 Yang, Z., Wong, W. S. W., & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under  
922 positive selection. *Molecular Biology and Evolution*, 22(4), 1107-1118. doi:10.1093/molbev/msi097
- 923 Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y. (2012). dbCAN: a web resource for automated  
924 carbohydrate-active enzyme annotation. *Nucleic Acids Research*, 40(W1), 445-451.
- 925 Zhang, J., Zhang, Y.-P., & Rosenberg, H. F. (2002). Adaptive evolution of a duplicated pancreatic  
926 ribonuclease gene in a leaf-eating monkey. *Nature Genetics*, 30, 411–415. doi:10.1038/ng852
- 927 Zheng, P., Xia, Y., Xiao, G., Xiong, C., Hu, X., Zhang, S. *et al.* (2011). Genome sequence of the insect  
928 pathogenic fungus *Cordyceps militaris*, a valued traditional chinese medicine. *BMC Genome Biology*,  
929 12(11), R116.
- 930 Zhou, K., Huang, B., Zou, M., Lu, D., He, S., & Wang, G. (2015). Genome-wide identification of lineage-  
931 specific genes within *Caenorhabditis elegans*. *Genomics*, 106(4), 242–248.  
932 doi:10.1016/j.ygeno.2015.07.002

933

### 934 **Data Accessibility**

935 The *de-novo* assemblies of *Ophiocordyceps camponoti-leonardi* (NCBI Biosample  
936 SAMN07662903) and *O. camponoti-saundersi* (NCBI Biosample SAMN07662932) have been  
937 deposited with the NCBI as whole-genome shotgun (WGS) projects with accession numbers  
938 PDHP000000000 and PDHQ000000000, respectively. The *O. polyrhachis-furcata* genome was  
939 updated based on the improved assembly from this study (LKCN020000000).

940

### 941 **Author Contributions**

942 N.K., J.J.L. and T.G. designed the study. N.K. and N.A. conducted sampling and DNA extraction.  
943 N.K., D.W. and RCRSLV analyzed sequencing and comparative genomic data. N.K. and T.G.  
944 wrote the manuscript, with contributions from all the authors.

945

### 946 **Supporting information**

947 Table S1. Orthologous clusters of genes encoding heat-labile enterotoxins in species of the  
948 Hypocreales.

949 Table S2. List of genes unique to one of the three complexes: *Ophiocordyceps unilateralis*, *O.*  
950 *australis*, *O. subramanianii*.

951 Table S3. List of species-specific genes for the *Ophiocordyceps unilateralis* complex.

952 Table S4. List of species-specific genes for the *Ophiocordyceps australis* complex.

953 Table S5. List of the 5% of genes with the highest non-synonymous-to-synonymous substitution  
954 rate ratios (dN/dS) values, with functional annotations.

955 Table S6. Complete list of genes with significant likelihood ratio tests (LRTs) after false  
956 discovery rate (fdr) correction.

957 Table S7. Results of likelihood ratio tests (LRTs) for categories of genes thought to be involved  
958 in adaptation.

959 Table S8. Results of likelihood ratio tests (LRTs) for putative neuromodulator genes (after De  
960 Bekker et al. 2017).

961

962

## 963 **Figure captions**

964 **Figure 1.** Inference of orthologous groups: Venn diagram showing the number of orthologous  
965 groups common to and specific to species complexes and species a. between the three ant-infecting  
966 *Ophiocordyceps* species complexes used in this study; b. between the species in the *O. unilateralis*  
967 complex (OPF = *O. polyrhachis-furcata*, OCL = *O. camponoti-leonardi*, OCS = *O. camponoti-*

968 *saundersi*, OKi = *O. kimflemingiae*, OCR = *O. camponoti-rufipedis*), b. between the species in the  
969 *O. australis* complex (OAG = *O. australis* from Ghana, OAB = *O. australis* from Brazil).

970

971 **Figure 2.** The best maximum-likelihood tree based on 4,014 single-copy orthologous groups with  
972 bootstrap supports. The horizontal scale bar represents the branch length based on substitution  
973 rates.

974

975 **Figure 3.** Distribution of pairwise non-synonymous-to-synonymous substitution ratios (dN/dS) for  
976 the genes in all orthologous groups in the *Ophiocordyceps unilateralis s.l.* species complex. (a)  
977 Whole-genome dN/dS distributions, (b) Boxplots of pairwise dN/dS values for the whole genome  
978 (small secreted protein-coding genes or SSPs, in blue, versus non-SSPs, in red) and between  
979 different categories of genes suspected *a priori* to be involved in pathogenesis and virulence, i.e.,  
980 with the putative functions of enterotoxins, core proteins of secondary metabolism (SM), lipases,  
981 proteases (including subtilisin-like, trypsin and aspartyl proteases) and trehalases. The dotted line  
982 represents the mean dN/dS value for the whole genome (0.145).

983

984 **Figure 4.** Percentages of genes in various functional categories for which likelihood ratio tests  
985 (LRTs) for positive selection (M7 vs. M8 and M8a vs. M8) yielded significant results (false  
986 discovery rate-corrected *p*-value < 0.05). Proteases = subtilisin, trypsin and aspartyl proteases, SM  
987 = core genes of secondary metabolites. The total number of genes in each category is indicated  
988 above the bars.

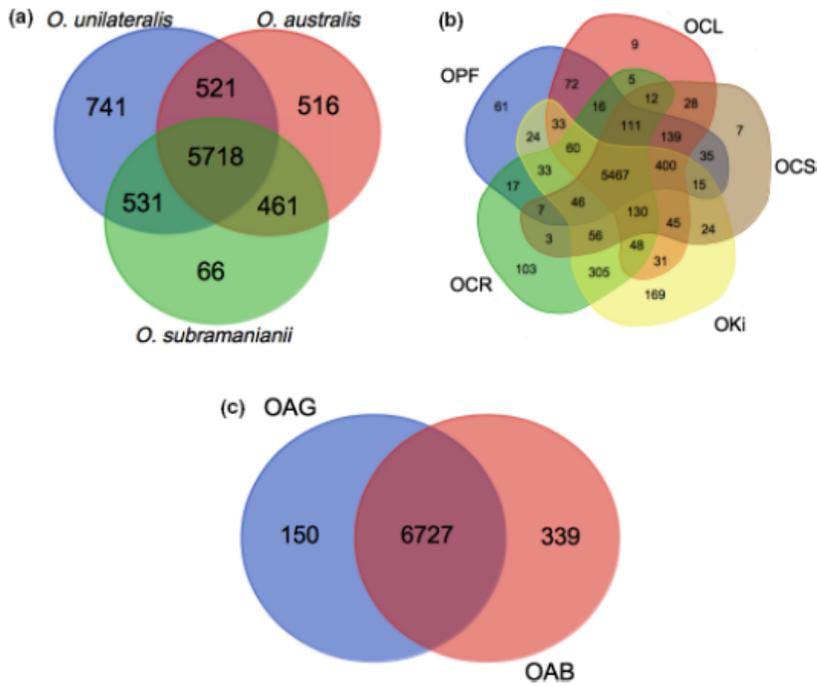
989

990 **Figure 5.** Homology of putative secondary metabolic gene clusters (SMGCs) with the core gene  
991 under positive selection according to log likelihood ratio tests (M7 vs. M8 models and M8a vs. M8  
992 models). The dashed lines indicate orthology between the putative polyketide synthase (PKS)-like  
993 core gene. The phylogenetic tree was inferred from Jaccard similarity indices between alignments  
994 of common gene domains within families. OCS = *Ophiocordyceps camponoti-saundersi*, OCL  
995 = *O. camponoti-leonardi*, OPF = *O. polyrhachis-furcata*, OKi = *O. kimflemingiae*, OCR = *O.*  
996 *camponoti-rufipedis*, OAG = *O. australis* from Ghana, OAB = *O. australis* from Brazil, OSS = *O.*  
997 *subramanianii*.

998

999 **Figure 6.** The best RAxML tree based on the GTRCAT model for the orthologous group  
1000 ORTHAgEnt13 of putative heat-labile enterotoxins in entomopathogenic and nematode-killing  
1001 fungi of the order Hypocreales. The numbers above the nodes are bootstrap supports. The numbers  
1002 below the branches are the ratios of non-synonymous-to-synonymous substitution rates (dN/dS).

1003

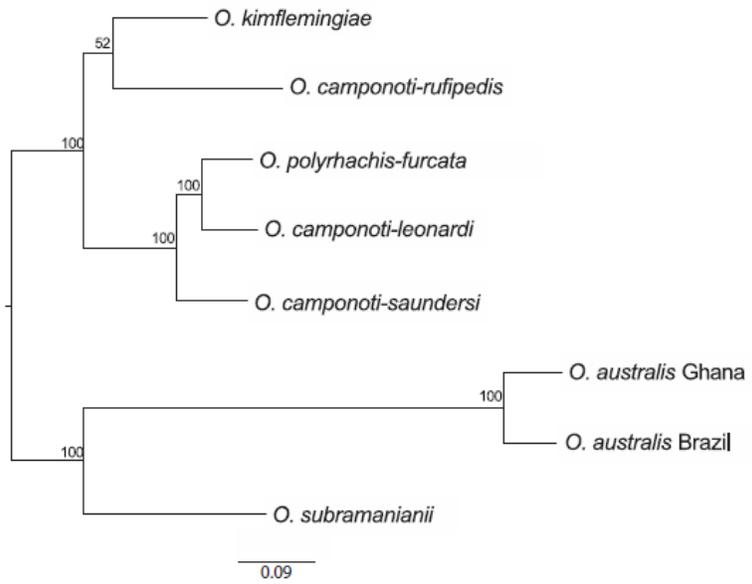


**FIGURE 1** Inference of orthologous groups: Venn diagram showing the number of orthologous groups common to and specific to species complexes and species a. between the three ant-infecting *Ophiocordyceps* species complexes used in this study; (b) between the species in the *O. unilateralis* complex (OPF = *O. polyrhachis-furcata*, OCL = *O. camponoti-leonardi*, OCS = *O. camponoti-saundersi*, OKi = *O. kimflemingiae*, OCR = *O. camponoti-rufipedis*), (c) between the species in the *O. australis* complex (OAG = *O. australis* from Ghana, OAB = *O. australis* from Brazil)

1004

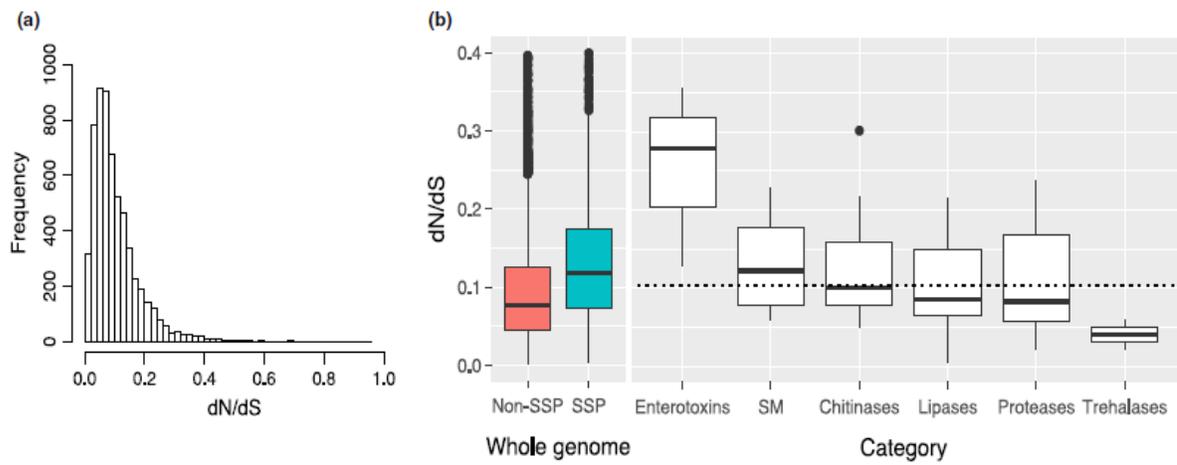
1005

**FIGURE 2** The best maximum-likelihood tree based on 4,014 single-copy orthologous groups with bootstrap supports. The horizontal scale bar represents the branch length based on substitution rates



1006

1007

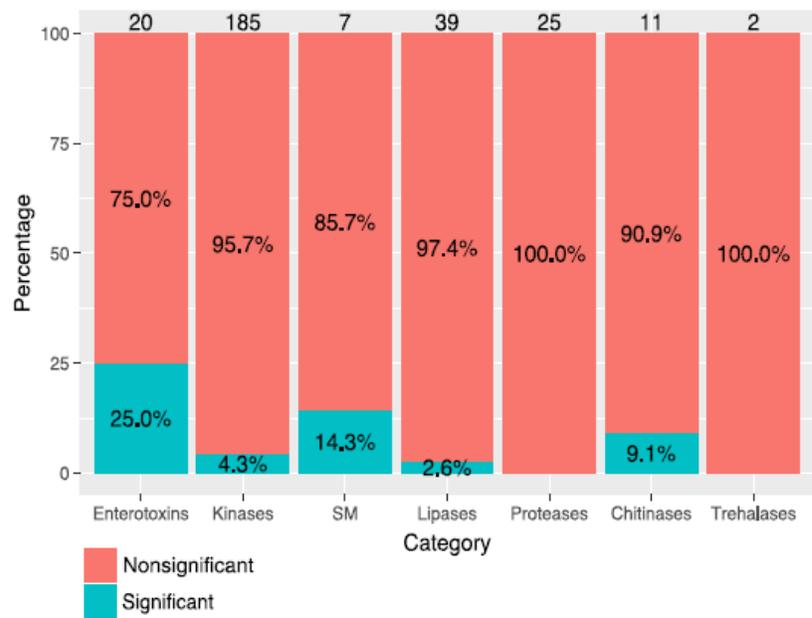


**FIGURE 3** Distribution of pairwise nonsynonymous-to-synonymous substitution ratios (dN/dS) for the genes in all single-copy orthologous groups with at least four species of ant-infecting *Ophiocordyceps* represented. (a) Whole-genome dN/dS distributions, (b) Boxplots of pairwise dN/dS values for the whole genome (small secreted protein-coding genes or SSPs, in blue, vs. non-SSPs, in red) and between different categories of genes suspected a priori to be involved in pathogenesis and virulence, that is, with the putative functions of enterotoxins, core proteins of secondary metabolism (SM), lipases, proteases (including subtilisin-like, trypsin and aspartyl proteases) and trehalases. The dotted line represents the mean dN/dS value for the whole genome (0.145)

1008

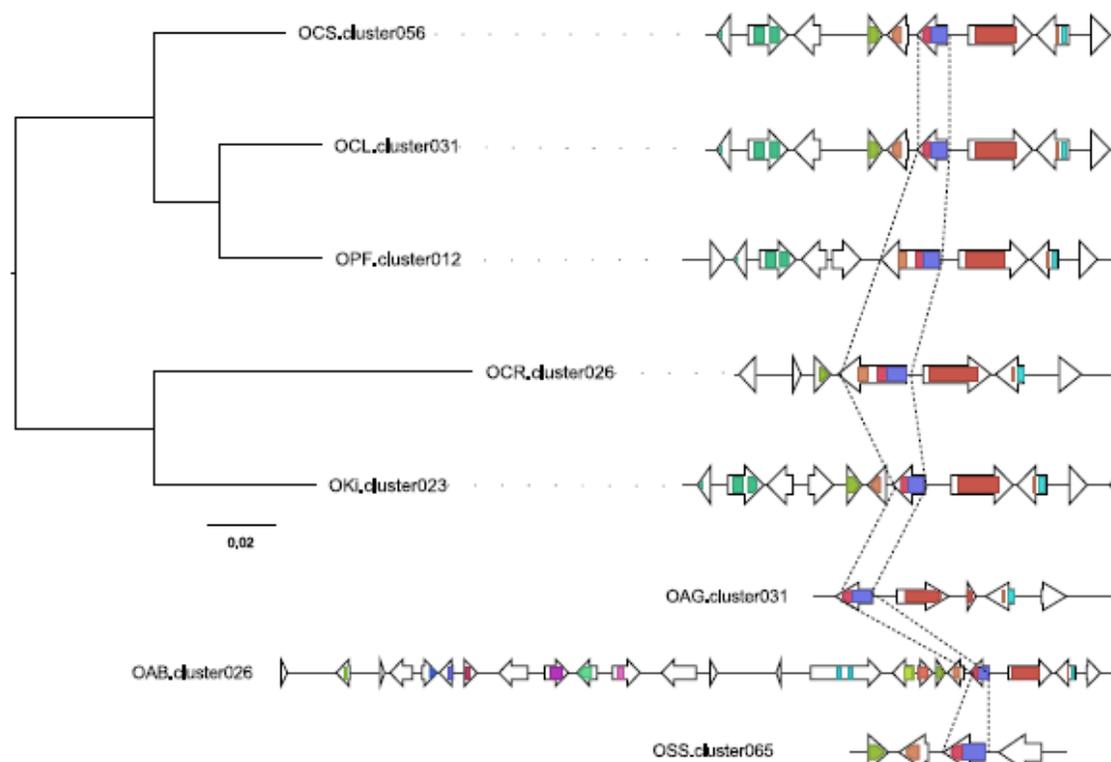
1009

**FIGURE 4** Percentages of genes in various functional categories for which likelihood ratio tests (LRTs) for positive selection (M7 vs. M8 and M8a vs. M8) yielded significant results (false-discovery rate-corrected  $p$ -value < 0.05). Proteases = subtilisin, trypsin and aspartyl proteases, SM = core genes of secondary metabolites. The total number of genes in each category is indicated above the bars



1010

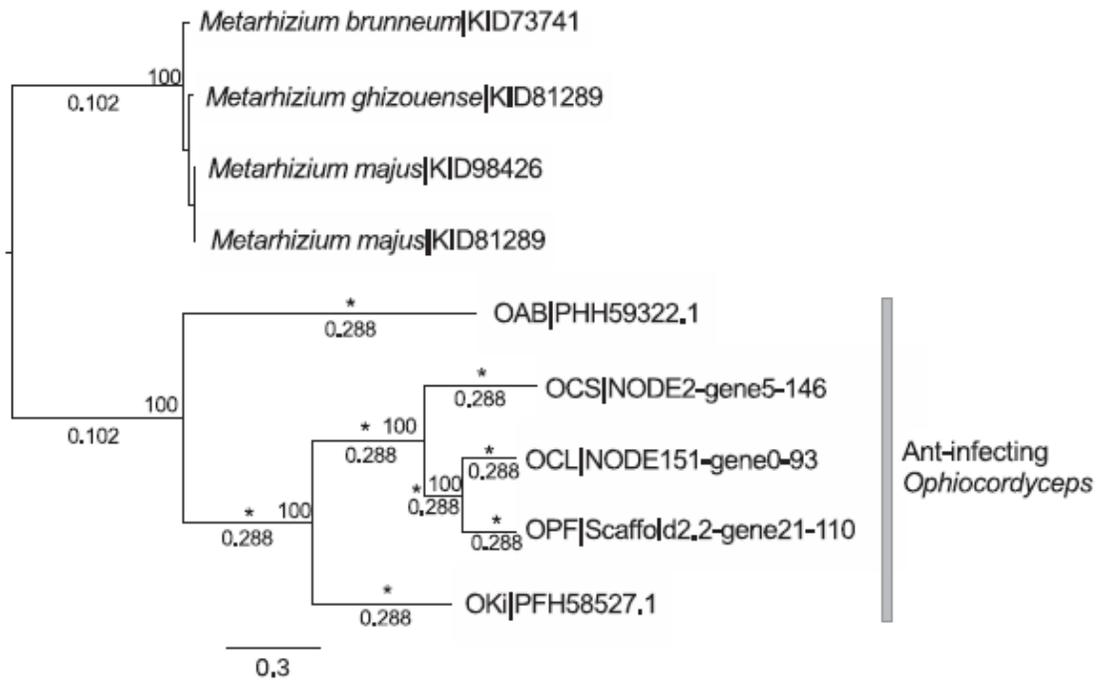
1011



**FIGURE 5** Homology of putative secondary metabolic gene clusters (SMGCs) with the core gene under positive selection according to log-likelihood ratio tests (M7 vs. M8 models and M8a vs. M8 models). The dashed lines indicate orthology between the putative polyketide synthase (PKS)-like core gene. The phylogenetic tree was inferred from Jaccard similarity indices between alignments of common gene domains within families. OCS = *Ophiocordyceps camponoti-saundersi*, OCL = *O. camponoti-leonardi*, OPF = *O. polyrhachis-furcata*, OKI = *O. kimflemingiae*, OCR = *O. camponoti-rufipedis*, OAG = *O. australls* from Ghana, OAB = *O. australls* from Brazil, OSS = *O. subramanianii*

1012

1013



**FIGURE 6** The best RAxML tree based on the GTRCAT model for the orthologous group ORTHAgEnt13 of putative heat-labile enterotoxins in entomopathogenic and nematode-killing fungi of the order Hypocreales. The numbers above the nodes are bootstrap supports. The numbers below the branches are the ratios of nonsynonymous-to-synonymous substitution rates (dN/dS)