

# UC Riverside

## UC Riverside Previously Published Works

### Title

Comparative Genomics Reveals the Core Gene Toolbox for the Fungus-Insect Symbiosis

### Permalink

<https://escholarship.org/uc/item/1q9479qp>

### Journal

mBio, 9(3)

### ISSN

2161-2129

### Authors

Wang, Yan

Stata, Matt

Wang, Wei

et al.

### Publication Date

2018-07-05

### DOI

10.1128/mbio.00636-18

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, available at

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Peer reviewed



# Comparative Genomics Reveals the Core Gene Toolbox for the Fungus-Insect Symbiosis

 Yan Wang,<sup>a,b,c,d</sup> Matt Stata,<sup>a</sup> Wei Wang,<sup>a</sup>  Jason E. Stajich,<sup>c,d</sup> Merlin M. White,<sup>e</sup> Jean-Marc Moncalvo<sup>a,b</sup>

<sup>a</sup>Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada

<sup>b</sup>Department of Natural History, Royal Ontario Museum, Toronto, Ontario, Canada

<sup>c</sup>Department of Microbiology and Plant Pathology, University of California, Riverside, Riverside, California, USA

<sup>d</sup>Institute for Integrative Genome Biology, University of California, Riverside, Riverside, California, USA

<sup>e</sup>Department of Biological Sciences, Boise State University, Boise, Idaho, USA

**ABSTRACT** Modern genomics has shed light on many entomopathogenic fungi and expanded our knowledge widely; however, little is known about the genomic features of the insect-commensal fungi. Harpellales are obligate commensals living in the digestive tracts of disease-bearing insects (black flies, midges, and mosquitoes). In this study, we produced and annotated whole-genome sequences of nine Harpellales taxa and conducted the first comparative analyses to infer the genomic diversity within the members of the Harpellales. The genomes of the insect gut fungi feature low (26% to 37%) GC content and large genome size variations (25 to 102 Mb). Further comparisons with insect-pathogenic fungi (from both Ascomycota and Zoopagomycota), as well as with free-living relatives (as negative controls), helped to identify a gene toolbox that is essential to the fungus-insect symbiosis. The results not only narrow the genomic scope of fungus-insect interactions from several thousands to eight core players but also distinguish host invasion strategies employed by insect pathogens and commensals. The genomic content suggests that insect commensal fungi rely mostly on adhesion protein anchors that target digestive system, while entomopathogenic fungi have higher numbers of transmembrane helices, signal peptides, and pathogen-host interaction (PHI) genes across the whole genome and enrich genes as well as functional domains to inactivate the host inflammation system and suppress the host defense. Phylogenomic analyses have revealed that genome sizes of Harpellales fungi vary among lineages with an integer-multiple pattern, which implies that ancient genome duplications may have occurred within the gut of insects.

**IMPORTANCE** Insect guts harbor various microbes that are important for host digestion, immune response, and disease dispersal in certain cases. Bacteria, which are among the primary endosymbionts, have been studied extensively. However, fungi, which are also frequently encountered, are poorly known with respect to their biology within the insect guts. To understand the genomic features and related biology, we produced the whole-genome sequences of nine gut commensal fungi from disease-bearing insects (black flies, midges, and mosquitoes). The results show that insect gut fungi tend to have low GC content across their genomes. By comparing these commensals with entomopathogenic and free-living fungi that have available genome sequences, we found a universal core gene toolbox that is unique and thus potentially important for the insect-fungus symbiosis. This comparative work also uncovered different host invasion strategies employed by insect pathogens and commensals, as well as a model system to study ancient fungal genome duplication within the gut of insects.

**KEYWORDS** FISCoG, phylogenomics, Trichomycetes, Zoopagomycota, Zygomycota

Received 2 April 2018 Accepted 13 April 2018 Published 15 May 2018

**Citation** Wang Y, Stata M, Wang W, Stajich JE, White MM, Moncalvo J-M. 2018. Comparative genomics reveals the core gene toolbox for the fungus-insect symbiosis. *mBio* 9:e00636-18. <https://doi.org/10.1128/mBio.00636-18>.

**Editor** John W. Taylor, University of California, Berkeley

© Crown copyright 2018. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Yan Wang, [yanxw.wang@gmail.com](mailto:yanxw.wang@gmail.com).

Many microfungi have obligate symbiotic relationships with other eukaryotes (1–3). Insects, the most species-rich group of animals, are often associated with fungal symbionts (4, 5). Fungus-insect relationships have mostly been documented and studied for the disease-bearing behavior of the insects and effects of transmission to humans and livestock (6, 7). Several fungi are well known to cause insect disease or death (8–10), although there is a suite of complex interactions that range from parasitism to commensalism and mutualism, according to the results of feedback with their hosts (11).

Various symbiotic interactions between insects and fungi have been recorded. For example, Attine ant farm cultivars of fungi (2) and aphids acquire fungal genes to make their own carotenoids (12). On the other hand, fungi are able to consume insects and even control their behaviors by forming peripheral networks encircling the host muscle and using detailed mechanisms to secure their own dispersal success, including examples such as *Entomophthora muscae* found in house flies (13) and *Ophiocordyceps unilateralis* in arboreal ants (14, 15). Some fungi kill pest insects such as *Beauveria bassiana* (16), *Metarhizium* spp. (17), and *Smittium morbosum* (18), and in some cases they have inspired the development of multiple biocontrol pesticides that use the fungal products (3, 19, 20). Furthermore, complex interactions between fungi and insects, in the form of bioactive by-products, have long been considered for medicinal usages, such as *Ophiocordyceps sinensis* (synonym *Cordyceps sinensis*) (21, 22) and *Cordyceps militaris* (10).

Recent advances in next-generation sequencing (NGS) and bioinformatics techniques are enabling detailed examinations and searches for genetic elements that are responsible for fungus-insect association in a novel and direct way (23–25). de Bekker et al. have shown that the genetic tool used by *O. unilateralis* to manipulate the behavior of ant hosts is a compound made of ergot, alkaloids, polyketides, and nonribosomal peptides, which have effects on central nervous systems (26). Xiao et al. found that *B. bassiana* has active gene sets that work as sensors to detect various environments and host types (27). Comparisons of two closely related *Metarhizium* fungi, the locust-specific pathogen *M. acridum* and the generalist *M. robertsii* (synonym *M. anisopliae*) (28), revealed that both produce strikingly larger proportions of secreted proteins than other fungi but that the genome of *M. robertsii* includes expanded families of genes encoding proteases, chitinase, polyketide synthases, and nonribosomal peptide synthetases, in response to its wider need for cuticle degradation, detoxification, and toxin biosynthesis for various host types (29). Most of the available whole-genome sequences of insect-associated fungi are from pathogens (27, 29–31). However, to obtain a broader and fundamental understanding of the biology of insect-fungus symbioses, it is important to include fungal commensals as well. By comparing the genome sequences of fungal pathogens and commensals, we aim to identify the universal gene toolbox that is available to various fungal symbionts and that is presumably critical in establishing the symbiotic relationship with the insects.

Harpellales fungi obligately associate with gut linings of aquatic larvae or nymphs of insects (lower Diptera) (5, 32). It has been estimated that their symbiotic relationship has existed since the Permian period (about 270 million years ago) (33). The Harpellales fungi compose a traditional order of Trichomycetes (Zygomycota) (5) and have been recently classified under the new phylum of Zoopagomycota based on genome-scale data (34). Most species within the Harpellales are considered commensals, and they are better known for their harmonious partnerships with the insect hosts, mostly due to their efficient synchronized development (32). One exception would be the species *Smittium morbosum*, which was reported with a unique parasitic lifestyle and which can kill mosquito larvae by penetrating the host gut linings from the inside, anchoring to the exoskeleton, and preventing the host from molting (18). The parasitic stage is also known in the life history of a number of Harpellales genera (i.e., *Genistellospora*, *Harpella*, and *Pennella*) that associate with black flies (35). It has been suggested that certain circumstances (e.g., nutritional stress, pH pressure, etc.) are important for these

**TABLE 1** Genome features and statistics of the nine Harpellales taxa

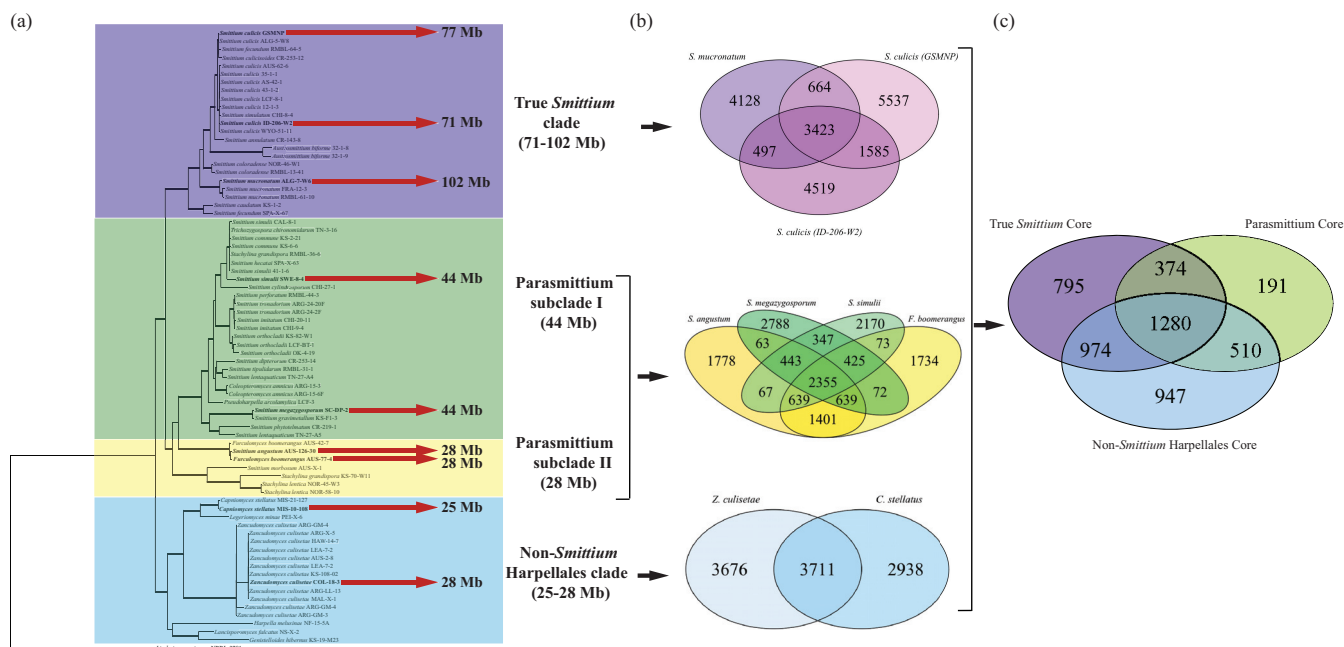
Strain	No. of scaffolds (>1 kb)	Genome size by scaffolds (Mb)	% CEGMA (248 in total)	GC ratio (%)	No. of predicted gene models	Repeat ratio (%)	SNP ratio (%)	NCBI accession no.
<i>Smittium culicis</i> GSMNP	6,137	77.12	97.98	28.61	11,209	3.34	0.45	LSSN00000000
<i>Smittium culicis</i> ID-206-W2	7,749	71.05	97.58	29.46	10,024	3.64	0.68	LSSM00000000
<i>Smittium mucronatum</i>	7,797	102.35	93.55	26.05	8,712	2.94	0.75	LSSL00000000
<i>Zancudomyces culisetae</i>	1,954	28.70	92.74	35.52	7,387	4.29	0.64	LSSK00000000
<i>Smittium megazygosporum</i>	3,927	43.63	96.77	32.49	7,132	4.60	0.41	MBFS00000000
<i>Smittium angustum</i>	1,283	28.05	99.19	32.40	7,385	1.60	0.43	MBFU00000000
<i>Furculomyces boomerangus</i>	1,312	28.13	99.19	32.37	7,338	1.58	0.43	MBFT00000000
<i>Capniomyces stellatus</i>	72	24.85	97.18	37.82	6,649	4.54	0.06	LUVW00000000
<i>Smittium simuli</i>	1,131	43.91	94.76	28.38	6,519	3.38	0.58	MBFR00000000

commensal fungi to shift the relationship with insects between ally and enemy boundaries in both the short and long terms (36–38).

We have sequenced and annotated four new Harpellales genomes in this study. We used these data and five additional genomes to compare a broad range of Harpellales lineages (39, 40). The objectives were to describe the genome features of Harpellales commensals, to conduct analyses of comparative genomics among the nine Harpellales taxa, and to identify the fungus-insect symbiotic core gene (FISCoG) set by further comparing these commensals with the entomopathogenic fungi and free-living relatives. The results of those analyses and the identification of a FISCoG toolbox will help to improve understanding of the basic biology of and the evolutionary relationships between fungi and insects. These resources will be fundamental to efforts to discover the genetic boundaries of symbiosis among parasitism, mutualism, and commensalism.

## RESULTS

**General genome features and comparative genomics of Harpellales.** Four Harpellales species genomes were sequenced to 165× to 230× coverage and assembled into 1,131 to 3,927 scaffolds (>1 kb). Genome sizes were estimated from the assemblies. There were two classes of genome sizes: one class was 44 Mb (*Smittium simuli* and *S. megazygosporum*), and one was approximately 28 Mb (*S. angustum* and *Furculomyces boomerangus*). The GC (guanine-cytosine) ratios across the whole genomes range from 28% to 33%. The core eukaryotic gene mapping approach (CEGMA) recovered more than 94% of core eukaryotic genes in all four genome assemblies. The *ab initio* protein-coding gene prediction identified 6,519 to 7,385 genes. Detailed genome features and statistics for these four taxa are listed in Table 1 along with those of previously sequenced members of this clade. The genome size variation among the nine Harpellales taxa presented an integer-multiple pattern among lineages of the 5-gene phylogenetic tree (Fig. 1a; reconstructed from reference 41). The names of “true *Smittium*,” “Parasmittium,” and “non-*Smittium* Harpellales” are labels adapted from a previous study (41) as well to refer to the major divergences among the members of this section of the Harpellales tree of life (32, 41, 42) (Fig. 1a). Venn diagrams show the number of shared and uncommon genes for subclades (Fig. 1b) and for Harpellales as a whole group (Fig. 1c). Based on these data, 3,423, 2,355, and 3,711 protein-coding genes are shared within the “true *Smittium*,” “Parasmittium,” and “non-*Smittium* Harpellales” subclades, respectively, and 1,280 of the genes were recovered from all three subclades and thus are proposed to be the genes that are common among the members of Harpellales (“Harpellales common genes”) (Fig. 1c). By allowing up to three missing taxa (i.e., by requiring any given gene to be present among six of the nine Harpellales taxa at minimum), the number of Harpellales common genes was found to have increased to 3,094. The genome-level relatedness of the nine Harpellales taxa was shown by aligning and plotting the scaffolds against each other (Fig. 2). The most closely related pairs were suggested by both high-density dot plots and dot alignment proximity to the diagonal curve. The detailed comparison statistics listed in the circles in Fig. 2 indicate both the number and identity level of the matches. Seven pairs were

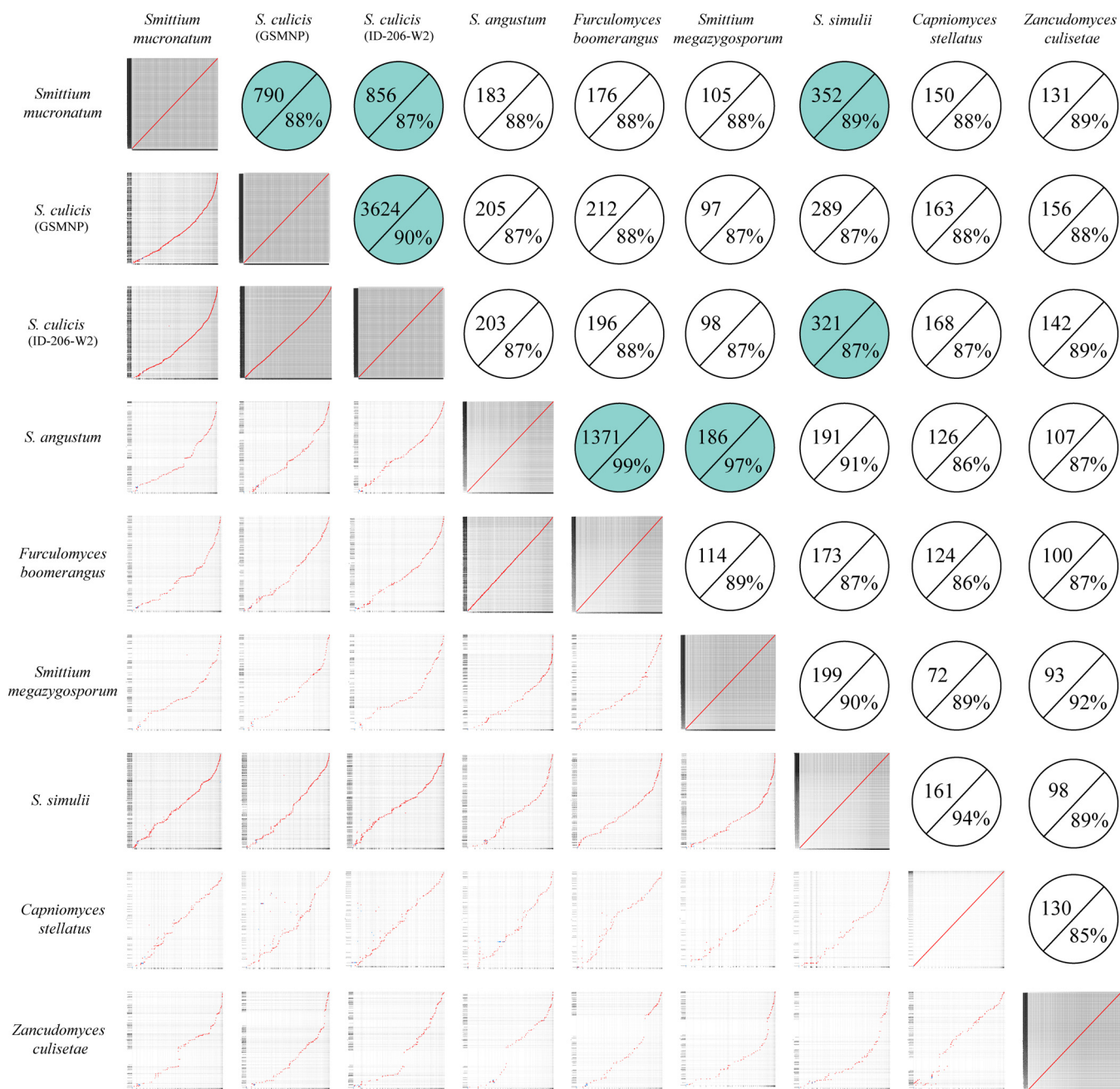


**FIG 1** Genome size variation across recognized subclades and Venn diagrams showing homologues across the nine genome-sequenced members of the Harpellales. (a) Harpellales phylogenetic tree based on 5 genes (reconstructed using the data set from reference 41 by adding the strains of *S. culicis* ID-206-W2 and *Capniomyces stellatus* MIS-10-108). Branches indicated in bold are considered strongly supported, with Bayesian posterior probability (BPP) values of >95% and maximum-likelihood bootstrap probability (MLBP) values of >0.70. Genome sizes of the recently sequenced 9 taxa were mapped with subclade information (non-*Smittium* Harpellales, true *Smittium*, Parasmitium subclades I and II). (b) Venn diagrams for each subclade derived from analysis of reciprocal best matches of protein-coding genes, showing relatedness and homologous comparisons across the subclades of Harpellales. (c) Identification of the Harpellales feature genes. Clade-specific genes were also identified in comparisons of the three major subclades of Harpellales.

highlighted since their matched bases were longer than 100 kb and are thus suggested to be closely related taxa among the total of 36 comparison pairs (Fig. 2). All three members of the true *Smittium* clade exhibit close relationships with each other, and the two species from the Parasmitium subclade II were revealed to feature the highest identity level (99%). Interestingly, the two individuals from the Parasmitium subclade I presented divergent affinity results; *S. megazygosporum* shows high identity to one member of the Parasmitium subclade II (*S. angustum*), and *S. similia* is more similar to the two members of the true *Smittium* clade (*S. mucronatum* and *S. culicis* ID-206-W2) than to members of the other taxa included here.

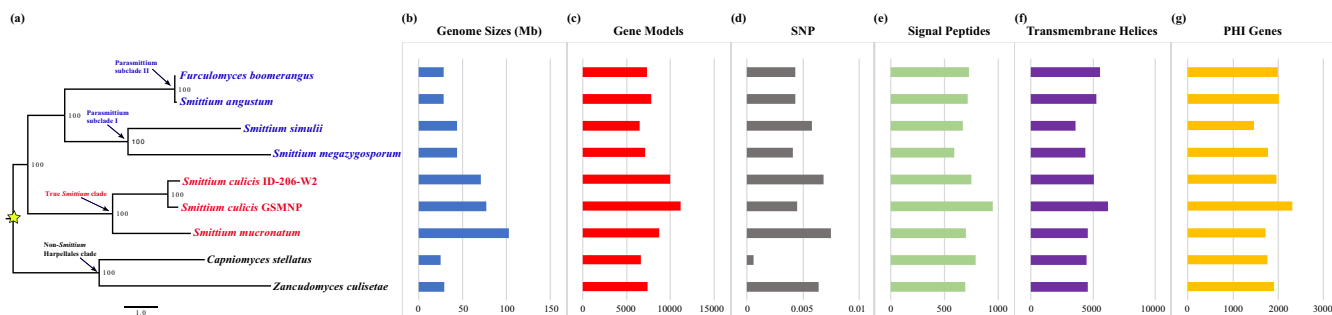
**Phylogenomics of the Harpellales.** The Harpellales phylogenomic tree was reconstructed based on 1,241 homologous protein sequences (Fig. 3a). The tree provides a robust phylogeny of the Harpellales and supports the topology proposed earlier on the basis of analyses performed using five genes (41). With the genomic information mapped to the phylogenomic tree (Fig. 3b to g), the data clearly show a pattern of lineage-specific genome size variation—the two lineages depicted in Fig. 3a (Parasmitium subclade II and Non-*Smittium* Harpellales clade), including *Capniomyces stellatus*, *Furculomyces boomerangus*, *Smittium angustum*, and *Zancudomyces culisetae*, have genome sizes between 25 and 28 Mb, whereas the taxa among other lineages show the genome sizes close to intervals in multiples of 25 Mb. This pattern implies that whole-genome duplication (WGD) might have occurred more than once and that the initial event might have been close to the time of divergence between the *Smittium* and non-*Smittium* clades (Fig. 3a; labeled with a star). Regardless of their genome size differences, there is no striking distinction with respect to the numbers of gene models, signal peptide genes, transmembrane helix genes, and pathogen-host interaction genes (Fig. 3b and c and 3e to g). Among the four taxa whose genome sizes are close to 25 Mb, we identified 1,071 single-copy genes. Almost half of them (582 of the 1,071) are found with multiple copies in at least one of the remaining five taxa (genome size, ≥44 Mb) (see Table S1 in the supplemental material). In total, 527 phylogenetic trees





**FIG 2** Whole-genome dot plots among the nine Harpellales genome sequences (centered diagonally, from lower left corner; determined using MUMmer plotting). Circles with detailed outputs of comparisons with exact match numbers (left) and the identity level of the matches (right) (centered diagonally, from upper right corner) are shown. Light blue circles indicate the pairs with matched regions longer than 100 kb. A default minimum cluster length of 65 bp was used for the comparison pairs, except for *S. mucronatum* and *S. culicis* (GSMNP) (75 bp), *S. mucronatum* and *S. culicis* (ID-206-W2) (75 bp), *S. culicis* (GSMNP) and *S. culicis* (ID-206-W2) (350 bp), *S. angustum* and *Furculomyces boomerangus* (4,000 bp), and *S. angustum* and *S. simulii* (70 bp), as well as *S. simulii* and *Capniomyces stellatus* (74 bp). Self-comparisons were performed using a minimum cluster length of 500 bp.

were reconstructed based on the well-recovered “single-copy” genes (55 of the 582 homologous groups were disregarded due to short alignment length) to infer their phylogenetic relationships (see Fig. S1 in the supplemental material). Most of the 527 homologous copies from the same taxon group with each other, and none of them exhibits divergence with statistical support. This result implies that the identified genome-size duplications may represent independent events and might be less likely to be due to interspecies hybridizations. The sites with single nucleotide polymorphisms (SNPs) were persistently found by the use of the binary version of the sequence

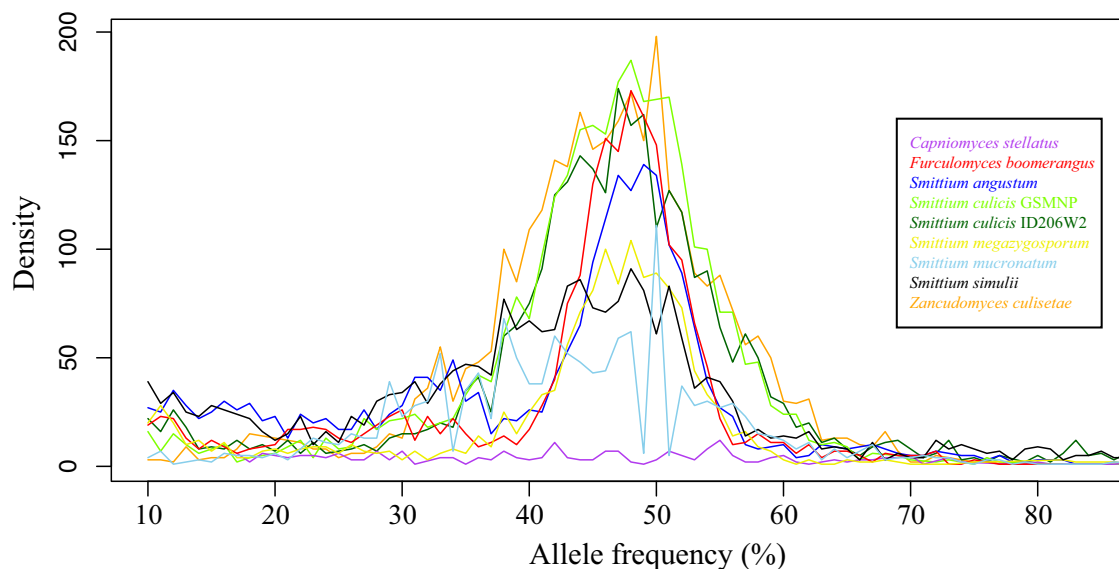


**FIG 3** Phylogenomics and genome statistics of Harpellales. (a) The phylogenetic tree was reconstructed based on a concatenated alignment of 1,241 homologues using IQ-TREE v1.5.3 for maximum-likelihood analysis and ultrabootstrap analysis performed with 1,000 replications (true *Smittium* and *Paramyttium* members are colored in red and blue, respectively, while non-*Smittium* Harpellales taxa are in black). (b to g) Genomic feature of the Harpellales in the order of genome sizes (b), predicted gene models (c), single nucleotide polymorphism sites (d), signal peptide numbers (e), transmembrane helix numbers (f), and numbers of genes that have homologues in the Pathogen-Host Interaction (PHI) database (g).

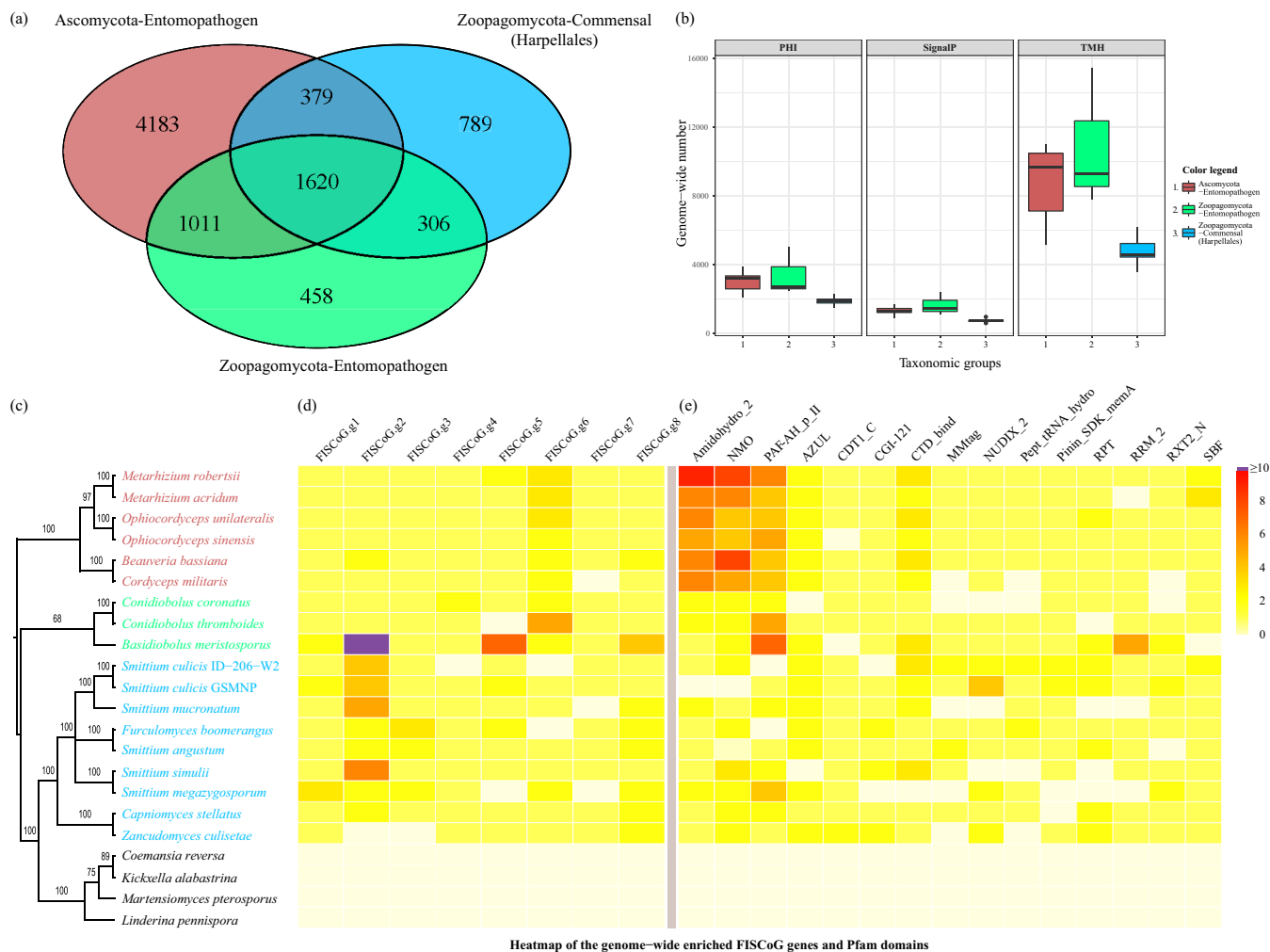
alignment map files of the Harpellales genomes. The percentages of the SNP sites across the eight genomes of Harpellales species (except that of *Capniomyces stellatus*) range from 0.41% to 0.75%, while that of *Capniomyces stellatus* has a much lower value (0.06%; Table 1) (Fig. 3d). These ratios (except that corresponding to *C. stellatus*) are comparable to those of the following diploid organisms from various kingdoms: 0.19% in *Pisum sativum* L. (43), 0.32% in *Homo sapiens* (44), 0.35% in *Edhazardia aedis* (31), and 1.2% in *Candida albicans* (45). The genome-wide allele frequency plot for the single-copy orthologs additionally supports this peculiar finding and suggests that *Capniomyces stellatus* is likely haploid whereas the other eight Harpellales may be all diploid, with a noticeable peak at the position of 50% allele frequency (Fig. 4).

**Fungus-insect symbiotic core gene/domain toolbox.** Representatives of the Ascomycota entomopathogenic fungi (*Beauveria bassiana*, *Cordyceps militaris*, *Metarhizium acridum*, *Metarhizium robertsii*, *Ophiocordyceps sinensis*, and *Ophiocordyceps uni-*

### Allele frequency plot



**FIG 4** Genome-wide allele frequency distribution among the single-copy orthologs of the nine Harpellales taxa (applied to 460 to 484 transcripts individually, allowing one taxon to be missing from among the nine). Eight of the nine taxa (except *C. stellatus*) exhibited a cumulative percentage around the 50% position, suggesting a disomic tendency of the genomes. Specifically, 1,398 to 3,235 variable nucleotide positions were analyzed and plotted for the eight Harpellales (allele frequency interval of 10% to 90%) but only 238 for the *C. stellatus*.



**FIG 5** Comparative genomics between the entomopathogenic fungi (Ascomycota in red and Zoopagomycota in green) and insect commensals of the Harpellales (in blue). (a) Venn diagram derived from interphylum homologues with the aim to sort out fungus-insect symbiotic core genes (FISCOGs), using pathogenic representatives both from Ascomycota and Zoopagomycota and commensals from Harpellales. (b) Box plot comparisons of genome-wide PHI genes, signal peptides, and transmembrane helices among the three groups. (c) Cladogram exhibiting the phylogenetic relationship of the included taxa based on 29 shared single-copy genes. (d) Heat map enrichment of the FISCOG toolbox among the insect-associated fungi (analyzed by removing the 1,612 false-positive hits with non-insect-associated Zoopagomycota genomes from those corresponding to the 1,620 shared genes in panel a). (e) Heat map comparison showing the enrichment pattern of genome-wide Pfam domains (detailed information for the fungus-insect symbiotic core domains is listed in Table S3).

*lateralis*) share 7,193 protein-coding genes (with no more than 2 of the 6 taxa allowed to be missing from those sharing the genes) (26, 27, 29, 46–48). Zoopagomycota entomopathogenic representatives (*Basidiobolus meristosporus*, *Conidiobolus coronatus*, and *Conidiobolus thromboides*) share 3,395 protein-coding genes (with no taxa allowed to be missing) (Fig. S2) (49–51). By comparing the 3,094 Harpellales common genes (with no more than 3 of the 9 taxa allowed to be missing) with both the Ascomycota and Zoopagomycota entomopathogenic core genes, we identified 1,620 protein-coding genes that are shared among both the insect commensals and pathogens (Fig. 5a). In screening out universal genes that are not specific to the fungus-insect associations, we found that 1,612 of them (among 1,620 protein-coding genes) had similarity hits (with an E-value of less than  $1E^{-5}$ ) in at least one of the negative-control proteomes among free-living Zoopagomycota fungi (*Coemansia reversa*, *Kickxella alabastrina*, *Linderina pennisporea*, and *Martensiomycetes pterosporus*) (49, 50). Only eight genes were found to be unique to these insect-associated fungi. Here we propose these eight protein-coding genes as candidates for the fungus-insect symbiotic core gene (FISCOG) toolbox. Functional annotation indicates that these FISCOGs are mainly re-



**TABLE 2** Information and detailed annotations of the FISCoG toolbox<sup>a</sup>

FISCoG	Description	GO name(s) (with GO-Slim)	Homologue with demonstrated function(s)	Subcellular location prediction (probability)	PHI hit(s) (1E−3)
FISCoG.g1	Peroxisomal NADH pyrophosphatase NUDT12	F, hydrolase activity	Regulation of concn of peroxisomal nicotinamide nucleotide cofactors required for oxidative metabolism (72)	Peroxisomal (0.65/1.00)	N/A
FISCoG.g2	Fasciclin domain-containing protein (beta-Ig-H3)	C, fungal vacuole membrane, extracellular space, membrane, integral component of membrane; P, macroautophagy	Cell adhesion protein (58–61)	Extracellular (0.80/1.00)	PHI:4231
FISCoG.g3	Acyl-CoA N-acyltransferase	F, N-acetyltransferase activity (transferring acyl groups)	Involvement in intestinal colonization and systemic infection (57)	Cytoplasmic (0.96/1.00)	PHI:5571
FISCoG.g4	Nuclear movement protein NudC	N/A	Nuclear migration and distribution (52)	Cytoplasmic (0.73/1.00)	PHI:2524
FISCoG.g5	F-box/LRR-repeat protein 2	F, protein kinase activity, ATP binding, kinase activity, ligase activity; P, protein phosphorylation, phosphorylation	Ubiquitin ligase complex F-box protein that mediates the ubiquitination and subsequent proteasomal degradation of target proteins (69, 70)	Cytoplasmic (0.81/1.00)	PHI:733; PHI:6752
FISCoG.g6	Platelet-activating factor acetylhydrolase	F, 1-alkyl-2-acetylglycerophosphocholine esterase activity; P, lipid catabolic process	Enzyme that catabolizes platelet-activating factor (65, 66)	Cytoplasmic (0.68/1.00)	N/A
FISCoG.g7	Putative SET-like protein	N/A	Related to growth control, gene transcription, and chromatin structure (73)	Nuclear (0.88/1.00)	N/A
FISCoG.g8	RNA-binding protein Nrd1	F, RNA binding; C, cytoplasm; P, negative regulation of conjugation with cellular fusion by regulation of transcription from RNA polymerase II promoter (reproduction, biosynthetic process, cellular nitrogen compound metabolic process)	Nucleotide binding; nucleic acid binding (71)	Nuclear (0.94/1.00)	N/A

<sup>a</sup>F, molecular function; C, cellular component; P, biological process; N/A, not available.

sponsible for adhesion, binding, gene transcription, and protein degradation in various subcellular locations (Table 2). Four of the eight genes were found to have homologues in the pathogen-host interaction (PHI) database (PHI:733/PHI:6752, PHI:2524, PHI:4231, and PHI:5571), matching the evidence seen with the invasive pulmonary pathogen *Aspergillus fumigatus* (52), rice blast fungus *Magnaporthe oryzae* (53–55), and cereal-devastating pathogens *Fusarium verticillioides* and *F. graminearum* (56), as well as *Salmonella enterica*, the causative agent of a spectrum of diseases (57). In a broader comparison, Harpellales genomes harbor fewer transmembrane helix genes, pathogen-host interaction genes, and signal peptide genes than entomopathogenic species of both Ascomycota and Zoopagomycota (Fig. 5b). Interestingly, *O. sinensis*, like Harpellales species, possesses small amounts of the products mentioned above, which may reflect their biological distinction in their use of the oral-gut pathway during the fungal-insect infection rather than of the exoskeleton route used by other fungal pathogens (Table S2). The FISCoG enrichment analyses (Fig. 5d) further revealed that Harpellales and *Basidiobolus meristosporus* have the most genes encoding the cell adhesion proteins (FISCoG.g2), whereas entomopathogenic fungi tend to have more copies of genes encoding “platelet-activating factor acetylhydrolase” (PAF-AH) (FISCoG.g6) than Harpellales commensals. The genome-wide protein family (Pfam) domain enrichment pattern suggests that 15 domains are fungus-insect symbiotic core domains (FISCoD) and that 3 of them—“Amidohydro\_2,” nitronate monooxygenase

(“NMO”), and “PAF-AH\_p\_II”—are specifically enriched in entomopathogenic fungi (Fig. 5e).

## DISCUSSION

**Fungus-insect symbiotic core genes/domains toolbox.** By comparing the insect-symbiotic fungal genomes, including both commensals and pathogens corresponding to the taxa included in this study, we identified an eight-gene toolbox consisting of genes that are unique to the fungus-insect associations and missing from the closely related free-living relatives. The genome-wide enrichment analyses of these eight genes also highlighted several players that are important for the fungi during symbiosis. FISCoG.g2 was found to be specifically enriched in Harpellales and *Basidiobolus*, whereas FISCoG.g6 is instead enriched in entomopathogenic fungi (Fig. 5b). The FISCoG.g2 gene is a novel gene encoding 562 amino acids (aa), and the N-terminal 300 aa contain two repeats of the fasciclin (fas1) domain in close succession, while the C-terminal region (~262 aa) has no known function assigned (according to the results of a BLAST search, with the best hit with the *beta-ig-h3/fasciclin* gene; Table 2). Fas1 domain-containing proteins participate in cell adhesion and communication and are present in many eukaryotes (58–61). FISCoG.g2 is thus suggested to have similar adhesion functions at a minimum. Many fungal pathogens utilize proteins to adhere to host cells and in formation of biofilms. Some well-characterized genes in *Candida albicans* and *Metarhizium anisopliae* include *Als1* (62), *Flo11* (63), and *Mad1* and *Mad2* (64). A search of these genes found only a few (*Flo11* and *Mad1*) or none (*Als1* and *Mad2*) in the Harpellales genomes. These disparate data with respect to the preference for an adhesion protein reflect the different tools utilized by these symbiotic fungi in establishing the relationship with their animal hosts. In addition, FISCoG.g2 is suggested to be an extracellular protein (Table 2) and Harpellales asexual spores were earlier found to release glue in certain pH environments during the passage through the host gut (38). Given the evidence, it is reasonable to expect that a novel gut-attaching strategy is utilized by the Harpellales fungi. In that regard, FISCoG.g2 is suggested to be an important excreted protein for the gut-dwelling lifestyle of Harpellales. Homologues of the FISCoG.g6 are annotated with the function of being platelet-activating factor acetylhydrolases (PAF-AH). The platelet-activating factor (PAF) is often referred to specifically as a proinflammatory messenger and is widely utilized in both vertebrate and invertebrate animals (65, 66). The biological functions of the PAF-AH include regulation of inflammation through the inactivation or deconstruction of PAF in the animal system and thus play an important role during the initial step of fungal invasion. From the independent view of the comparative Pfam domains (Fig. 5c), the corresponding “PAF-AH\_p\_II” domain has also been identified as an important functional domain uniquely maintained by the insect-associated fungi and enriched in the entomopathogens. In addition, the Amidohydro\_2 and NMO domains present a similar pattern but are enriched only in the Ascomycota entomopathogens (Fig. 5c). Nitronate monooxygenase (NMO) was recently proven to be a novel factor in suppressing the host defense and promoting invasive hyphal growth and development during the pathogenic invasion using the example of the rice pathogen *Magnaporthe oryzae* (67). As a result, similar biological roles for the NMO-containing proteins during the fungus-insect interactions are anticipated, with the evidence that NMO was uniquely identified in insect-associated fungal proteomes and enriched in the entomopathogenic species. The specific function of the Amidohydro\_2 domain is still unknown; however, the amidohydrolase family includes adenine deaminase, which has been found to be important for adenine utilization and for providing a nitrogen source (68). Both FISCoG.g6 and PAF-AH\_p\_II and additional domains of NMO and Amidohydro\_2 are suggested to serve important biological functions for the interaction of entomopathogens (especially Ascomycota representatives) with their insect hosts.

Interestingly, the *Basidiobolus meristosporus* genome is enriched for FISCoG.g2, FISCoG.g5, and FISCoG.g8. Similarly to the aquatic Harpellales fungi, *B. meristosporus* seems to favor this novel fasciclin adhesion protein as well to maintain its residency

within amphibian and insect hosts. FISCoG.g5 is suggested to have the function of mediating ubiquitination and subsequent proteasomal degradation of targeted proteins (69, 70). FISCoG.g8 is predicted to be a differentiation regulator and has outstanding nucleotide/nucleic acid binding ability (according to the results of a BLAST search, with the best hit with Nrd1) (71). It is implied by the enrichment data that these three FISCoGs play bigger roles in the interaction of *B. meristosporus* with the animal hosts, emphasizing the abilities of adhesion, protein degradation, and nucleotide binding. The other four FISCoGs present no obvious enrichment patterns, although they have been suggested to contribute to oxidative metabolism, nuclear migration, and gene transcriptions during the fungus-insect interactions (52, 57, 72, 73).

The Harpellales fungal spores enter immature aquatic insect hosts via oral ingestion, and they germinate in the midgut or hindgut, where they finish the rest of the life cycle and release asexual or sexual spores at maturity during the intermolt phases of their insect hosts (5). This route of entry throughout ingestion and in-gut development exemplifies the prominent differences between Harpellales and most entomopathogenic fungi, including *B. bassiana*, *Conidiobolus coronatus*, *C. thromboides*, *Cordyceps militaris*, *M. robertsii*, and *O. unilateralis*, which heavily rely on their ability to degrade and penetrate the chitinous exoskeleton of potential insect hosts (27, 74). Usually, the secreted proteins that allow these penetration processes also play major roles in immune evasion (31). *Ophiocordyceps sinensis* is special in the way that it infects insect hosts through spiracles or the mouth and thus avoids the cuticle degradation step (75). As a result, the presentation of smaller amounts of signal peptide genes, transmembrane helix genes, and PHI genes in both Harpellales and *O. sinensis* than in the rest of the entomopathogens studied might be explained by their similar host invasion strategies in the form of taking the route of available openings of the host (see Table S2 in the supplemental material). The recent identification of protein family expansions (30, 75) corroborated the suggestion that *O. sinensis* has a much smaller number of CYP52 enzymes, subtilisins, trypsins, and aspartyl proteases than *B. bassiana*, *C. militaris*, or *M. robertsii* and that all are utilized for degradation of the insect cuticles.

**Host specificity.** It was suggested previously that a high proportion of secreted proteins positively correlate with the fungal parasitic lifestyle (27). Here we show that PHI genes, signal peptide genes, and transmembrane helix genes are all found in greater numbers in both Ascomycota and Zoopagomycota pathogens than in the Harpellales commensals (Fig. 5b). Genomes of host generalists tend to be equipped with genes that encode expanded numbers of protein families, while specialists encode greater numbers of species-specific proteins but lack the diversity of genes that encode the secretory signal peptides that are used to interact with various hosts (31). Surprisingly, a few host-specialized fungi are found with large genome sizes, although the number of gene models does not increase proportionally (31, 39, 48, 76). One interpretation suggests that the genome size variation could be a consequence of variation in telomeric regions represented by noncoding and small genomic repetitions and, thus, that the larger genome size is accompanied by increased genomic complexity (76).

*Ophiocordyceps sinensis* requires the ghost moth (Hepialidae) as a host to complete its life cycle, and both the fungus and insect hosts are endemic in the Qinghai-Tibet Plateau in western China (77). *Ophiocordyceps sinensis* was reported to have potential hosts, including 57 species of the ghost moth (78), although low genetic diversity among these potential hosts was suggested by the results of studies performed using the mitochondrial cytochrome C oxidase subunit I (COI) marker (77). Contemporarily, Zhang et al. (79) revealed higher genetic variation in both the fungus and the ghost moth using 7 fungal and 3 insect markers and further suggested that *O. sinensis* cospeciated with the ghost moth on the basis of significant cophylogenetic congruence and similar divergence times. In accordance with the aforementioned results, we assume that *O. sinensis* is highly restricted to the ghost moth hosts; thus, we treated it

as a host-specialized fungal pathogen in this study (Table S2). *O. unilateralis* is a well-known specialized fungal pathogen that manipulates and kills formicine ants (74).

The generalist-specialist pairs in taxonomic clade present comparable patterns—the genome of the generalist *M. robertsii* includes more signal peptide genes (1,707 versus 1,212), transmembrane helix genes (11,022 versus 9,136), and PHI genes (3,858 versus 3,268) than the specialist *M. acridum*; similarly, genes of the members of the specialist group consisting of *O. sinensis* and *O. unilateralis* encode noticeably fewer signal peptides, transmembrane helices, and PHI proteins than the genes of the other Ascomycota generalist entomopathogens (Table S2). The higher signal peptide number in *O. unilateralis* than in *O. sinensis* may contribute to a process described in a recent finding in which *O. unilateralis* cells invade host muscle fibers and form networks throughout the body to perform the manipulations (15). Similar patterns have been revealed in the Harpellales (Zoopagomycota; Kickxellomycotina). The hosts of *Smittium mucronatum* were found to be restricted to *Psectrocladius* (midge) (5), and the numbers of its signal peptide genes (700), transmembrane helix genes (4,550), and PHI genes (1,725) were all found to be lower than the corresponding Harpellales averages (733, 4,837, and 1,880) (Table S2). *Capniomyces stellatus*, as another host-specialized Harpellales, was reported to be present specifically in winter stoneflies (Capniidae and Taeniopterygidae) (80, 81). Similarly to *S. mucronatum*, *Capniomyces stellatus* harbors lower numbers of transmembrane helix genes (4,440) and PHI genes (1,763) than the Harpellales averages but a higher number of signal peptide genes (792). The data imply that *C. stellatus* may maintain a more intensive interaction with the stonefly hosts than we previously thought. The noncoincidental large genome sizes of *O. sinensis* and *S. mucronatum* may be the result of host specialization and increased genome complexity, a conjecture that was also strengthened by several independent findings (31, 48, 76).

**Harpellales genome evolution.** We used the newly produced Harpellales genome sequences to reconstruct the first phylogenomic tree of Harpellales using 1,241 orthologous genes (Fig. 3a). Analysis of the phylogenomic tree indicates a pattern implying that multiple genome-level duplications may have occurred (Fig. 3b). One parsimonious possibility is that the genome size of the Harpellales ancestor is similar with that of the members of the basal non-*Smittium* Harpellales clade (*Capniomyces* and *Zancudomyces*). The sizes remain the same in both clades of non-*Smittium* Harpellales and in Parasmittium subclade II. The size nearly doubled (from ~25 Mb to ~44 Mb) when Parasmittium subclade I diverged from the Parasmittium subclade II. The true *Smittium* genome sizes (71 to 102 Mb) approximately tripled in comparison to that of the Harpellales ancestor (~25 Mb), and then the genome size of *S. mucronatum* enlarged further due to the host specialization. Alternatively, the genome size of the Parasmittium ancestor might have experienced duplication followed by reduction of the levels of Parasmittium subclade II members when their genomes further evolved, rearranged, and got rid of the redundant regions. The availability of genome sequences of *Smittium morbosum* and *Stachylina* from a sister clade would help confirm this explanation (Fig. 1a). Unfortunately, the lack of culture of both strains prevented us from getting their genome information easily. The true *Smittium* clade holds the largest genome sizes in the Harpellales. It is also likely that the genome sizes of the true *Smittium* members underwent more than one whole-genome duplication from the Harpellales ancestor (Table S1).

It is still relatively rare to discover whole-genome duplication (WGD) events in fungal genomes, although studies of the members from both Dikarya and early-diverging groups have resulted in the recent publication of independent reports of the finding of WGD, with increasing number of genome sequences and improving technology (82–86). The yeast (*Saccharomyces cerevisiae*) genome has been intensively studied since its production (87). Found with many duplicated genes, yeasts were thought to have undergone genome duplication that was followed by massive gene loss events (88), although this was not officially confirmed until the production of the genome of a more

basal relative, *Kluyveromyces waltii* (82). Phylogenetic evidence recently found that the yeast genome doubling was the result of a contemporaneous interspecies hybridization within the baker's yeast lineage (89). Immediately after that, the genome hybridization model was proposed (90). Multiple Harpellales species have been identified from the same host gut, and in some cases, they may influence their relative locations along the gut (91). It would be worthwhile to further test whether the interspecies hybridization model might explain the Harpellales genome multiplication. In total, 1,071 single-copy genes were identified from the four smaller-genome-size taxa (genome sizes of between 25 and 28 Mb). The finding of various numbers of copies (0 to 4) present in the larger-genome-size taxa (genome sizes of  $\geq 44$  Mb) suggests that the duplication events were so ancient that many of the copies have been lost (Table S1). The 527 independent phylogenetic trees were built using the single-copy markers, and none supports the interspecies hybridization hypothesis (Fig. S1). However, we still cannot exclude the possibility of ancient hybridization due to the fact that the signal might be too weak to correspond to the ancient age of the event. In addition, this method also cannot exclude the idea of the possibility of hybridization between closely related species. The *Smittium* and non-*Smittium* clades diverged much earlier (270 Ma) than the yeast hybridization event (100 to 140 Ma), which could be one of the major obstacles to detection of such a signal (33, 89, 92). The draft quality of Harpellales genomes also prevents us from comparing their syntenic data at the chromosome level. With future research efforts designed to refine high-quality genome assemblies, new comparative genomic methods will serve to help unravel the mystery of this symbiotic system.

## MATERIALS AND METHODS

**Fungal strains, DNA extraction, and whole-genome sequencing.** *Furculomyces boomerangus* (AUS-77-4; ARSEF 9021), *Smittium angustum* (AUS-126-30; ARSEF 9241), *S. megazygosporum* (SC-DP-2; ARSEF 9037), and *S. simulii* (SWE-8-4; ARSEF 9139) were obtained from the USDA-ARS Collection of Entomopathogenic Fungal Cultures (ARSEF). Fungal cultures were grown and DNA was extracted following earlier protocols (39). TruSeq Nano paired-end (PE) libraries with an insertion size of 500 bp were prepared for each of *Furculomyces boomerangus*, *S. angustum*, and *S. megazygosporum*. *Smittium simulii* was prepared using one PCR-free PE library (500-bp insertion size) and two Nextera mate-pair (MP) libraries (3-kb and 5-kb insertion sizes). All were sequenced using an Illumina HiSeq 2500 platform (2 × 125-bp read length) at the Donnelly Sequencing Center, University of Toronto (Canada).

**Genome assembly and annotation.** Raw FASTQ sequence reads were subjected to adapter trimming using Trim Galore v0.4.1 ([http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) and were quality checked using FASTQC v0.11.4 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Genomes were assembled with RAY v2.3.1 (93). The scaffolds of *S. simulii* were built using SSPACE (94). Satellites, simple repeats, and low-complexity sequences were annotated with RepeatMasker v4.0.5 (<http://www.repeatmasker.org>) and Tandem Repeat Finder v4.07b (95), corresponding to fungal sequences from RepBase (96). The genomes were annotated with the Funannotate v0.6.2 pipeline (<https://github.com/nextgenusfs/funannotate>), employing tools of Augustus (97), GeneMark.hmm-E5 (98), and EVM (99). Gene function was inferred from matches to the databases of Pfam (100), Merops (101), CAZy (102), InterProScan (103), and Swiss-Prot (104). Product descriptions were assigned with homologues with 60% similarity across 60% of the protein length (105). CEGMA v2.4.010312 was used to identify the presence of core eukaryotic-protein-coding genes and for subsequent evaluation of genome coverage (106). Secreted proteins were predicted using SignalP v4.1 (no truncation to the sequence length) (107), and transmembrane helices were predicted using TMHMM v2.0 (108). Potential pathogenic proteins were identified using BLASTP against pathogen-host interaction (PHI) database v4.4 (with 4,376 entries) (109).

**Homologue identification and phylogenomics of Harpellales.** Putative homologues among the nine Harpellales genomes were identified using two independent methods. The first method employed reciprocal similarity searches using BLASTP v2.2.30 (cutting E value set to  $1E^{-5}$ ) and a Perl script ("Find\_mutual\_BestHit.pl"; available from GitHub) and filtering for the reciprocal best hit. The second method employed MCL v14-137 (110) and a Python script ("Ortho\_Rep.py"; available from GitHub), selecting the representatives of each clustering group based on a theoretical graph approach—the best sequence of each taxon from each cluster was chosen based on its connectivity (weighted by  $\log_{10}$ -transformed E-value) to all other sequences. Venn diagrams were produced using R v3.1.3 to visualize the results of comparisons. A group of all 1,280 of the shared Harpellales homologues retrieved as described above was used for phylogenomic analyses. Multiple-sequence alignment was performed using MUSCLE v3.6 (111). Gaps and poorly aligned regions were excluded using trimAl v1.4 (112). Aligned protein sequences that were longer than 50 amino acids (1,241 of the 1,280 homologues) were selected using a Python script ("Filter\_Len.py"; available from GitHub) and then concatenated using FASconCAT-G v1.02 (113). Data corresponding to appropriate substitutional models, the best partition scheme, phylog-



enomic tree reconstruction, and ultrafast bootstrap analyses were all inferred and analyses conducted within the IQ-TREE v1.5.3 package (114–116). In order to test whether the revealed integer-multiple pattern of genome sizes was due to interspecies hybridization, we estimated the phylogenetic relationship of the multiple-copy homologous genes within the larger genomes (44 to 102 Mb). We used the comparative genomic pipeline ([https://github.com/stajichlab/Comparative\\_pipeline](https://github.com/stajichlab/Comparative_pipeline)) to identify 582 single-copy protein-coding genes that were all shared among the four taxa with the smallest genome sizes (25 to 28 Mb) and that also were found to be present with a minimum of two copies in at least one of the remaining five taxa (genome sizes between 44 and 102 Mb). Individual homologous alignments were prepared similarly to those described above. Finally, 527 phylogenetic trees were reconstructed using the maximum-likelihood analysis and the ultrabootstrap method and 1,000 replications with the IQ-TREE tool (55 of the 582 alignments were discarded due to having a length shorter than 50 aa).

**Single nucleotide polymorphisms, allele frequencies, and dot plot comparisons.** The whole-genome-wide SNPs were identified using BWA (117), SAMtools (118), Picard Tools (<http://broadinstitute.github.io/picard>), and the GATK toolkit (119), according to the previously described pipeline method (120). A total of 460 to 484 single-copy orthologs among the nine Harpellales (with one taxon allowed to be missing from the total of nine taxa) were discovered from the earlier step using the comparative pipeline. The individual allele frequency was calculated using the read counts of the binary alignment map (BAM) files against the single-copy orthologs with the “pileup2snp” function of VarScan v2.3.9 (125) (default parameters were determined using a –min-coverage value of 10 and a *P* value of 0.05). The allele frequency value was then rounded to an integer value before being plotted with the number of occurrences and level of density using R v3.1.3 (<http://www.r-project.org>). The Harpellales genome dot plot comparisons were generated using the MUMmer (v3.23) package (121). The “nucmer” algorithm and the “delta-filter” function were employed to find the reciprocal best matches between the genome scaffolds. Total numbers of matches and aligned bases were calculated using the “show-coords” function. The percentage of identity was calculated using a Python script (“Calcul\_Identity.py”; available from GitHub).

**FISCOG toolbox.** The FISCOG toolbox was identified by comparing the Harpellales core genes (including those allowed to be mostly missing but minimally present in six of the nine taxa) with those corresponding to the representatives of entomopathogenic fungi from the Ascomycota and Zoopagomycota groups (27, 29, 46, 49–51). These included six taxa from Ascomycota—*Beauveria bassiana*, *Cordyceps militaris*, *Metarhizium acridum*, *Metarhizium robertsii*, *Ophiocordyceps sinensis*, and *Ophiocordyceps unilateralis* (allowing each gene to be missing from up to two taxa) and three from Zoopagomycota—*Basidiobolus meristosporus*, *Conidiobolus coronatus*, and *Conidiobolus thromboides* (with no genes allowed to be missing from any taxa). The FISCOG candidates were then subjected to a BLAST search (cutting E-value set to  $1E^{-5}$ ) against the non-insect-associated Zoopagomycota genomes (*Cemansia reversa*, *Kickella alabastrina*, *Linderina pennispora*, and *Martensiomycetes pterosporus*) to filter out potential false-positive hits. FISCOG enrichment in each fungal taxon was assessed using BLASTP. The enrichment heat map was produced using the “aheat map” function in R package “NMF” (122). The cladogram of the included taxa was reconstructed using IQ-TREE, based on 29 proteins (longer than 50 aa) encoded by single-copy genes. The GO names and subcellular protein locations of the FISCOGs were predicted using Blast2GO v5.0 and the MultiLoc2 webserver against the fungal high-resolution database, respectively (123, 124).

**Data availability.** All four Harpellales taxa and corresponding genome information have been deposited in DDBJ/ENA/GenBank under the following accession numbers: BioProject identifier (ID) [PRJNA329411](https://doi.org/10.1128/PRJNA329411); BioSamples IDs SAMN05412443, SAMN05412446, SAMN05412450, and SAMN05412451; and Whole-Genome Shotgun IDs [MBFU00000000](https://doi.org/10.1128/MBFU00000000), [MBFT00000000](https://doi.org/10.1128/MBFT00000000), [MBFS00000000](https://doi.org/10.1128/MBFS00000000), and [MBFR00000000](https://doi.org/10.1128/MBFR00000000). Multiple-sequence alignment and phylogenetic tree files are archived in TreeBASE (study ID S22516). Script files created for this study are available at <https://github.com/YanWangTF>.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.00636-18>.

**FIG S1**, PDF file, 1.3 MB.

**FIG S2**, PDF file, 0.4 MB.

**TABLE S1**, PDF file, 0.5 MB.

**TABLE S2**, PDF file, 0.1 MB.

**TABLE S3**, PDF file, 0.1 MB.

## ACKNOWLEDGMENTS

We thank J. Anderson, D. Currie, S. Kvist, H. López-Fernández, S. Wright, and J. P. Xu for their critical comments on the research design. We also thank JGI for the permission to use genomes ahead of publication.

This work was supported by a University of Toronto Fellowship (to Y.W.) and Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant 453847 to J.-M.M. Y.W. sincerely thanks the Mycological Society of America for the Robert W. Lichtwardt Research Award and acknowledges the SciNet staff at the University of



Toronto and the Biocluster staff at the University of California, Riverside, for facilitating access to the supercomputing infrastructure. We gratefully acknowledge support from the ZyGoLife project (National Science Foundation [NSF] award DEB1441715) for support of ongoing research. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under contract number DE-AC02-05CH11231.

## REFERENCES

- James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Guaidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung GH, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schüssler A, Longcore JE, O'Donnell K, Mozley-Standridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkman-Kohlmeyer B, Spotts RA, et al. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443:818–822. <https://doi.org/10.1038/nature05110>.
- Currie CR, Wong B, Stuart AE, Schultz TR, Rehner SA, Mueller UG, Sung GH, Spatafora JW, Straus NA. 2003. Ancient tripartite coevolution in the attine ant-microbe symbiosis. *Science* 299:386–388. <https://doi.org/10.1126/science.1078155>.
- Scholte EJ, Ng'habi K, Kihonda J, Takken W, Paaijmans K, Abdulla S, Killeen GF, Knols BGJ. 2005. An entomopathogenic fungus for control of adult African malaria mosquitoes. *Science* 308:1641–1642. <https://doi.org/10.1126/science.1108639>.
- Wheeler Q, Blackwell M. 1984. *Fungus-insect relationships: perspectives in ecology and evolution*. Columbia University Press, New York, NY.
- Lichtwardt RW. 1986. *The Trichomyces, fungal associates of arthropods*, 1st ed. Springer-Verlag, New York, NY.
- Sessions OM, Barrows NJ, Souza-Neto JA, Robinson TJ, Hershey CL, Rodgers MA, Ramirez JL, Dimopoulos G, Yang PL, Pearson JL, Garcia-Blanco MA. 2009. Discovery of insect and human dengue virus host factors. *Nature* 458:1047–1050. <https://doi.org/10.1038/nature07967>.
- Moran NA. 2006. Symbiosis. *Curr Biol* 16:R866–R871. <https://doi.org/10.1016/j.cub.2006.09.019>.
- Scholte EJ, Knols BGJ, Samson RA, Takken W. 2004. Entomopathogenic fungi for mosquito control: a review. *J Insect Sci* 4:19.
- Butt TM, Greenfield BPJ, Greig C, Maffei TGG, Taylor JWD, Piasecka J, Dudley E, Abdulla A, Dubovskiy IM, Garrido-Jurado I, Quesada-Moraga E, Penny MW, Eastwood DC. 2013. *Metarhizium anisopliae* pathogenesis of mosquito larvae: a verdict of accidental death. *PLoS One* 8:e81686. <https://doi.org/10.1371/journal.pone.0081686>.
- Paterson RRM. 2008. *Cordyceps*: a traditional Chinese medicine and another fungal therapeutic biofactory? *Phytochemistry* 69:1469–1495. <https://doi.org/10.1016/j.phytochem.2008.01.027>.
- McCreddie JW, Beard CE, Adler PH. 2005. Context-dependent symbiosis between black flies (Diptera: Simuliidae) and trichomycete fungi (Harpellales: Legeriomycetaceae). *Oikos* 108:362–370. <https://doi.org/10.1111/j.0030-1299.2005.13417.x>.
- Moran NA, Jarvik T. 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* 328:624–627. <https://doi.org/10.1126/science.1187113>.
- Brobyn PJ, Wilding N. 1983. Invasive and developmental processes of *Entomophthora muscae* infecting houseflies (*Musca domestica*). *Trans Br Mycol Soc* 80:1–8. [https://doi.org/10.1016/S0007-1536\(83\)80157-0](https://doi.org/10.1016/S0007-1536(83)80157-0).
- Andersen SB, Gerritsma S, Yusah KM, Mayntz D, Hywel-Jones NL, Billen J, Boomsma JJ, Hughes DP. 2009. The life of a dead ant: the expression of an adaptive extended phenotype. *Am Nat* 174:424–433. <https://doi.org/10.1086/603640>.
- Fredericksen MA, Zhang Y, Hazen ML, Loreto RG, Mangold CA, Chen DZ, Hughes DP. 2017. Three-dimensional visualization and a deep-learning model reveal complex fungal parasite networks in behaviorally manipulated ants. *Proc Natl Acad Sci U S A* 114:12590–12595. <https://doi.org/10.1073/pnas.1711673114>.
- Steinhaus EA. 1956. Microbial control—the emergence of an idea: a brief history of insect pathology through the nineteenth century. *Hilgardia* 26:107–160. <https://doi.org/10.3733/hilg.v26n02p107>.
- Roberts DW, St Leger RJ. 2004. *Metarhizium* spp., cosmopolitan insect-pathogenic fungi: mycological aspects. *Adv Appl Microbiol* 54:1–70. [https://doi.org/10.1016/S0065-2164\(04\)54001-7](https://doi.org/10.1016/S0065-2164(04)54001-7).
- Sweeney AW. 1981. An undescribed species of *Smittium* (Trichomyces) pathogenic to mosquito larvae in Australia. *Trans Br Mycol Soc* 77:55–60. [https://doi.org/10.1016/S0007-1536\(81\)80179-9](https://doi.org/10.1016/S0007-1536(81)80179-9).
- Wang C, St Leger RJ. 2007. A scorpion neurotoxin increases the potency of a fungal insecticide. *Nat Biotechnol* 25:1455–1456. <https://doi.org/10.1038/nbt1357>.
- Knols BGJ, Bukhari T, Farenhorst M. 2010. Entomopathogenic fungi as the next-generation control agents against malaria mosquitoes. *Future Microbiol* 5:339–341. <https://doi.org/10.2217/fmb.10.11>.
- Sung GH, Hywel-Jones NL, Sung JM, Luangsa-Ard JJ, Shrestha B, Spatafora JW. 2007. Phylogenetic classification of *Cordyceps* and the clavicipitaceous fungi. *Stud Mycol* 57:5–59. <https://doi.org/10.3114/sim.2007.57.01>.
- Zhu JS, Halpern GM, Jones K. 1998. The scientific rediscovery of an ancient Chinese herbal medicine: *Cordyceps sinensis*: part I. *J Altern Complement Med* 4:289–303. <https://doi.org/10.1089/acm.1998.4.3-289>.
- Stajich JE. 2017. Fungal genomes and insights into the evolution of the Kingdom. *Microbiol Spectr* 5:1–15. <https://doi.org/10.1128/microbiolspec.FUNK-0055-2016>.
- Spatafora JW. 2011. 1000 fungal genomes to be sequenced. *IMA Fungus* 2:41. <http://www.imafungus.org/issue/4/04.pdf>.
- Wang C, Wang S. 2017. Insect pathogenic fungi: genomics, molecular interactions, and genetic improvements. *Annu Rev Entomol* 62:73–90. <https://doi.org/10.1146/annurev-ento-031616-035509>.
- de Bekker C, Ohm RA, Loreto RG, Sebastian A, Albert I, Meroow M, Brachmann A, Hughes DP. 2015. Gene expression during zombie ant biting behavior reflects the complexity underlying fungal parasitic behavioral manipulation. *BMC Genomics* 16:620. <https://doi.org/10.1186/s12864-015-1812-x>.
- Xiao G, Ying SH, Zheng P, Wang ZL, Zhang S, Xie XQ, Shang Y, St Leger RJ, Zhao GP, Wang C, Feng MG. 2012. Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Sci Rep* 2:483. <https://doi.org/10.1038/srep00483>.
- Bischoff JF, Rehner SA, Humber RA. 2009. A multilocus phylogeny of the *Metarhizium anisopliae* lineage. *Mycologia* 101:512–530. <https://doi.org/10.3852/07-202>.
- Gao Q, Jin K, Ying SH, Zhang Y, Xiao G, Shang Y, Duan Z, Hu X, Xie XQ, Zhou G, Peng G, Luo Z, Huang W, Wang B, Fang W, Wang S, Zhong Y, Ma LJ, St Leger RJ, Zhao GP, Pei Y, Feng MG, Xia Y, Wang C. 2011. Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genet* 7:e1001264. <https://doi.org/10.1371/journal.pgen.1001264>.
- Wichadakul D, Kobmoo N, Ingsriswang S, Tangphatsornruang S, Chantasingh D, Luangsa-Ard JJ, Eurwilaichitr L. 2015. Insights from the genome of *Ophiocordyceps polyrhachis-furcata* to pathogenicity and host specificity in insect fungi. *BMC Genomics* 16:881. <https://doi.org/10.1186/s12864-015-2101-4>.
- Desjardins CA, Sanscrainte ND, Goldberg JM, Heiman D, Young S, Zeng Q, Madhani HD, Becnel JJ, Cuomo CA. 2015. Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nat Commun* 6:7121. <https://doi.org/10.1038/ncomms8121>.
- White MM. 2006. Evolutionary implications of a rRNA-based phylogeny of Harpellales. *Mycol Res* 110:1011–1024. <https://doi.org/10.1016/j.mycres.2006.06.006>.
- Wang Y. 2016. *Genome evolution of Smittium and allies (Harpellales)*. PhD dissertation. University of Toronto, Toronto, Canada.
- Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N, Grigoriev I, Gryganskyi A, James TY, O'Donnell K, Roberson RW, Taylor TN, Uehling J, Vilgalys R, White MM, Stajich JE.

2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* 108:1028–1046. <https://doi.org/10.3852/16-042>.
35. White MM, Lichtwardt RW, Colbo MH. 2006. Confirmation and identification of parasitic stages of obligate endobionts (Harpellales) in blackflies (Simuliidae) by means of rRNA sequence data. *Mycol Res* 110: 1070–1079. <https://doi.org/10.1016/j.mycres.2006.06.008>.
  36. Horn BW, Lichtwardt RW. 1981. Studies on the nutritional relationship of larval *Aedes aegypti* (Diptera: Culicidae) with *Smittium culisetae* (Trichomycetes). *Mycologia* 73:724–740. <https://doi.org/10.2307/3759499>.
  37. Williams MC. 2001. Trichomycetes: a brief review of research, p 15–25. In Misra JK, Horn B (ed), *Trichomycetes and other fungal groups*. Science Publishers Inc., Enfield, NH.
  38. Horn BW. 1989. Requirement for potassium and pH shift in host-mediated sporangiospore extrusion from trichospores of *Smittium culisetae* and other *Smittium* species. *Mycol Res* 93:303–313. [https://doi.org/10.1016/S0953-7562\(89\)80157-1](https://doi.org/10.1016/S0953-7562(89)80157-1).
  39. Wang Y, White MM, Kvist S, Moncalvo JM. 2016. Genome-wide survey of gut fungi (Harpellales) reveals the first horizontally transferred ubiquitin gene from a mosquito host. *Mol Biol Evol* 33:2544–2554. <https://doi.org/10.1093/molbev/msw126>.
  40. Wang Y, White MM, Moncalvo JM. 2016. Draft genome sequence of *Capniomyces stellatus*, the obligate gut fungal symbiont of stonefly. *Genome Announc* 4:e00761-16. <https://doi.org/10.1128/genomeA.00761-16>.
  41. Wang Y, Tretter ED, Johnson EM, Kandel P, Lichtwardt RW, Novak SJ, Smith JF, White MM. 2014. Using a five-gene phylogeny to test morphology-based hypotheses of *Smittium* and allies, endosymbiotic gut fungi (Harpellales) associated with arthropods. *Mol Phylogenet Evol* 79:23–41. <https://doi.org/10.1016/j.ympev.2014.05.008>.
  42. Tretter ED, Johnson EM, Benny GL, Lichtwardt RW, Wang Y, Kandel P, Novak SJ, Smith JF, White MM. 2014. An eight-gene molecular phylogeny of the Kickxellomycotina, including the first phylogenetic placement of Asellariales. *Mycologia* 106:912–935. <https://doi.org/10.3852/13-253>.
  43. Leonforte A, Sudheesh S, Cogan NOI, Salisbury PA, Nicolas ME, Materne M, Forster JW, Kaur S. 2013. SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.). *BMC Plant Biol* 13:161. <https://doi.org/10.1186/1471-2229-13-161>.
  44. LaFramboise T. 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res* 37:4181–4193. <https://doi.org/10.1093/nar/gkp552>.
  45. Forche A, Magee PT, Magee BB, May G. 2004. Genome-wide single-nucleotide polymorphism map for *Candida albicans*. *Eukaryot Cell* 3:705–714. <https://doi.org/10.1128/EC.3.3.705-714.2004>.
  46. Zheng P, Xia Y, Xiao G, Xiong C, Hu X, Zhang S, Zheng H, Huang Y, Zhou Y, Wang S, Zhao GP, Liu X, St Leger RJ, Wang C. 2011. Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional Chinese medicine. *Genome Biol* 12:R116. <https://doi.org/10.1186/gb-2011-12-11-r116>.
  47. Hu X, Xiao G, Zheng P, Shang Y, Su Y, Zhang X, Liu X, Zhan S, St Leger RJ, Wang C. 2014. Trajectory and genomic determinants of fungal-pathogen speciation and host adaptation. *Proc Natl Acad Sci U S A* 111:16796–16801. <https://doi.org/10.1073/pnas.1412662111>.
  48. Hu X, Zhang YJ, Xiao GH, Zheng P, Xia YL, Zhang XY, St Leger RJ, Liu XZ, Wang CS. 2013. Genome survey uncovers the secrets of sex and lifestyle in caterpillar fungus. *Chin Sci Bull* 58:2846–2854. <https://doi.org/10.1007/s11434-013-5929-5>.
  49. Chang Y, Wang S, Sekimoto S, Aerts AL, Choi C, Clum A, LaButti KM, Lindquist EA, Yee Ngan CY, Ohm RA, Salamov AA, Grigoriev IV, Spatafora JW, Berbee ML. 2015. Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biol Evol* 7:1590–1601. <https://doi.org/10.1093/gbe/evv090>.
  50. Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, Haridas S, Kuo A, Salamov A, Ahrendt SR, Lau R, Bowen BP, Lipzen A, Sullivan W, Andreopoulos BB, Clum A, Lindquist E, Daum C, Northen TR, Kunde-Ramamoorthy G, Schmitz RJ, Gryganskyi A, Culley D, Magnuson J, James TY, O'Malley MA, Stajich JE, Spatafora JW, Visel A, Grigoriev IV. 2017. Widespread adenine N6-methylation of active genes in fungi. *Nat Genet* 49:964–968. <https://doi.org/10.1038/ng.3859>.
  51. Arnesen JA, Malagocka J, Gryganskyi A, Grigoriev IV, Voigt K, Stajich JE, Licht HHDF. 2018. Early diverging insect pathogenic fungi of the order Entomophthorales possess diverse and unique subtilisin-like serine proteases. *bioRxiv* <https://doi.org/10.1101/247858>.
  52. Romero B, Turner G, Olivás I, Laborda F, De Lucas JR. 2003. The *Aspergillus nidulans* *alca* promoter drives tightly regulated conditional gene expression in *Aspergillus fumigatus* permitting validation of essential genes in this human pathogen. *Fungal Genet Biol* 40:103–114. [https://doi.org/10.1016/S1087-1845\(03\)00090-2](https://doi.org/10.1016/S1087-1845(03)00090-2).
  53. Liu TB, Chen GQ, Min H, Lin FC. 2009. *MoFLP1*, encoding a novel fungal fasciclin-like protein, is involved in conidiation and pathogenicity in *Magnaporthe oryzae*. *J Zhejiang Univ Sci B* 10:434–444. <https://doi.org/10.1631/jzus.B0920017>.
  54. Jeon J, Rho H, Kim S, Kim KS, Lee YH. 2014. Role of *MoAND1*-mediated nuclear positioning in morphogenesis and pathogenicity in the rice blast fungus, *Magnaporthe oryzae*. *Fungal Genet Biol* 69:43–51. <https://doi.org/10.1016/j.fgb.2014.05.002>.
  55. Kershaw MJ, Talbot NJ. 2009. Genome-wide functional analysis reveals that infection-associated fungal autophagy is necessary for rice blast disease. *Proc Natl Acad Sci U S A* 106:15967–15972. <https://doi.org/10.1073/pnas.0901477106>.
  56. Shim WB, Sagaram US, Choi YE, So J, Wilkinson HH, Lee YW. 2006. FSR1 is essential for virulence and female fertility in *Fusarium verticillioides* and *F. graminearum*. *Mol Plant Microbe Interact* 19:725–733. <https://doi.org/10.1094/MPMI-19-0725>.
  57. Sang Y, Ren J, Ni J, Tao J, Lu J, Yao YF. 2016. Protein acetylation is involved in *Salmonella enterica* serovar Typhimurium virulence. *J Infect Dis* 213:1836–1845. <https://doi.org/10.1093/infdis/jiw028>.
  58. Skonier J, Bennett K, Rothwell V, Kosowski S, Plowman G, Wallace P, Edelhoff S, Disteche C, Neubauer M, Marquardt H. 1994. Beta-g-h3: a transforming growth factor- $\beta$ -responsive gene encoding a secreted protein that inhibits cell attachment *in vitro* and suppresses the growth of CHO cells in nude mice. *DNA Cell Biol* 13:571–584. <https://doi.org/10.1089/dna.1994.13.571>.
  59. Kim JE, Kim SJ, Lee BH, Park RW, Kim KS, Kim IS. 2000. Identification of motifs for cell adhesion within the repeated domains of transforming growth factor- $\beta$ -induced gene, betaig-h3. *J Biol Chem* 275: 30907–30915. <https://doi.org/10.1074/jbc.M002752200>.
  60. Clout NJ, Tisi D, Hohenester E. 2003. Novel fold revealed by the structure of a FAS1 domain pair from the insect cell adhesion molecule fasciclin I. *Structure* 11:197–203. [https://doi.org/10.1016/S0969-2126\(03\)00002-9](https://doi.org/10.1016/S0969-2126(03)00002-9).
  61. Huber O, Sumper M. 1994. Algal-CAMs: isoforms of a cell adhesion molecule in embryos of the alga *Volvox* with homology to *Drosophila* fasciclin I. *EMBO J* 13:4212–4222.
  62. Fu Y, Ibrahim AS, Sheppard DC, Chen YC, French SW, Cutler JE, Filler SG, Edwards JE. 2002. *Candida albicans* Als1p: an adhesin that is a downstream effector of the EFG1 filamentation pathway. *Mol Microbiol* 44:61–72. <https://doi.org/10.1046/j.1365-2958.2002.02873.x>.
  63. Guo B, Styles CA, Feng Q, Fink GR. 2000. A *Saccharomyces* gene family involved in invasive growth, cell-cell adhesion, and mating. *Proc Natl Acad Sci U S A* 97:12158–12163. <https://doi.org/10.1073/pnas.220420397>.
  64. Wang C, St Leger RJ. 2007. The *MAD1* adhesin of *Metarhizium anisopliae* links adhesion with blastospore production and virulence to insects, and the *MAD2* adhesin enables attachment to plants. *Eukaryot Cell* 6:808–816. <https://doi.org/10.1128/EC.00409-06>.
  65. Derewenda ZS, Derewenda U. 1998. The structure and function of platelet-activating factor acetylhydrolases. *Cell Mol Life Sci* 54: 446–455. <https://doi.org/10.1007/s00180050172>.
  66. Stafforini DM, Tjoelker LW, McCormick SPA, Vaitkus D, McIntyre TM, Gray PW, Young SG, Prescott SM. 1999. Molecular basis of the interaction between plasma platelet-activating factor acetylhydrolase and low density lipoprotein. *J Biol Chem* 274:7018–7024. <https://doi.org/10.1074/jbc.274.11.7018>.
  67. Marroquin-Guzman M, Hartline D, Wright JD, Elowsky C, Bourret TJ, Wilson RA. 2017. The *Magnaporthe oryzae* nitrooxidative stress response suppresses rice innate immunity during blast disease. *Nat Microbiol* 2:17054. <https://doi.org/10.1038/nmicrobiol.2017.54>.
  68. Nygaard P, Duckert P, Saxild HH. 1996. Role of adenine deaminase in purine salvage and nitrogen metabolism and characterization of the *ade* gene in *Bacillus subtilis*. *J Bacteriol* 178:846–853. <https://doi.org/10.1128/jb.178.3.846-853.1996>.
  69. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C, Bennett J, Bowyer P, Chen D, Collins M, Coulson R, Davies R, Dyer PS, Farman M, Fedorova N,

- Fedorova N, Feldblyum TV, Fischer R, Fosker N, Fraser A, García JL, García MJ, Goble A, Goldman GH, Gomi K, Griffith-Jones S, Gwilliam R, Haas B, Haas H, Harris D, Horiuchi H, Huang J, Humphray S, Jiménez J, Keller N, Khouri H, Kitamoto K, Kobayashi T, Konzack S, Kulkarni R, Kumagai T, Lafon A, Latgé JP, Li W, Lord A, Lu C, et al. 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 438:1151–1156. <https://doi.org/10.1038/nature04332>.
70. Jaquenoud M, Gulli MP, Peter K, Peter M. 1998. The Cdc42p effector Gic2p is targeted for ubiquitin-dependent degradation by the SCF-(Grr1) complex. *EMBO J* 17:5360–5373. <https://doi.org/10.1093/emboj/17.18.5360>.
71. Düring L, Thorsen M, Petersen DSN, Køster B, Jensen TH, Holmberg S. 2012. MRN1 implicates chromatin remodeling complexes and architectural factors in mRNA maturation. *PLoS One* 7:e44373. <https://doi.org/10.1371/journal.pone.0044373>.
72. Abdelraheem SR, Spiller DG, McLennan AG. 2003. Mammalian NADH diphosphatases of the Nudix family: cloning and characterization of the human peroxisomal NUDT12 protein. *Biochem J* 374:329–335. <https://doi.org/10.1042/BJ20030441>.
73. Cui X, De Vivo I, Slany R, Miyamoto A, Firestein R, Cleary ML. 1998. Association of SET domain and myotubularin-related proteins modulates growth control. *Nat Genet* 18:331–337. <https://doi.org/10.1038/ng0498-331>.
74. Evans HC, Elliot SL, Hughes DP. 2011. *Ophiocordyceps unilateralis*: a keystone species for unraveling ecosystem functioning and biodiversity of fungi in tropical forests? *Commun Integr Biol* 4:598–602.
75. Wang JB, St Leger RJ, Wang C. 2016. Advances in genomics of insect pathogenic fungi. *Adv Genet* 94:67–105. <https://doi.org/10.1016/bs.adgen.2016.01.002>.
76. Corradi N, Haag KL, Pombert JF, Ebert D, Keeling PJ. 2009. Draft genome sequence of the *Daphnia* pathogen *Ooctosporea bayeri*: insights into the gene content of a large microsporidian genome and a model for host-parasite interactions. *Genome Biol* 10:R106. <https://doi.org/10.1186/gb-2009-10-10-r106>.
77. Quan QM, Chen LL, Wang X, Li S, Yang XL, Zhu YG, Wang M, Cheng Z. 2014. Genetic diversity and distribution patterns of host insects of caterpillar fungus *Ophiocordyceps sinensis* in the Qinghai-Tibet Plateau. *PLoS One* 9:e92293. <https://doi.org/10.1371/journal.pone.0092293>.
78. Wang XL, Yao YJ. 2011. Host insect species of *Ophiocordyceps sinensis*: a review. *Zookeys* 127:43–59. <https://doi.org/10.3897/zookeys.127.802>.
79. Zhang Y, Zhang S, Li Y, Ma S, Wang C, Xiang M, Liu X, An Z, Xu J, Liu X. 2014. Phylogeography and evolution of a fungal-insect association on the Tibetan Plateau. *Mol Ecol* 23:5337–5355. <https://doi.org/10.1111/mec.12940>.
80. Peterson SW, Lichtwardt RW. 1983. *Capniomyces stellatus* and *Simulio-myces spica*: new taxa of Harpellales (Trichomycetes) from winter-emerging stoneflies. *Mycologia* 75:242–250. <https://doi.org/10.2307/3792808>.
81. White MM, Siri A, Lichtwardt RW. 2006. Trichomycete insect symbionts in Great Smoky Mountains National Park and vicinity. *Mycologia* 98:333–352.
82. Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428:617–624. <https://doi.org/10.1038/nature02424>.
83. Sinha S, Flibotte S, Neira M, Formby S, Plemenitaš A, Cimerman NG, Lenassi M, Gostinčar C, Stajich JE, Nislow C. 2017. Insight into the recent genome duplication of the halophilic yeast *Hortaea werneckii*: combining an improved genome with gene expression and chromatin structure. *G3* 7:2015–2022. <https://doi.org/10.1534/g3.117.040691>.
84. Ma LJ, Ibrahim AS, Skory C, Grabherr MG, Burger G, Butler M, Elias M, Iduurm A, Lang BF, Sone T, Abe A, Calvo SE, Corrochano LM, Engels R, Fu J, Hansberg W, Kim JM, Kodira CD, Koehrsen MJ, Liu B, Miranda-Saavedra D, O'Leary S, Ortiz-Castellanos L, Poulter R, Rodriguez-Romero J, Ruiz-Herrera J, Shen YQ, Zeng Q, Galagan J, Birren BW, Cuomo CA, Wickes BL. 2009. Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet* 5:e1000549. <https://doi.org/10.1371/journal.pgen.1000549>.
85. Corrochano LM, Kuo A, Marcet-Houben M, Polaino S, Salamov A, Villalobos-Escobedo JM, Grimwood J, Álvarez MI, Avalos J, Bauer D, Benito EP, Benoit I, Burger G, Camino LP, Cánovas D, Cerdá-Olmedo E, Cheng JF, Domínguez A, Eliáš M, Eslava AP, Glaser F, Gutiérrez G, Heitman J, Henrissat B, Iturriaga EA, Lang BF, Lavín JL, Lee SC, Li W, Lindquist E, López-García S, Luque EM, Marcos AT, Martin J, McCluskey K, Medina HR, Miralles-Durán A, Miyazaki A, Muñoz-Torres E, Oguiza JA, Ohm RA, Olmedo M, Orejas M, Ortiz-Castellanos L, Pisabarro AG, Rodríguez-Romero J, Ruiz-Herrera J, Ruiz-Vázquez R, et al. 2016. Expansion of signal transduction pathways in Fungi by extensive genome duplication. *Curr Biol* 26:1577–1584. <https://doi.org/10.1016/j.cub.2016.04.038>.
86. Stajich JE. 2016. Fungal evolution: *Mucor* and *Phycomyces* see double. *Curr Biol* 26:R775–RR777. <https://doi.org/10.1016/j.cub.2016.06.049>.
87. Goffeau AA, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Tettelin H, Oliver SG. 1996. Life with 6000 genes. *Science* 274:546, 563–546, 567. <https://doi.org/10.1126/science.274.5287.546>.
88. Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387:708–713. <https://doi.org/10.1038/42711>.
89. Marcet-Houben M, Gabaldón T. 2015. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol* 13:e1002220. <https://doi.org/10.1371/journal.pbio.1002220>.
90. Wolfe KH. 2015. Origin of the yeast whole-genome duplication. *PLoS Biol* 13:e1002221. <https://doi.org/10.1371/journal.pbio.1002221>.
91. McCreadie JW, Beard CE. 2003. The microdistribution of the trichomycete *Smittium culisetae* in the hindgut of the black fly host *Simulium vittatum*. *Mycologia* 95:998–1003. <https://doi.org/10.1080/15572536.2004.11833015>.
92. Albertin W, Marullo P. 2012. Polyploidy in fungi: evolution after whole-genome duplication. *Proc Biol Sci* 279:2497–2509. <https://doi.org/10.1098/rspb.2012.0434>.
93. Boisvert S, Laviolette F, Corbeil J. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol* 17:1519–1533. <https://doi.org/10.1089/cmb.2009.0238>.
94. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
95. Benson G. 1999. Tandem repeats finder: a program to analyse DNA sequences. *Nucleic Acids Res* 27:573–580. <https://doi.org/10.1093/nar/27.2.573>.
96. Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11. <https://doi.org/10.1186/s13100-015-0041-9>.
97. Keller O, Kollmar M, Stanke M, Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* 27:757–763. <https://doi.org/10.1093/bioinformatics/btr010>.
98. Ter-Ter-Hovhannisyán V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res* 18:1979–1990. <https://doi.org/10.1101/gr.081612.108>.
99. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Rvís J, White O, Robin CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 9:R7. <https://doi.org/10.1186/gb-2008-9-1-r7>.
100. Finn RD, Cogill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285. <https://doi.org/10.1093/nar/gkv1344>.
101. Rawlings ND, Morton FR. 2008. The Merops batch BLAST: a tool to detect peptidases and their non-peptidase homologues in a genome. *Biochimie* 90:243–259. <https://doi.org/10.1016/j.biochi.2007.09.014>.
102. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henriks B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:D490–D495. <https://doi.org/10.1093/nar/gkt1178>.
103. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
104. The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45:D158–D169. <https://doi.org/10.1093/nar/gkw1099>.
105. Coleine C, Masonjones S, Selbmann L, Zucconi L, Onofri S, Pacelli C, Stajich JE. 2017. Draft genome sequences of the Antarctic endolithic fungi *Rachicladosporium antarcticum* CCFEE 5527 and *Rachicladospo-*



- rium* sp. CCFEE 5018. *Genome Announc* 5:e00397-17. <https://doi.org/10.1128/genomeA.00397-17>.
106. Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067. <https://doi.org/10.1093/bioinformatics/btm071>.
  107. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786. <https://doi.org/10.1038/nmeth.1701>.
  108. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>.
  109. Urban M, Cuzick A, Rutherford K, Irvine A, Pedro H, Pant R, Sadanadan V, Khamari L, Billal S, Mohanty S, Hammond-Kosack KE. 2017. PHI-base: a new interface and further additions for the multi-species pathogen-host interactions database. *Nucleic Acids Res* 45:D604–D610. <https://doi.org/10.1093/nar/gkw1089>.
  110. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30:1575–1584. <https://doi.org/10.1093/nar/30.7.1575>.
  111. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
  112. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
  113. Kück P, Longo GC. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool* 11:81. <https://doi.org/10.1186/s12983-014-0081-x>.
  114. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Le SV. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>.
  115. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>.
  116. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
  117. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
  118. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
  119. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>.
  120. Altmann A, Weber P, Bader D, Preuss M, Binder EB, Müller-Myhsok B. 2012. A beginners guide to SNP calling from high-throughput DNA-sequencing data. *Hum Genet* 131:1541–1554. <https://doi.org/10.1007/s00439-012-1213-z>.
  121. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* 5:R12. <https://doi.org/10.1186/gb-2004-5-2-r12>.
  122. Gaujoux R, Seoighe C. 2010. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 11:367. <https://doi.org/10.1186/1471-2105-11-367>.
  123. Blum T, Briesemeister S, Kohlbacher O. 2009. MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC Bioinformatics* 10:274. <https://doi.org/10.1186/1471-2105-10-274>.
  124. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>.
  125. Koboldt D, Zhang Q, Larson D, Shen D, McLellan M, Lin L, Miller C, Mardis E, Ding L, Wilson R. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* <https://doi.org/10.1101/gr.129684.111>.