

# Revealing the genome of the microsporidian *Vairimorpha bombi*, a potential driver of bumble bee declines in North America

Victoria L. Webster,<sup>1</sup> Samuel Hemmings ,<sup>2</sup> Marta Pérez,<sup>1</sup> Matthew C. Fisher ,<sup>2</sup> Mark J.F. Brown,<sup>1,\*</sup> Rhys A. Farrer <sup>3,\*</sup>

<sup>1</sup>Department of Biological Sciences, Royal Holloway University of London, London TW20 0EX, UK

<sup>2</sup>MRC Centre for Global Infectious Disease Analysis, Imperial College London, London W2 1PG, UK

<sup>3</sup>MRC Centre for Medical Mycology, University of Exeter, Exeter EX4 4QD, UK

\*Corresponding authors: Mark J. F. Brown, Department of Biological Sciences, Royal Holloway University of London, Egham Hill, Egham, Surrey TW20 0EX, UK. Email: Mark.Brown@rhul.ac.uk; Rhys A. Farrer, MRC Centre for Medical Mycology, University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter EX4 4QD, UK. Email: r.farrer@exeter.ac.uk

Pollinators are vital for food security and the maintenance of terrestrial ecosystems. Bumblebees are important pollinators across northern temperate, arctic, and alpine ecosystems, yet are in decline across the globe. *Vairimorpha bombi* is a parasite belonging to the fungal class Microsporidia that has been implicated in the rapid decline of bumblebees in North America, where it may be an emerging infectious disease. To investigate the evolutionary basis of pathogenicity of *V. bombi*, we sequenced and assembled its genome using Oxford Nanopore and Illumina technologies and performed phylogenetic and genomic evolutionary analyses. The genome assembly for *V. bombi* is 4.73 Mb, from which we predicted 1,870 protein-coding genes and 179 tRNA genes. The genome assembly has low repetitive content and low GC content. *V. bombi*'s genome assembly is the smallest of the *Vairimorpha* and closely related *Nosema* genera, but larger than those found in the *Encephalitozoon* and *Ordospora* sister clades. Orthology and phylogenetic analysis revealed 18 core conserved single-copy microsporidian genes including the histone acetyltransferase (HAT) *GCN5*. Surprisingly, *V. bombi* was unique to the microsporidia in not encoding the second predicted HAT *ESA1*. The *V. bombi* genome assembly annotation included 265 unique genes (i.e. not predicted in other microsporidia genome assemblies), 20% of which encode a secretion signal, which is a significant enrichment. Intriguingly, of the 36 microsporidian genomes we analyzed, 26 also had a significant enrichment of secreted signals encoded by unique genes, ranging from 6 to 71% of those predicted genes. These results suggest that microsporidia are under selection to generate and purge diverse and unique genes encoding secreted proteins, potentially contributing to or facilitating infection of their diverse hosts. Furthermore, *V. bombi* has 5/7 conserved spore wall proteins (SWPs) with its closest relative *V. ceranae* (that primarily infects honeybees), while also uniquely encoding four additional SWPs. This gene class is thought to be essential for infection, providing both environmental protection and recognition and uptake into the host cell. Together, our results show that SWPs and unique genes encoding a secretion signal are rapidly evolving in the microsporidia, suggesting that they underpin key pathobiological traits including host specificity and pathogenicity.

**Keywords:** *Vairimorpha bombi*; microsporidia; genome assembly; secretome; spore wall proteins; histone acetyltransferase

## Introduction

Microsporidia are globally ubiquitous and obligate pathogens that infect and subsequently impact the health of a wide range of animal species. Infection can be caused by both vertical and horizontal transmission, with the most common mode in humans (causing the disease Microsporidiosis) via zoonosis or ingestion of spores in food or water (Han, Takvorian, et al. 2020). Microsporidia infect their hosts by adherence of the spore wall proteins (SWPs) to the host cell surface prior to spore activation (Southern et al. 2007). Next, the spore is activated through polarization and changes to the spore wall, causing the polar tube to rapidly discharge out of the spore and pierce the host cell membrane, where it serves as a conduit for sporoplasm passage into the new host cell (Xu and Weiss 2005). The genomes of numerous

species of Microsporidia have been sequenced, in order to understand both their evolution and their pathogenicity, revealing that they possess highly reduced genomes and consequently reduced numbers of protein-coding genes (Jespersen et al. 2022).

Microsporidia can have impacts beyond host health when they infect species that provide essential ecosystem services. Pollination, which is predominantly carried out by bees, is essential for crop production and ecosystem health (IPBES 2016). Honeybees and bumblebees, which provide the majority of pollination services, are infected by at least three species of Microsporidia—*Vairimorpha apis*, *V. bombi*, and *V. ceranae*—all of which impact host health (Grupe and Quandt 2020). The genomes of the predominantly honeybee-infecting microsporidia (*V. apis* and *V. ceranae*) have been sequenced (Chen et al. 2013; Huang

Received on 07 January 2024; accepted on 04 February 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

et al. 2021), revealing important features of their genome evolution, including reduced genome length, reduced number of protein-coding genes, and loss of the RNAi protein Dicer. Indeed, Dicer has been reported to be lost in approximately half of all microsporidian species (Huang et al. 2021), where it otherwise would be involved in RNAi-driven silencing of transposons and viruses. *V. ceranae* has also been found in bumblebees, despite being primarily a honeybee-infecting parasite (Arbulo et al. 2015). Further comparative genomics work is needed to understand the evolutionary history and selective pressures that are acting on these *Vairimorpha* species, and this requires sequencing of the *V. bombi* genome.

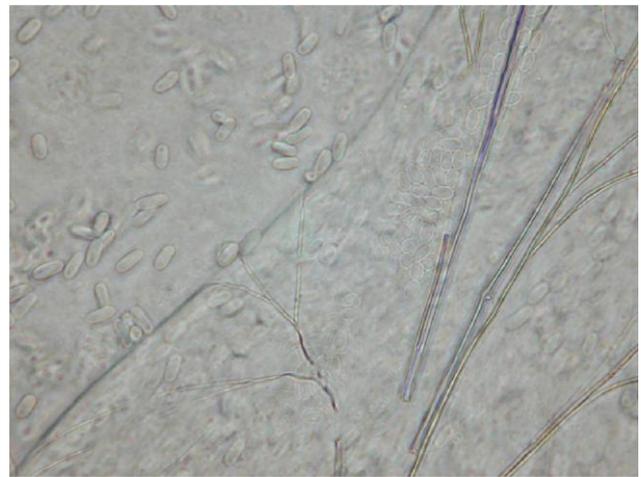
*Vairimorpha bombi* (previously known as *Nosema bombi*; Tokarev et al. 2020) is an obligate fungal parasite of bumblebees in the Microsporidia family and has been endemic in European bumblebees for over a century (Fantham and Porter 1914). It significantly reduces colony fitness in its European bumblebee hosts (Otti and Schmid-Hempel 2007; Rutrecht and Brown 2009), and thus is likely to play an important role in bumblebee population dynamics. However, recently *V. bombi* has been proposed to be the cause of massive population declines in bumblebee species native to the United States (Whittington and Winston 2004). More specifically, it has been suggested that commercial bumblebee production facilities resulted in the spillover of European strains of *V. bombi* into North American bumblebees and that this spillover drove collapses in the populations of numerous native bumblebees (Thorp and Shepherd 2005). Analyses of temporal patterns of bumblebee decline and *V. bombi* prevalence provide support for this hypothesis (Cameron et al. 2011, 2016). However, the small subunit rRNA from *V. bombi* collected in Europe and the United States in 2010–2011 was genetically indistinct from North American museum samples collected from 1979 to 2011 (Cameron et al. 2016), leaving the origin of this putative spillover event unclear.

The genome of *V. bombi* has not yet been sequenced, but the development of a high-quality annotated assembly for *V. bombi* is essential to assess the risk it poses to global bumblebee populations. Such an assembly would facilitate a comparative genomics study exploring the pathogenicity and host specificity of bumblebee-infecting microsporidia, as well as whether a cross-continental spillover event is behind the collapse in North American bumblebee populations. Furthermore, an assembly is essential for future epidemiological investigations that may reveal mitigation strategies to limit the impact and continuing spread of *V. bombi*. Here, we have sequenced and assembled the *V. bombi* genome and used comparative genomics across the wider family of Microsporidia to predict the evolutionary origin of the species and the genetic causes of its pathogenicity and host specificity.

## Methods

### Spore purification

A wild *B. terrestris* queen that was naturally infected with *V. bombi* (strain VR1) was collected from Windsor Great Park, UK (SU992703) in 2016 (Folly et al. 2020) and used to inoculate eight *B. terrestris* colonies (Folly et al. 2021) which were then stored at  $-80^{\circ}\text{C}$  until use. The abdomens from 200 of these infected workers were individually homogenized in 2 ml of 0.01 M  $\text{NH}_4\text{Cl}$  to inhibit spore germination (Undeen and Avery 1988; Fries et al. 2013) and strained through layered cheesecloth to remove large particles. The spores were then isolated using density gradient centrifugation (10.11G for 10 minutes at  $4^{\circ}\text{C}$ , after which  $\text{NH}_4\text{Cl}$  was removed and replaced, with this process being repeated until the pellet was pure white) and resuspended in sterile  $\text{NH}_4\text{Cl}$ . Individual suspensions were pooled into two samples, each



**Fig. 1.** A photograph of a squash of Malpighian tubules from a bumblebee infected with *Vairimorpha bombi* at 400 $\times$  magnification under phase contrast, showing spores both within and outside the tubules. © Mark J. F. Brown.

containing spores from 100 bees, and their final concentrations were quantified using a hemocytometer (Neubauer cell-counter chamber) under 400 $\times$  magnification. Both suspensions were then pelleted again as before.

### DNA extraction and sequencing

To extract spore DNA for short-read sequencing, the pellet was suspended in yeast cell lysis solution (Biosearch technologies) with 1.0 mm Zirconia/Silica beads (Thistle Scientific) and disrupted in a TissueLyser II at 30 m/s for 45 seconds six times to mechanically break the spore walls. We then centrifuged the sample at 20.82G for two minutes, added 1  $\mu\text{l}$  of RNase A, incubated at  $65^{\circ}\text{C}$  for 15 minutes, and then incubated on ice for 15 minutes. The DNA was then extracted and purified using the DNeasy Blood and Tissue Kit (Qiagen) and quantified using a Qubit dsDNA BR Assay (Qubit). We obtained 223 ng/ $\mu\text{l}$  in 38  $\mu\text{l}$  from  $1.29 \times 10^7$  cells which we split into six tubes, each containing 25 ng/ $\mu\text{l}$  in 25  $\mu\text{l}$ . Three tubes were sent to Novogene (Cambridge, UK) for sequencing using the Illumina HiSeq 4000 (San Diego, CA) using microbial whole genome library preparation (350 bp), and the remaining three were sent to the Earlham Institute (Norwich, UK) for sequencing using the Illumina NovaSeq 6000 (San Diego, CA) with reagent kit SP v1.5.

Long-read sequencing was performed on an Oxford Nanopore MinION flow cell. We minimized shearing of the DNA by encouraging the spores to germinate rather than breaking open the spore wall. We did this by desiccating the pellet of  $4.37 \times 10^7$  spores overnight at room temperature and then re-suspending them in 0.01 M Ringer's solution and incubating them for 3 hours at  $30^{\circ}\text{C}$  to simulate the environment inside the bee. We then extracted the DNA from this suspension using the DNeasy Plant Mini Kit (Qiagen). From this, we obtained 62 ng/ $\mu\text{l}$  of DNA in 18  $\mu\text{l}$  which we prepared for sequencing using the library protocol SQK-LSK112. This DNA was loaded into a flow cell and sequenced on an Oxford Nanopore MinION flow cell (FLO-MIN106D; Oxford Nanopore Technologies) for 58 hours in total over the course of three runs, with no adaptive sampling. We set a Phred quality score (Q-score) threshold of 8 and a minimum read length of 200 bp.

### Genome assembly and annotation

We removed lambda sequences from the Nanopore reads added in the library preparation step using NanoLyse version 1.1.0

(De Coster *et al.* 2018). We trimmed the Nanopore adapters from each read using Porechop version 0.2.4 (Wick *et al.* 2017). Quality filtering was achieved using NanoFilter version 2.3.0 (De Coster *et al.* 2018) with a quality score of  $\geq 5$  and a minimum read length of  $\geq 1,000$  base pairs. We searched the raw Illumina and Nanopore reads for species-specific primer sequences used for the molecular identification of *V. apis*, *V. ceranae*, and *V. bombi* before starting any analysis and only found evidence of *V. bombi*. Similarly, we aligned all of the quality-filtered Nanopore reads to the *B. terrestris* genome (GenBank accession ID: GCA\_910591885.2) using BWA-mem version 0.7.17-r1188 (Li 2013) and removed all reads which matched bumblebee DNA (~95% of the reads). As an additional filtering step, we screened a preliminary assembly using Canu version 2.2 (Koren *et al.* 2017) (the same steps as described next) through NCBI MegaBLAST (Morgulis *et al.* 2008) and filtered Nanopore reads that matched bacterial sequences using default settings.

All quality-filtered and non-bumblebee Nanopore reads were assembled using Canu version 2.2 (Koren *et al.* 2017) with parameters corOutCoverage = 100, corMhapSensitivity = high, minInput Coverage = 0, corMinCoverage = 0, and the predicted genome size as 10 Mb, which is approximately the same length of the other two bee-infecting microsporidia—*V. apis* (8.5 Mb; Chen *et al.* 2013) and *V. ceranae* (7.9–8.8 Mb; Cormman *et al.* 2009; Huang *et al.* 2021).

We trimmed adapter sequences from the Illumina reads using Trimmomatic (Bolger *et al.* 2014) and aligned them to the draft Nanopore assembly using BWA-mem (Li 2013). We down sampled and normalized the combined aligned reads to ~120 $\times$  coverage using bbnorm from the BBTools suite version 39.01 (Bushnell 2014) and then iteratively polished the assembly three times using Pilon version 1.24 (Walker *et al.* 2014). We then ran through the assembly and polishing process again using these filtered reads and made final sequence adjustments according to the NCBI database requirements. We evaluated our assembly using the quality assessment tool QUAST v5.2.0 (Gurevich *et al.* 2013).

We annotated the genome using Braker2 version 2.1.6 (Brüna *et al.* 2021) with esmode (for gene prediction with genome sequence data only) and softmasking enabled which combines GeneMark-ET version 3.68 (Lukashin and Borodovsky 1998) and AUGUSTUS version 3.5.0 (Stanke *et al.* 2006). From the resulting annotation, we generated protein and coding sequence files using gffread (Pertea and Pertea 2020).

To assess annotation completeness, we searched the *V. bombi* VR1 predicted protein sequences and the predicted proteins from an additional 35 microsporidia genome assemblies (Supplementary Table 1) for universal Benchmarking Universal Single-Copy Orthologs (BUSCOs) version 4.1.2 (Manni *et al.* 2021) specific to the microsporidia (microsporidia\_odb10). BLASTn (gene sequence) and tblastn (protein sequence) were used to search for *V. ceranae* BRL HAT1 against the *V. bombi* genome ( $E < 1e-5$ ). tblastn was used to search for the 265 uniquely identified genes from orthology clustering against every other microsporidian genome.

Gene ontology mapping for *V. bombi* VR1 was achieved using Blast2Go (Conesa *et al.* 2005) and PFAM domains and KEGG classifications were assigned to each gene using InterPro (Hunter *et al.* 2009). We predicted the presence of signal peptides using SignalP4 (Petersen *et al.* 2011). tRNAs were predicted using tRNAscan-SE eufindtRNA version 1.1 (Lowe and Eddy 1997), and rRNAs using RNAmmer version 1.2 (Lagesen *et al.* 2007). We identified repeated elements in the assembly with RepeatMasker version 4.0.8 (Smit *et al.* 2013–2015) using a *de novo* repeat library generated with RepeatModeler version 1.0.11 (Flynn *et al.* 2020) with LTRStruct.

## Orthology prediction and phylogenetics

Orthologs were predicted between our new *V. bombi* VR1 annotation and the annotation for 36 microsporidia species (1 strain representative with the highest  $N_{50}$  and contiguity) available in MicrosporidiaDB (Aurrecochea *et al.* 2011) and NCBI using the Synima pipeline (Farrer 2017) with OrthoMCL (Li *et al.* 2003) (Supplementary Table 1). Specifically, the Synima pipeline outputs Orthogroups, which can be divided into categories of interest, including single-copy orthologs, which we used to construct phylogenetic trees. We also separated out Orthogroups that had genes from every microsporidia species apart from *V. bombi* (those that we describe are uniquely lost in *V. bombi*) and those that are only found in a given species (those that we describe as unique to a given species). The script (Ortholog\_dist\_per\_genome\_summary.pl) used to separate out these categories of interest has now been added to the Synima Github repository.

We generated phylogenetic trees using 18 microsporidia single-copy orthologs. Briefly, Orthogroups were aligned using MUSCLE version 2.1.11 (Edgar 2004) and trimmed using trimAl version v1.4.rev15 (Capella-Gutiérrez *et al.* 2009) with the parameters -w 3 -gt 0.95 -st 0.01. We concatenated the alignments into a single FASTA using a bespoke script and converted that into nexus format. We used ProtTest version 3.4.2 (Darriba *et al.* 2011) to identify the best-fitting model for protein evolution (GAMMAILGF) and used that to generate a maximum likelihood phylogenetic tree using RAXML version 7.7.8 (Stamatakis 2014) with 1,000 bootstraps.

## Selection and enrichment analyses

$dN/dS$  using the Yang and Nielsen method (Yang and Nielsen 2000) was calculated for each of the five conserved SWPs in *V. bombi* compared with *V. ceranae* using the Prank v.170427 (Löytynoja 2021) multiple alignment tool, and PAML version 4.10.6 Codeml (Yang 2007). GO-term enrichment analyses were conducted using the two-tailed Fisher's exact test with q-value FDR. Multiple testing corrections were achieved using the Storey–Tibshirani method (Storey and Tibshirani 2003) (requiring q values of  $< 0.05$ ). Hypergeometric tests were conducted using the R phyper function to test for enrichment of genes with secretion signals among orthogroups unique to a single species.

## Results

### 5.1 Genome assembly of *V. bombi* VR1

We sequenced and assembled the genome of *V. bombi* (isolate name VR1) using Oxford Nanopore and Illumina technologies. The genome assembly of *V. bombi* VR1 that we generated in this study comprises 57 contigs and has a total length of 4.73 Mb with an  $N_{50}$  of 216 kb and with a low GC content (28.7%) (Table 1). From this genome assembly, we performed gene predictions, which identified 1,870 protein-coding genes and 179 tRNA genes. The protein-coding genes predicted encompass 2.23 Mb of sequence (~61% of the genome), while repetitive content comprises only a small portion of the genome (44 kb; 1.2% of the genome). The remainder of the genome assembly is intergenic (985 kb; 27%) and intronic (449 kb; 12%). Gene completeness analysis using the BUSCO approach with the microsporidia\_odb10 dataset revealed a completeness score of 90%. This BUSCO score is similar to its closest relative *V. ceranae* (also 90%) and with fewer duplicated complete BUSCOs (only 4 found in *V. bombi*; 27 found in *V. ceranae*). Compared with 35 additional microsporidia species, we found that the *V. bombi* genome assembly is the smallest of the

*Vairimorpha* and *Nosema* genera, but larger than those found in the *Encephalitozoon* and *Ordospora* sister clades, and that our genome assembly quality is similar or better than those of its closest relatives (Fig. 2).

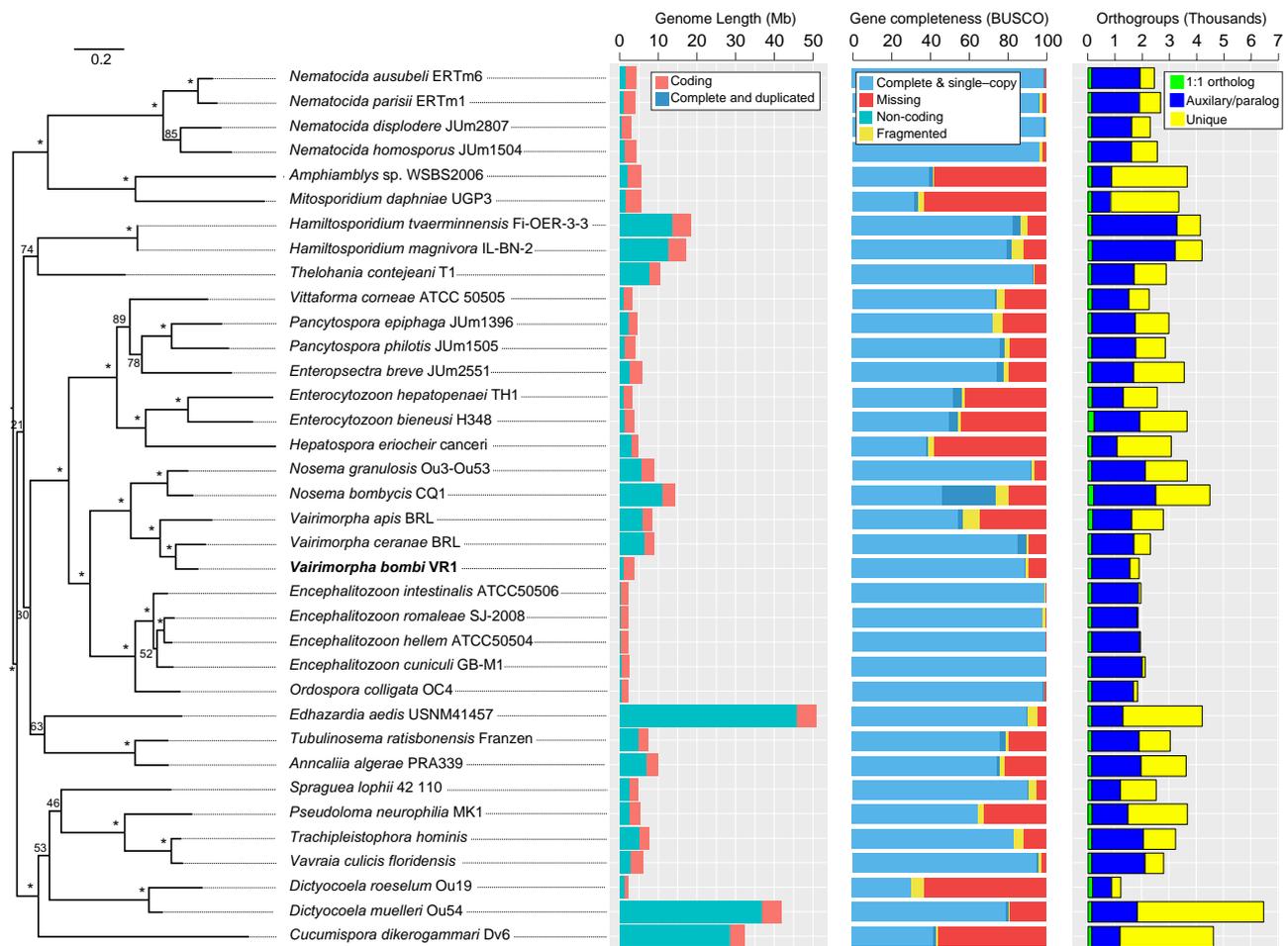
**Table 1.** *V. bombi* VR1 genome and annotation summary. Mb = Megabase, kb = kilobase, bp = basepair.

Assembly	
Total length (Mb)	4.73
Total length without N's (Mb)	3.66
GC content (%)	28.7
Contig N50 (kb)	216
Contig number	57
BUSCO completeness score (%)	96.8
Annotation	
Total length of repeats (bp)	43,850
Number of protein-coding genes	1,870
Total length of genes (Mb)	2.23
Mean gene length (bp)	1,189
Mean exon length (bp)	733
Mean intron length (bp)	386
Average exons per gene	1.6
BUSCO completeness score (%)	90

## Genome evolution, orthology, and phylogenetics of microsporidia

We identified 18 single-copy 1:1 orthologs that were conserved across all 36 species of microsporidia that we analyzed (Supplementary Tables 1 and 2). Phylogenetic analysis using these orthologs placed *V. ceranae* as the sister species to *V. bombi*, giving new insights into the relationships between these microsporidian parasites of bees. More broadly, our phylogenetic inference grouped the *Vairimorpha* species with the *Nosema* genus, which includes *N. granulosis*, a pathogen of the crustacean *Gammarus duebeni*, and *N. bombycis* which causes Pébrine or “pepper disease,” in silkworms. More distantly related is the *Encephalitozoon* genus that includes pathogens of diverse animal hosts including rabbits, humans, and grasshoppers. The *Encephalitozoon* clusters closely with *Ordospora colligata*, which is an intracellular parasite of the crustacean *Daphnia magna*.

As expected, the 18 highly conserved orthologs included several ribosomal proteins, eukaryotic translation initiation factor, protein kinases, kinesin, and perhaps less expectedly, the Histone acetyltransferase (HAT) GCN5. These orthologs were used to infer a maximum likelihood that supported the previous grouping of taxa including among and between *Vairimorpha*, *Nosema*, and *Encephalitozoon* (Fig. 2). Most of the genes across



**Fig. 2.** A maximum likelihood phylogenetic tree constructed using multiple alignment based on 18 single-copy core orthologs and RAXML (using the best-fitting model PROT + GAMMA + ILGF) with 1,000 bootstraps. To the right of the phylogenetic tree are three bar charts showing for each microsporidian genome assembly the genome length (separated into coding and non-coding), the BUSCO gene completeness score (separated into complete and single copy, complete and duplicated, fragmented and missing), and the thousands of predicted Orthogroups (separated into 1:1 core ortholog, auxiliary/paralogs and unique). \* Indicates 100% bootstrap support, while numbers indicate <100% bootstrap support.

the 36 microsporidia fell within auxiliary/paralog clusters, encompassing duplicated genes, some of which are conserved across multiple species. In contrast to all other microsporidian genome assemblies analyzed, orthogroup clustering indicated that the *V. bombi* VR1 genome does not encode the HAT ESA1 (e.g. *V. ceranae* gene = G9O61\_00G005020). To explore this gene absence further, we performed a tblastn of the *V. ceranae* gene to the *V. bombi* genome, which found two significant hits, neither of which were HAT ESA1. The first hit was to contig 2 (positions 245,128–246,240; E = 1.52e-169), which overlaps the protein-coding gene g363 (HrpA-like helicase) that clustered with genes from 34 other species including two genes from *N. bombycis*. The second hit was to contig 7 (positions 24,960–25,568; E = 5.26e-24) that overlaps the protein-coding gene g130 (HAT MST2) that clustered with genes from 20 other species including two genes from *N. bombycis*. These analyses suggest that ESA1 is absent in *V. bombi*, although other genes share sequence similarity with it.

Orthology clustering identified 265 unique genes in *V. bombi* VR1 (i.e. no orthology to other microsporidian species genes), which is fewer than the number of unique annotated genes in most other microsporidia genome assemblies. To validate and analyze these genes further, tblastn was used to search for sequence similarity in all 36 microsporidian genome assemblies (Supplementary Table 3). This analysis revealed that 158 genes (60%) had novel sequences and were not in any of the other sequenced microsporidian genome assemblies. A further 107 genes (40%) did have sequence similarity to  $\geq 1$  microsporidian genome assemblies, most of which ( $n = 96/107$ ; 90%) also overlapped annotated protein-coding genes. These unique genes are therefore likely to be mostly novel sequences, along with some paralogs or genes with sequence similarity with other genes and some inaccuracies associated with genome sequencing, assembly, gene annotation, and orthology clustering.

All 265 unique genes were searched for functional enrichment compared with the rest of the protein-encoded genes (1:1 core orthologs and auxiliary), identifying only two high-level GO term enrichments among the unique genes (cellular process and intracellular anatomical structure). Surprisingly, 51/265 (19%) of *V. bombi*'s unique genes encoded a secretion signal, which was significantly enriched, given that only 95 genes were predicted to be secreted in total across all 1,775 genes (hypergeometric test upper tail  $P$ -value = 7.5E-22). Intriguingly, of the 36 microsporidian genome assemblies we compared, 26 also had a significant enrichment ( $P < 0.01$ ) of secreted signals encoded by unique genes, ranging from 6% in *E. romaleae* to 71% in *E. aedis* (Supplementary Table 4).

### Dynamically evolving SWPs

We compared the repertoire of *V. bombi* VR1 genes to the SWPs and polar tube proteins of its closest relative *V. ceranae* (Huang et al. 2021) (Supplementary Table 5). *V. bombi* is predicted to encode most ( $n = 5/7$ ) of the SWP's described in *V. ceranae*. SWPs encoded by *V. bombi* include two genes that are unique to *V. ceranae* and *V. bombi* (SWP1 and 2), one gene that is unique to *V. ceranae*, *V. bombi*, and *V. apis* (SWP3), and two that are more deeply conserved among microsporidia (SWP9 found in 10/36 species and SWP12 found in 27/36 species).  $dN/dS$  ( $\omega$ ) for the five conserved SWPs in *V. bombi* VR1 compared with *V. ceranae* were between 0.06 and 0.25 suggesting that those genes are undergoing purifying selection (Table 2).

The two SWPs not encoded by *V. bombi* are SWP4, which is uniquely encoded by *V. ceranae*, and SWP7, which is more widely conserved (encoded by 27/36 species). Using the 5 *V. bombi* SWPs

**Table 2.**  $dN$ ,  $dS$ , and  $\omega$  ( $dN/dS$ ) for the five conserved spore wall proteins (SWPs) in *V. bombi* VR1 compared with *V. ceranae*.

Gene	$dN$	$dS$	$\omega$
HSWP1	0.5355	3.0194	0.1774
HSWP2	0.4695	3.5485	0.1323
HSWP3	0.5715	2.3068	0.2477
HSWP9	0.3952	4.0004	0.0988
HSWP12	0.2216	3.506	0.0632

as a BLASTp database, we searched for additional SWPs that were not orthologous with *V. ceranae* genes, revealing four additional SWPs that all matched to SWP2 with E-value  $< 1E-5$  (IDs = g414, g1028, g1030, g1482). Together, these analyses suggest dynamically evolving spore walls (in terms of gene presence and absence) among *Vairimorpha* species, including both conserved and uniquely encoded SWPs.

### Discussion

We present the first genome assembly for *Vairimorpha bombi*, which is 4.73 Mb and has low GC content and low repetitive richness. This genome assembly is the smallest of the *Vairimorpha* and *Nosema* genera, but larger than those found in the *Encephalitozoon* and *Ordospora* sister clades. We identified 2,049 genes (including 1,870 protein-coding genes and 179 tRNA genes), which is lower than average for the microsporidia ( $\bar{x} = 3,054$ ). Indeed, only five species had fewer predicted genes including *Dictyocoela roeselium* (1,201 genes), *Encephalitozoon romaleae* (1,883), *Ordospora colligata* (1,879), *Encephalitozoon hellem* (2,006), and *Encephalitozoon intestinalis* (2,011). The impact and extent of gene reduction in the microsporidia is poorly understood. Understanding the genome evolution of the microsporidia is hampered by poorly resolved draft genome assemblies owing to relying on short sequence reads, and a challenge of separating host and pathogen DNA. Indeed, even with Oxford Nanopore reads, our assembly is still fragmented and includes ambiguous bases, despite its small size. Despite these challenges, our long and short read-based hybrid assembly achieved a 90% BUSCO gene completeness metric, which is higher than the average across all species ( $\bar{x} = 78\%$ ).

Our phylogenetic analysis based on 18 conserved single-copy orthologs and previous phylogenetic trees suggests that microsporidian evolution consists of both congruent and incongruent phylogenies with their host species, suggesting frequent host-jumps, especially among human-infecting microsporidia (Park and Poulin 2021). Whole genome or ortholog trees, such as the one we generated here, have the potential for providing greater resolution to microsporidian evolution, and also providing future insight regarding their overall placement in the fungal kingdom.

The HAT GCN5 is a conserved ortholog across all 36 microsporidia analyzed. In other eukaryotes, GCN5 is a subunit of the SAGA (Spt-Ada-Gcn5 acetyltransferase) transcriptional coactivator complex, which catalyzes acetylation of histone H3 and H2B N-terminal tails, with a corresponding association with gene activation. Given that neither H3 nor H2B were identified as microsporidian core orthologs, this raises the question of what the function of microsporidian GCN5 might have. An intriguing hypothesis is that it functions as a host protein to disrupt or interfere with normal gene regulation, including innate and adaptive immune responses (Han, Gao, et al. 2020). Previously, it was suggested that Tra1p (a core component of the Ada and Spt proteins of the SAGA complex) is absent in microsporidia

(Miranda-Saavedra et al. 2007). Additionally, the LisH domain that is part of Taf5 (a subunit of the transcription initiation factor TFIID) is absent in microsporidia (Jespersen et al. 2022), where it normally interacts with the Spt20 subunit of the SAGA complex (Miranda-Saavedra et al. 2007). Therefore, microsporidian encoded GCN5 may function by itself, or apart from the SAGA complex. Our gene annotation and a BLASTn search of the *V. ceranae* BRL (closest relative) HAT1 gene to the *V. bombi* genome suggested that *V. bombi* has uniquely lost HAT ESA1 compared with all the other microsporidia species. In other fungi, such as *Saccharomyces cerevisiae*, ESA1 is essential for growth (Smith et al. 1998). It is unclear how or why *V. bombi* would dispense with this gene that is conserved across all other microsporidian and even *Vairimorpha* species. Further work should explore the functional roles that conserved microsporidian HATs have and the cellular locations of those functions.

We discovered a significant enrichment of secretion signals among genes that were predicted to be uniquely encoded by *V. bombi*, and indeed a significant enrichment of uniquely encoded genes that also encode a secretion signal in the majority of microsporidian species ( $n = 26/36$ ; 72%). Secreted proteins have been shown to be important for infection, including, for example, the nematode-specific protein AAIM-1, whose loss of function confers resistance during *Nematocida parisii* infection. AAIM-1 is secreted into the intestinal lumen where it ensures spore orientation during intestinal cell invasion, as well as limiting bacterial colonization (Tamim El Jarkass et al. 2022). The function of all 2,818 total unique secreted proteins across the microsporidia that we identified is likely to be key for infection or competition and requires further study.

The SWPs and polar tube proteins are among the best-studied microsporidian virulence factors. Our genomic comparisons with *V. ceranae* suggested that most of the genes involved in the spore wall and polar tube are conserved (Huang et al. 2021). However, we also identified several SWPs that appear to be lost in many microsporidian species (e.g. SWP9 is encoded and conserved in 10/36 species, while SWP12 is found in 27/36 species). The two SWPs not encoded by *V. bombi* are SWP4, which is uniquely encoded by *V. ceranae*, and SWP7 that is conserved more widely (encoded by 27/36 species). We identified four additional putative SWPs that had sequence similarity to SWP2 (IDs = g414, g1028, g1030, g1482). Together, these analyses suggest a dynamically evolving spore wall among *Vairimorpha* species, including both conserved and uniquely encoded SWPs, which may be a signal of coevolution with their hosts.

Future work should aim to classify SWPs and genes that encode secretion signals according to their orthology, detailing how gene duplications or new genes have emerged in each genus or species. Additional genome sequencing, assembly, and annotation are also required, including efforts to improve current assemblies with additional long-read sequencing or short-read polishing to ensure repeat-content and gene annotation are accurately represented and fairly compared. Here, our sequencing, assembly, and comparative genomics of *V. bombi* have revealed unique genomic features of the pathogen, along with general features of genome evolution, both of which provide clues about the mechanisms underpinning microsporidian infection and host specificity.

## Data availability

The raw Illumina and Nanopore reads were submitted to NCBI GenBank (accession ID: SAMN33783333 and SAMN33864217). The annotated genome was submitted to NCBI (accession ID:

PRJNA945438) and FigShare (10.6084/m9.figshare.24354313). Code used for genome assembly and analysis is provided on Github ([https://github.com/VikiBlanchard/2022-23\\_Vairimorpha\\_bombi\\_de\\_novo\\_genome](https://github.com/VikiBlanchard/2022-23_Vairimorpha_bombi_de_novo_genome)).

Supplemental material available at G3 online.

## Funding

VW was supported by a BBSRC DTP PhD Studentship. RAF is part of the Medical Research Council Centre for Medical Mycology MR/N006364/2. RAF is supported by a Wellcome Trust Career Development Award (225303/Z/22/Z). MCF is a member of the Medical Research Council Centre for Global Infectious Disease Analysis MR/MR/R015600/1. MCF is funded by the CIFAR Fungal Kingdoms program.

## Conflicts of interest

The author(s) declare no conflicts of interest.

## Literature cited

- Arbulo N, Antúnez K, Salvarrey S, Santos E, Branchiccela B, Martín-Hernández R, Higes M, Invernizzi C. 2015. High prevalence and infection levels of *Nosema ceranae* in bumblebees *Bombus atratus* and *Bombus bellicosus* from Uruguay. *J Invertebr Pathol.* 130:165–168. doi:10.1016/j.jip.2015.07.018.
- Aurrecochea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, Gajria B, Gao X, Gingle A, Grant G, et al. 2011. AmoebaDB and MicrosporidiaDB: functional genomic resources for amoebozoa and microsporidia species. *Nucleic Acids Res.* 39(Database): D612–D619. doi:10.1093/nar/gkq1006.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics.* 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.
- Brüna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform.* 3(1):lqaa108. doi:10.1093/nargab/lqaa108.
- Bushnell B. 2014. BMAP: A Fast, Accurate, Splice-Aware Aligner. United States. <https://www.osti.gov/biblio/1241166>
- Cameron SA, Lim HC, Lozier JD, Duennes MA, Thorp R. 2016. Test of the invasive pathogen hypothesis of bumble bee decline in North America. *Proc Natl Acad Sci U S A.* 113(16):4386–4391. doi:10.1073/pnas.1525266113.
- Cameron SA, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, Griswold TL. 2011. Patterns of widespread decline in North American bumble bees. *Proc Natl Acad Sci U S A.* 108(2): 662–667. doi:10.1073/pnas.1014743108.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 25(15):1972–1973. doi:10.1093/bioinformatics/btp348.
- Chen Yp, Pettis JS, Zhao Y, Liu X, Tallon LJ, Sadzewicz LD, Li R, Zheng H, Huang S, Zhang X, et al. 2013. Genome sequencing and comparative genomics of honey bee microsporidia, *Nosema apis* reveal novel insights into host-parasite interactions. *BMC Genom.* 14(1):451. doi:10.1186/1471-2164-14-451.
- 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(18): 3674–3676. doi:10.1093/bioinformatics/bti610.

- Cornman RS, Chen YP, Schatz MC, Street C, Zhao Y, Desany B, Egholm M, Hutchison S, Pettis JS, Lipkin WI, et al. 2009. Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. *PLoS Pathog.* 5(6):e1000466. doi:10.1371/journal.ppat.1000466.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics.* 27(8):1164–1165. doi:10.1093/bioinformatics/btr088.
- De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics.* 34(15):2666–2669. doi:10.1093/bioinformatics/bty149.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* 5(1):113. doi:10.1186/1471-2105-5-113.
- Fantham, Porter A. 1914. *Nosema bombi*. *Am J Trop Med.* 8:623–638. doi:10.15468/39omei.
- Farrer RA. 2017. Synima: a synteny imaging tool for annotated genome assemblies. *BMC Bioinform.* 18(1):507. doi:10.1186/s12859-017-1939-7.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 117(17):9451–9457. doi:10.1073/pnas.1921046117.
- Folly AJ, Koch H, Farrell IW, Stevenson PC, Brown MJF. 2021. Agri-environment scheme nectar chemistry can suppress the social epidemiology of parasites in an important pollinator. *Proc Biol Sci.* 288(1951):20210363. doi:10.1098/rspb.2021.0363.
- Folly AJ, Stevenson PC, Brown MJF. 2020. Age-related pharmacodynamics in a bumblebee-microsporidian system mirror similar patterns in vertebrates. *J Exp Biol.* 223(Pt 6):jeb217828. doi:10.1242/jeb.217828.
- Fries I, Chauzat M-P, Chen Y-P, Doublet V, Genersch E, Gisder S, Higes M, McMahon DP, Martín-Hernández R, Natsopoulou M, et al. 2013. Standard methods for *Nosema* research. *J Apic Res.* 52(1):1–28. doi:10.3896/IBRA.1.52.1.14.
- Grupe AC, Quandt CA. 2020. A growing pandemic: a review of *Nosema* parasites in globally distributed domesticated and native bees. *PLoS Pathog.* 16(6):e1008580. doi:10.1371/journal.ppat.1008580.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 29(8):1072–1075. doi:10.1093/bioinformatics/btt086.
- Han Y, Gao H, Xu J, Luo J, Han B, Bao J, Pan G, Li T, Zhou Z. 2020. Innate and adaptive immune responses against microsporidia infection in mammals. *Front Microbiol.* 11:1468. doi:10.3389/fmicb.2020.01468.
- Han B, Takvorian PM, Weiss LM. 2020. Invasion of host cells by microsporidia. *Front Microbiol.* 11:172. doi:10.3389/fmicb.2020.00172.
- Huang Q, Wu ZH, Li WF, Guo R, Xu JS, Dang XQ, Ma ZG, Chen YP, Evans JD. 2021. Genome and evolutionary analysis of *Nosema ceranae*: a microsporidian parasite of Honey Bees. *Front Microbiol.* 12:645353. doi:10.3389/fmicb.2021.645353.
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Res.* 37(Database):D211–D215. doi:10.1093/nar/gkn785.
- IPBES. 2016. The assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production. In: Potts SG, Imperatriz-Fonseca VL, Ngo HT, editors. Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn, Germany: Zenodo. doi:10.5281/zenodo.3402856.
- Jespersen N, Monroy L, Barandun J. 2022. Impact of genome reduction in microsporidia. In: Weiss LM, Reinke AW, editors. *Microsporidia: Current Advances in Biology*. Cham: Springer International Publishing. p. 1–42.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27(5):722–736. doi:10.1101/gr.215087.116.
- Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35(9):3100–3108. doi:10.1093/nar/gkm160.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio]. <https://doi.org/10.48550/arXiv.1303.3997>, preprint: not peer reviewed.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13(9):2178–2189. doi:10.1101/gr.1224503.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25(5):955–964. doi:10.1093/nar/25.5.955.
- Löytynoja A. 2021. Phylogeny-Aware alignment with PRANK and PAGAN. *Methods Mol Biol.* 2231:17–37. doi:10.1007/978-1-0716-1036-7\_2.
- Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* 26(4):1107–1115. doi:10.1093/nar/26.4.1107.
- Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: assessing genomic data quality and beyond. *Curr Protoc.* 1(12):e323. doi:10.1002/cpz1.323.
- Miranda-Saavedra D, Stark MJ, Packer JC, Vivares CP, Doerig C, Barton GJ. 2007. The complement of protein kinases of the microsporidium *Encephalitozoon cuniculi* in relation to those of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *BMC Genom.* 8(1):309. doi:10.1186/1471-2164-8-309.
- Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. 2008. Database indexing for production MegaBLAST searches. *Bioinformatics.* 24(16):1757–1764. doi:10.1093/bioinformatics/btn322.
- Otti O, Schmid-Hempel P. 2007. *Nosema bombi*: a pollinator parasite with detrimental fitness effects. *J Invertebr Pathol.* 96(2):118–124. doi:10.1016/j.jip.2007.03.016.
- Park E, Poulin R. 2021. Revisiting the phylogeny of microsporidia. *Int J Parasitol.* 51(10):855–864. doi:10.1016/j.ijpara.2021.02.005.
- Pertea G, Pertea M. 2020. GFF utilities: GffRead and GffCompare. *F1000Res.* 9:ISCB Comm J-304. doi:10.12688/f1000research.23297.1.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8(10):785–786. doi:10.1038/nmeth.1701.
- Rutrecht ST, Brown MJF. 2009. Differential virulence in a multiple-host parasite of bumble bees: resolving the paradox of parasite survival? *Oikos.* 118(6):941–949. doi:10.1111/j.1600-0706.2009.17392.x.
- Smit A, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. <https://www.repeatmasker.org/faq.html>.
- Smith ER, Eisen A, Gu W, Sattah M, Pannuti A, Zhou J, Cook RG, Lucchesi JC, Allis CD. 1998. ESA1 is a histone acetyltransferase that is essential for growth in yeast. *Proc Natl Acad Sci U S A.* 95(7):3561–3565. doi:10.1073/pnas.95.7.3561.
- Southern TR, Jolly CE, Lester ME, Hayman JR. 2007. Enp1, a microsporidian spore wall protein that enables spores to adhere to

- and infect host cells *in vitro*. *Eukaryot Cell*. 6(8):1354–1362. doi:[10.1128/EC.00113-07](https://doi.org/10.1128/EC.00113-07).
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9):1312–1313. doi:[10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033).
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res*. 34(Web Server):W435–W439. doi:[10.1093/nar/gkl200](https://doi.org/10.1093/nar/gkl200).
- Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*. 100(16):9440–9445. doi:[10.1073/pnas.1530509100](https://doi.org/10.1073/pnas.1530509100).
- Tamim El Jarkass H, Mok C, Schertzberg MR, Fraser AG, Troemel ER, Reinke AW. 2022. An intestinally secreted host factor promotes microsporidia invasion of *C. elegans*. *eLife*. 11:e72458. doi:[10.7554/eLife.72458](https://doi.org/10.7554/eLife.72458).
- Thorpe R, Shepherd M. 2005. Subgenus *Bombus* Latreille. In: Shepherd MD, Vaughan DM, Black SH, editors. 1802. Red List of Pollinator Insects of North America. Portland, OR: Xerces Society for Invertebrate Conservation. p. 5.
- Tokarev YS, Huang W-F, Solter LF, Malysh JM, Becnel JJ, Vossbrinck CR. 2020. A formal redefinition of the genera *Nosema* and *Vairimorpha* (Microsporidia: Nosematidae) and reassignment of species based on molecular phylogenetics. *J Invertebr Pathol*. 169:107279. doi:[10.1016/j.jip.2019.107279](https://doi.org/10.1016/j.jip.2019.107279).
- Undeen AH, Avery SW. 1988. Ammonium chloride inhibition of the germination of spores of *Nosema* algerae (Microspora: Nosematidae). *J Invertebr Pathol*. 52(2):326–334. doi:[10.1016/0022-2011\(88\)90142-5](https://doi.org/10.1016/0022-2011(88)90142-5).
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 9(11):e112963. doi:[10.1371/journal.pone.0112963](https://doi.org/10.1371/journal.pone.0112963).
- Whittington R, Winston ML. 2004. Comparison and examination of *Bombus occidentalis* and *Bombus impatiens* (Hymenoptera: Apidae) in tomato greenhouses. *J Econ Entomol*. 97(4):1384–1389. doi:[10.1093/jee/97.4.1384](https://doi.org/10.1093/jee/97.4.1384).
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom*. 3(10):e000132. doi:[10.1099/mgen.0.000132](https://doi.org/10.1099/mgen.0.000132).
- Xu Y, Weiss LM. 2005. The microsporidian polar tube: a highly specialised invasion organelle. *Int J Parasitol*. 35(9):941–953. doi:[10.1016/j.ijpara.2005.04.003](https://doi.org/10.1016/j.ijpara.2005.04.003).
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591. doi:[10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088).
- Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 17(1):32–43. doi:[10.1093/oxfordjournals.molbev.a026236](https://doi.org/10.1093/oxfordjournals.molbev.a026236).

Editor: A. Rokas