



Third Generation Genome Sequencing Reveals That Endobacteria in Nematophagous Fungi *Esteya vermicola* Contain Multiple Genes Encoding for Nematicidal Proteins

Ruizhen Wang^{1,2}, Leiming Dong^{1*}, Yuequ Chen³, Shuai Wang⁴ and Liangjian Qu^{2*}

OPEN ACCESS

Edited by:

Zhiyong Li,
Shanghai Jiao Tong University, China

Reviewed by:

Claudia S. L. Vicente,
University of Évora, Portugal
Shu-Jun Wei,
Beijing Academy of Agricultural
and Forestry Sciences, China
Chunyan Wang,
Northwest A&F University, China

*Correspondence:

Liangjian Qu
qujl2001@caf.ac.cn
Leiming Dong
dongleiming2008@126.com

Specialty section:

This article was submitted to
Microbial Symbioses,
a section of the journal
Frontiers in Microbiology

Received: 24 December 2021

Accepted: 28 March 2022

Published: 03 May 2022

Citation:

Wang R, Dong L, Chen Y, Wang S
and Qu L (2022) Third Generation
Genome Sequencing Reveals That
Endobacteria in Nematophagous
Fungi *Esteya vermicola* Contain
Multiple Genes Encoding for
Nematicidal Proteins.
Front. Microbiol. 13:842684.
doi: 10.3389/fmicb.2022.842684

¹ Beijing Floriculture Engineering Technology Research Centre, Beijing Botanical Garden, Beijing, China, ² The Key Laboratory of Forest Protection of National Forestry and Grassland Administration, Ecology and Nature Conservation Institute, Chinese Academy of Forestry, Beijing, China, ³ Forestry Resources Protection Institute, Jilin Provincial Academy of Forestry Sciences, Changchun, China, ⁴ School of Pharmacy, Liaocheng University, Liaocheng, China

Esteya vermicola is the first recorded endoparasitic nematophagous fungus with high infectivity capacity, attacking the pinewood nematode *Bursaphelenchus xylophilus* which causes pine wilt disease. Endosymbionts are found in the cytoplasm of *E. vermicola* from various geographical areas. We sequenced the genome of endobacteria residing in *E. vermicola* to discover possible biological functions of these widespread endobacteria. Multilocus phylogenetic analyses showed that the endobacteria form a previously unidentified lineage sister to *Phyllobacterium myrsinacearum* species. The number of genes in the endobacterium was 4542, with 87.8% of the proteins having a known function. It contained a high proportion of repetitive sequences, as well as more Acyl-CoA synthetase genes and genes encoding the electron transport chain, compared with plant-associated *P. zundukense* Tri 48 and *P. myrsinacearum* DSM 5893. Thus, this symbiotic bacterium is likely to be more efficient in regulating gene expression and energy release. Furthermore, the endobacteria in nematophagous fungi *Esteya vermicola* contained multiple nematicidal subtilase/subtilisin encoding genes, so it is likely that endobacteria cooperate with the host to kill nematodes.

Keywords: endobacteria, *Esteya vermicola*, pine wood nematode, nematode-killing capabilities, genome

INTRODUCTION

Pinewood nematode (*Bursaphelenchus xylophilus*) causes pine wilt disease, resulting in severe ecological and economic losses in Asia and Europe (Liou et al., 1999; Vicente et al., 2012; Futai, 2013). *Esteya vermicola* is endoparasitic nematophagous fungus with high infectivity capacity that attacks pinewood nematode. *Esteya vermicola* attracts pine wood nematode through specific volatile

compounds (α -pinene, β -pinene, and camphor) (Lin et al., 2013). The lunate adhesive spores of *E. vermicola*, whose concave surface has large surface area and thick adhesive layer, adhere to the cuticle of the attracted nematode and penetrate it. Then *E. vermicola* colonize in the body cavity and absorb the nematode's nutrient, and produce a large number of mycelia and lunate spores in preparation for next infection cycle (Wang et al., 2011, 2018).

Symbioses have played pivotal roles in biological, ecological, and evolutionary diversification (Archibald, 2014). Various symbiotic microorganisms provide a metabolic capability absent from the eukaryotes host, resulting in increased ecological amplitude and often evolutionary diversification of the eukaryotes host (Douglas, 2014). Symbioses between bacteria and insects have been researched in depth (Ferrari and Vavre, 2011). However, there are few studies on the function and genome of fungal endobacteria.

So far, seven fungal endobacteria genomes have been published, and they are *Candidatus Glomeribacter gigasporarum* colonizing arbuscular mycorrhizal fungi (AMF) *Gigaspora margarita* (Ghignone et al., 2012); *Burkholderia rhizoxinica* colonizing *Rhizopus microsporus*, which participates in the production of rhizoxin (Lackner et al., 2011); a betaproteobacterial endosymbiont associated with *Mortierella elongate* (Fujimura et al., 2014), mycoplasma-related endobacteria (MRE) living in *Rhizophagus clarus*, MRE living in *Racocetra verrucosa*, MRE living in *Claroideoglossum etunicatum* (Naito et al., 2015), MRE living in AMF *Dentiscutata heterogama* (Torres-Cortés et al., 2015).

Based on the sequenced genome of intracellular symbiotic bacteria, intracellular symbiotic bacteria have the following biological functions. *Ca. G. gigasporarum* expresses type II and type III secretion systems and synthesizes vitamin B12, antibiotics- and toxin-resistance molecules, that may contribute to the fungal host's ecological fitness (Ghignone et al., 2012). *Burkholderia rhizoxinica*, not its host fungus itself, which cause rice seedling blight, biosynthesize rhizoxin to destroy and kill rice (Partida-Martinez and Hertweck, 2005). The betaproteobacterial endosymbiont associated with *Mortierella elongate* encode the insecticidal toxin complex gene (Fujimura et al., 2014). Mycoplasma-related endobacteria genome suggests a high degree of adaptation to the fungal host (Naito et al., 2015; Torres-Cortés et al., 2015).

Esteya vermicola from various geographical locations, as endoparasitic nematophagous fungus, hosts endosymbionts, biotrophic endobacteria, in their cytoplasm (Wang et al., 2017). The endobacteria were coccoid, vertically hereditary, as yet uncultured, and essential symbionts (Wang et al., 2017). It was hypothesized that endobacteria play an important biological role in *E. vermicola*. We sequenced the genomes of endobacteria colonizing the *E. vermicola* CBS115803 to elucidate the endobacteria biology associated with nematocidal effect. We were particularly interested in determining the biological relevance of nematode-killing. We presented the annotation of the endobacteria genome draft, phylogenetic placement based on single copy genes, energy generation abilities, coevolution implications for nematode-killing abilities. Overall,

the endobacteria possibly exhibit a high adaptability to the fungal host for nematode-killing capabilities, showing the significance of endobacteria in the biology of *E. vermicola*.

MATERIALS AND METHODS

Biological Material and DNA Extraction

To acquire mycelium, *E. vermicola* CBS115803 was cultured in Potato sucrose medium. The Retsch Mixer Mill MM 400 was used to grind the mycelium at low temperature in order to break fungal cells and release intracellular symbiotic bacteria. The grounded mycelium were washed out with sterile 0.25M sucrose solution at 4°C. To remove ungrounded mycelium and a portion of the host's fungal nucleus, the mixture was centrifuged at 1200g. Then the supernatant was centrifuged with an angle rotor at 16000g for 30 min to obtain symbiotic bacteria. Endobacterial DNA was extracted using 8M sterile guanidine hydrochloride after residual host DNA was removed with Sigma benzonase nuclease (Merck KGaA, Germany). For genome sequencing, 15 μ g of DNA were used.

DNA Sequencing and Sequence Assembly

Genomic DNA was fragmented using G-tubes (Covaris) and then end-repaired to prepare SMRTbell DNA template libraries (fragment sizes > 10 kb selected using a bluepippin system) according to the manufacturer's instructions (Pacific Biosciences). Library quality was assessed by Qubit 4 Fluorometer (Life Technologies, Singapore), and average fragment size was estimated using a Agilent 2100 Bioanalyzer (Agilent, United States). SMRT sequencing was carried out with a Pacific Biosciences Sequel II sequencer (Pacbio, United States) at Frasergen Bioinformatics Co., Ltd, Wuhan, China.

The PacBio reads were *de novo* assembled using the Canu (v2.0) software (Koren et al., 2017). The depth of genome coverage were determined by palign tool (BLASR, v0.4.1) (Chaisson and Tesler, 2012). The sequencing data/the raw sequencing reads in this article have been deposited in the CNSA database¹ (project ID: CNP0002484). For Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis (Simão et al., 2015), version 5.2.2 was used. The lineage dataset was alphaproteobacteria_odb10 (Creation date: 2020-03-06, number of genomes: 744, number of BUSCOs: 432).

Genome Annotation

The genome was annotated using Glimmer (v3.02) (Delcher et al., 2007). The tRNA and rRNA genes were identified by tRNAscan-SE (v2.0) (Lowe and Eddy, 1996) and RNAmmer (v1.2) (Lagesen et al., 2007), respectively. The functional characterization of the identified genes was achieved by BLASTP with an E-value threshold of 1e-5 against the databases of the NCBI Non-Redundant protein database (NR), Swiss-Prot, Clusters of Orthologous Groups (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Ontology (GO)

¹<https://db.cngb.org/>

(Buchfink et al., 2015). The highest-scoring match was chosen as the final annotation for each gene. Tandem repeats finder and MISA-web were used to predict tandem repeats and simple sequence repeats (SSR) sequences in the genome, respectively (Benson, 1999; Beier et al., 2017). The predicted protein sequences of each genome were compared with each other, and protein alignment was performed using blastp (Altschul et al., 1990). OrthoMCL software was used to cluster and count the clustering results (Chen, 2006). The blast Ring Image Generator (BRIG) (Alikhan et al., 2011) was used to visualize the multiple copy genes incorporation in the genome of the *E. vermicola* symbiont being the reference, in comparison with the closest *Phyllobacterium* genomes. Signal peptide prediction tool signalp (v4.1) was used to annotate whether the protein sequence was a secretory protein (Parameters: - t gram + - f summary) (Nielsen, 2017).

Phylogenetic Analyses

Eight hundred and seventy-nine single copy genes for multilocus analyses were identified from 29 bacteria for the multilocus phylogenetic reconstruction. We used BUSCO v2.0 software to assess the integrity of 29 genomes (Simão et al., 2015). Multiple sequence alignment of single copy genes was accomplished by MUSCLE (v3.8.31) (default parameter) (Edgar, 2004). The maximum likelihood tree was constructed with RAxML (v8.2.12) (Stamatakis, 2014) using the automated protein model assignment algorithm and a gamma model of rate heterogeneity (-m PROTGAMMAAUTO) (Carroll et al., 2019; Sawada et al., 2021), based on concatenated 879 amino acid sequences from single copy genes of endobacteria and close relatives within the genus *Phyllobacterium* (Supplementary Data Sheet 1). The genes concatenated were shown in supporting information (genome_SingleCopy.phylip). 100 bootstrap replicates was used to produce the phylogenetic tree using amino acids sequence. Bootstrap values were given at each branch.

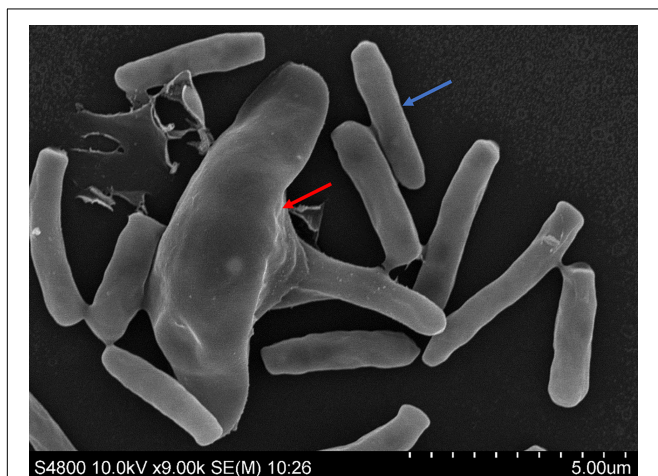


FIGURE 1 | Lunate spores (red arrow) and rod shaped spores (blue arrow) of *Esteya vermicola* CBS 115803 were used to produce mycelium for obtaining endobacteria.

To infer the evolutionary history of subtilases/peptidase S8, maximum likelihood method and a JTT matrix-based model was used. The JTT model was based on the assumption that the amino acid frequencies are in equilibrium and remain the same throughout the evolutionary process (Jones et al., 1992). The subtilases/peptidase S8 genes used were shown in supporting information (Supplementary Data Sheet 2). The evolutionary history of the taxa studied is represented by a bootstrap consensus tree generated from 1000 replicates (Felsenstein, 1985). The tree with the highest log likelihood (-26864.23) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances predicted using the JTT model, and then selecting the topology with the highest log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+ G, parameter = 4.1038)]. The rate variation model allowed for some sites to be evolutionarily invariable [(+ I), 0.00% sites]. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018).

RESULTS

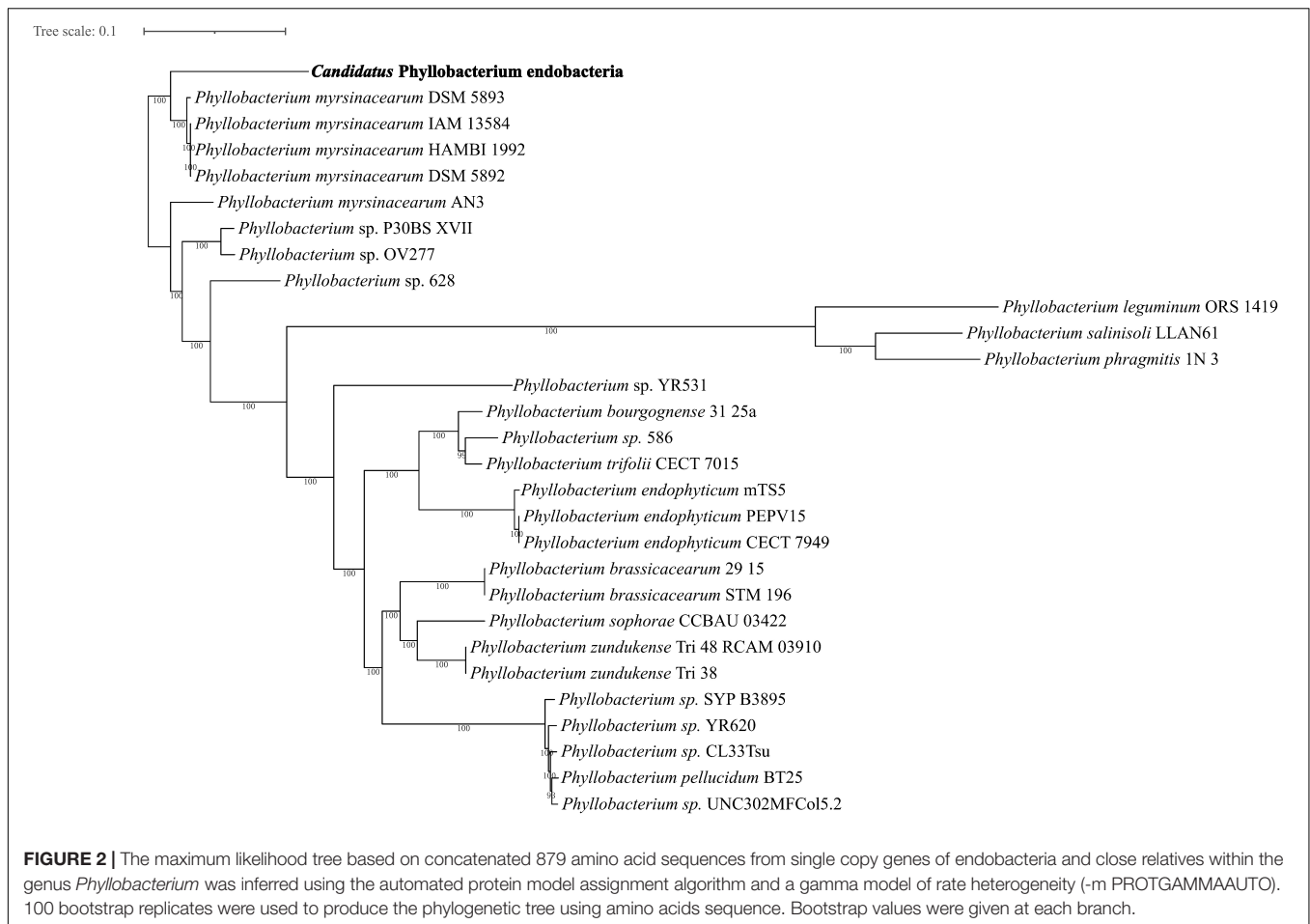
Genome Sequencing, and General Features of the Genome

The lunate spores and rod shaped spores were used to culture mycelium of *Esteya vermicola* CBS115803. Then the mycelium were used to purify intracellular symbiotic bacteria. The spores were observed by scanning electron microscopy to determine whether there were epiphytic bacteria or contaminated bacteria on the surface of the fungus. As can be seen from the image (Figure 1), there were no cell-surface bacteria of *E. vermicola*.

To increase the DNA proportion of endobacteria, nucleases (sigma, Benzonase E1014) were used to degrade host DNA. The third generation pacbio sequencing method was used to improve the read length and assembly accuracy of the reads. The proportion of symbiotic bacteria DNA data increased from 0.1 to 3%. Before filtering the host sequence, the entire sequencing amount was 22 Gb, with 0.664 Gb of bacterial data remaining (3% proportion relative host DNA content). The length of contigs ranged from 1 to 86 Kb. The average read depth was 129.12 \times , and the estimated read depth in genome ranged from 4.87 to 9793.60 \times . The coverage was from 99.74 to 100%. The endobacteria genome had a 58.93% guanine-cytosine (GC) content, according to the metagenomic assemblies. Furthermore, the genome size was about 3.96 Mb based on the number of contigs, and the number of annotated genes was 4542 (Supplementary Tables 1, 2), accounting for 74.18% of the genome. Total 86.79% genes were annotated (Supplementary Table 3).

Evolutionary Placement

The majority of the endobacteria sequences were identified as the genus *Phyllobacterium* according to nucleotide and protein database. A total of 879 single-copy genes found for 29 genomes including 28 genomes of *Phyllobacterium*



and the endobacteria in *E. vermicola*. To determine the phylogenetic placement of the endobacteria, a maximum likelihood phylogenetic analysis of amino acid sequences from 879 single copy genes were conducted. Genome features, background information and BUSCO assessment are shown in **Supplementary Tables 4, 5**. The endobacteria was firmly placed in the branch of *P. myrsinacearum* (**Figure 2**). As a result, the symbiotic bacterium was most closely related to *P. myrsinacearum* species. The symbiotic bacteria clustered with *P. myrsinacearum* species and unknown *Phyllobacterium* species, according to the phylogenetic tree constructed by using the conserved atpD and recA (**Supplementary Figure 1**), and again showed that the endobacteria and *P. myrsinacearum* were the closest relationship.

Gene Family Analysis

The genomes of some bacterial pathogens have gained genes through gene duplication, resulting in increased numbers of key gene clusters or the expansion of important protein families (Cole et al., 1998; Pallen and Wren, 2007). Gene duplication and protein family expansion are important genomic mechanisms that shape the evolution of pathogenic fungi (Powell et al., 2008). Similarly, gene duplication in bacteria facilitates adaptation to changing environmental conditions, which affects the fitness and

immunogenic properties of many bacterial pathogens (Craven and Neidle, 2007; Ahmed et al., 2008). To investigate the features presented in the endobacteria, a comparative gene family analysis against 28 completely sequenced *Phyllobacterium* genomes was performed. The 29 genomes yielded a total of 11,332 gene families. Focusing on the endobacteria genome in *E. vermicola*, 4542 genes were categorized into 2637 families. Interestingly, among the 29 *Phyllobacterium* genomes, multiple copy gene families and multiple copy genes in the endobacteria were the most abundant (**Figure 3A** and **Supplementary Table 6**). Multiple copy genes in the genome of the endobacteria living in *E. vermicola*, in comparison with its the closest *P. myrsinacearum* DSM 5893, indicated that the endobacteria had more multiple copy gene family (42) and multiple copy genes (91) (**Figure 3B** and **Supplementary Table 6**). These genes are mainly involved in amino acid metabolism, carbohydrate metabolism, membrane transport, nucleotide metabolism, replication and repair. Nucleotide excision repair and cell cycle were the specific metabolic pathways of multiple copy gene family in the symbiotic bacteria, which were distinct from those of other free-living bacteria (**Supplementary Tables 6, 7**), which metabolic pathways could be related to the ecological adaptability of the symbiotic bacteria in fungal cells. Eleven gene families (25 genes) of the endobacteria were unique (**Figure 3C** and

Supplementary Table 8), mainly involved in the synthesis of chemotaxis protein and mobility protein.

High Proportion of Repetitive Sequences

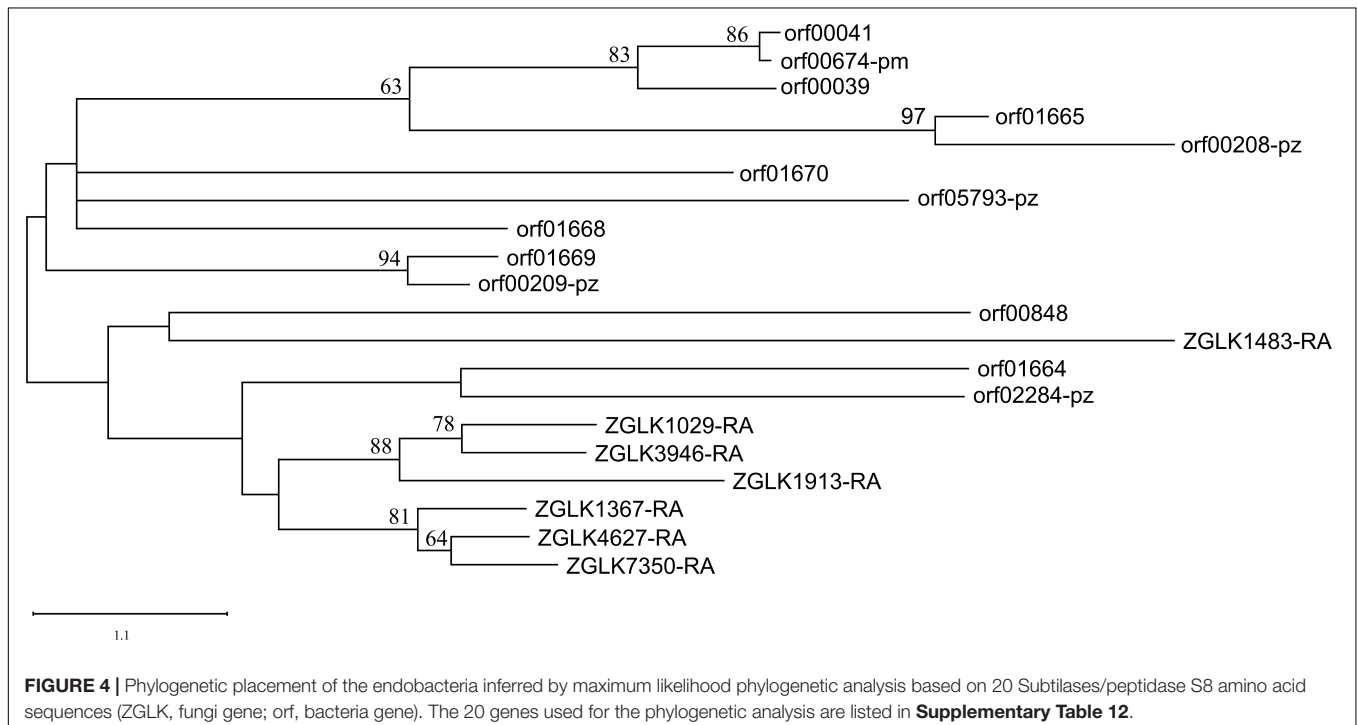
Tandem repeat may act as an engine in genetic variability and bacterial adaptation, as pointed out by accumulating evidence (Zhou et al., 2014). Repetitive sequences may improve the ability to regulate gene expression, also important for bacterial adaptation and increasing their phenotypic plasticity (Zhou et al., 2014). Repeat sequence analysis was performed in order to better understand the genomic structure. It was showed that the endobacteria genome existed 235 short tandem repeats (STRs), which total length was 61839 bp, accounting for 1.56% of the genome. There were 98 genes involved in the 235 STRs. Intragenic STRs were found more frequently near the termini (45% at 5' end, and 55% at 3' end) (**Supplementary Tables 9, 10**). Within the annotated genes, there are 35 STRs involving 29 genes. Two-component system [Biofilm formation (ko02020| ko02025), beta-Lactam resistance [Cationic antimicrobial peptide (CAMP) resistance (ko01501| ko01503), Bacterial secretion system (ko03070), base/nucleotide excision repair (ko03410/ko03420), Quorum sensing [Protein export| Bacterial secretion system (ko02024| ko03060| ko03070), RNA degradation| Longevity regulating pathway (ko03018| ko04212), arginine and proline metabolism (ko00330) were among the involved metabolic pathways. There are multiple genes including more than two STRs, of which the functions involved are chemosensory pili system protein (ChpA), type VI secretion system secreted protein (VgrG), and membrane fusion protein (acrA, mexA, adeI, smeD, mtrC, cmeA). *P. zundukense* (Accession numbers: GCF_002764115.1), on the other hand,

had only 98 STRs in its genome, involving in 72 genes. The total length was 8,395 bp, accounting for only 0.14% of the genome. Half number of STRs was intragenic, which were found more frequently near the termini (50% at 5' end, and 50% at 3' end) (**Supplementary Tables 9, 10**). Also, *P. myrsinacearum* (Accession numbers: GCF_004216655.1) had only 92 STRs (13,637 bp), accounting for only 0.24% of the genome, involving in 69 genes. 64% out of the 92 STRs was intragenic, which were also found more frequently near the termini (48% at 5' end, and 52% at 3' end).

Specially, both SSRs (CGC)5 and STR (5'-GGGGAGAG GGGGAGAGGGGGGAGA-3') was discovered in 3' end of the *luxR* family gene (bacterial regulatory) of the endobacteria, which regulates motility, synthesis of virulence determinants and biofilm architecture (Chen and Xie, 2011). The gene is highly regulated and could play a key role in a flexible response to the changing environment conditions.

Environmental Adaptation Pathway

Intriguingly, there were 8 Acyl-CoA synthetase genes in oxidative phosphorylation of metabolic pathway (ko04714) of the endobacteria. Only two Acyl-CoA synthetase genes were found in *P. zundukense*, and one gene was found in *P. myrsinacearum*, both of which were closely related to the endobacteria genome. The flow of electrons through the electron transport chain is an exergonic process. The energy from the redox reactions creates an electrochemical proton gradient, which drives the synthesis of adenosine triphosphate (ATP). The endobacteria genome had ten genes encoding electron transport chains, including *ctaA* (1), *ctaG* (2), *fbhH* (1), *petA* (2), *petB* (3), *petC* (1). While there were only six and five those genes in *P. zundukense*



and *P. myrsinacearum*, respectively. The endobacteria possessed QseC gene, which is involved in metabolic pathway of two-component system. QseC is a bacterial adrenergic receptor that activates virulence genes in response to interkingdom cross-signaling (Clarke et al., 2006). Furthermore, the endobacteria had great potential to use trehalose, with three neutral trehalase and four trehalose consumption proteins in its genome. In contrast, there was only one such gene in the genomes of the two closely related bacteria, *P. zundukense* and *P. myrsinacearum*. The role of trehalose in symbiosis is to enhance abiotic stress tolerance (Sharma et al., 2020).

Endobacteria Genome in Nematophagous Fungi *Esteya vermicola* Contain Genes Encoding Nematicidal Proteins

Subtilases/peptidase S8 and subtilisin-like were prevalent in nematophagous fungi and played a role in nematode lethality and infection process (Wang et al., 2018). Surprisingly, there were eight subtilase/subtilisin genes in the genome of endosymbiotic bacteria, five of which could be transported outside the endobacteria cell (Supplementary Table 11). The endobacteria had more subtilase/subtilisin genes than closely related *P. zundukense* and *P. myrsinacearum*, which had four and one genes, respectively. We used subtilases/peptidase S8 genes in the endobacteria (eight genes) and its fungal host *E. vermicola* (seven genes), *P. zundukense* (four genes) and *P. myrsinacearum* (one gene) (Supplementary Table 12) to construct a phylogenetic tree to explore their genetic and evolutionary relationships. Almost all the subtilases/peptidase S8 genes in the fungi *E. vermicola* were found to be clustered together, with the exception of one gene, ZGLK1483-RA, which encodes a secretory protein. Almost all of the subtilases/peptidase S8 genes from bacteria were clustered together (Figure 4). According to the phylogenetic tree, the bacteria's subtilases/peptidase S8 genes originated earlier. Four subtilases/peptidase S8 which are typically virulence factors, were found to be autotransporter associated with infection or virulence in pathogenic bacteria (Henderson and Nataro, 2001; Nishimura et al., 2010). Endobacteria genome in nematophagous fungi *E. vermicola* contain multiple nematicidal genes, which is most likely the outcome of coevolution, adapting to its niche and incorporates with the fungal host to kill nematodes.

DISCUSSION

The First Identification of *Candidatus Phyllobacterium* Endobacteria in the Cells of the Nematophagous Fungi *Esteya vermicola*

According to blast against nucleotide and protein database, 282 of 306 contigs of the genome can match with known sequences. The genus *Phyllobacterium* was the closest blast matches (121 of the 282 sequences), and *Acidovorax* was

the second closest blast matches (17 of the 282 sequences) (Supplementary Tables 13, 14). Therefore, the endosymbiotic bacteria were preliminarily identified as this genus *Candidatus Phyllobacterium*, so the genomes of this genus *Phyllobacterium* was used to construct phylogenetic tree. However, the 16S gene was not obtained by genome sequencing and assembly. Moreover, this preliminary genome-based identification result was different from the previous 16S identification result (Wang et al., 2017). The following are some plausible explanations: 1. More than one kind of endobacteria exists in fungal cells. 2. The endobacteria *Candidatus Phyllobacterium* may be enriched in this medium. The abundance and the thickness of cell wall of the endobacterium may alter when the culture medium changed.

Phyllobacterium have been found as free-living bacteria in soil, in water and associated with unicellular organisms, and *Phyllobacterium* strains were mainly isolated from plant rhizosphere and root nodules. Several strains are capable of communicating with plant tissues and promoting plant growth (Mantelin et al., 2006; Flores-Félix et al., 2013; Sánchez et al., 2014). For the first time, we discovered this endobacteria in the cells of the nematophagous fungi *E. vermicola*.

Genome Assembly Integrity

That 16S rRNA gene and other conserved (*ftsZ*, *atpD* and *recA*) gene absent or not was researched in endobacteria of 8 fungal endobacteria and 1 insect endobacterium (Table 1). Only 2 fungal and 1 insect endobacteria genome had one single copy 16S rRNA gene. Cell division protein (*ftsZ*) were absent in four MRE genome, consistent with Kuo's commentary (Kuo, 2015). *AtpD* gene, which encodes ATP synthase subunit beta, was not included in the assembled genome of the 7 endobacteria of fungi, except that of endobacteria of *E. vermicola* and *Wolbachia* endosymbiont. It is inferred that ATP may be provided by their host fungi. In contrast, *recA* gene, essential for the repair and maintenance of DNA, exists in all seven symbiotic bacteria and *Wolbachia* endosymbiont, except MRE living in the AMF *D. heterogama*, and this may be due to assembly problems, resulting in the deletion of the gene. These findings suggest that the cell functioning and proliferation of endobacteria may be under the control of their fungal hosts, that is likely the outcome of long-term coevolution (Kuo, 2015).

The genome assembly integrity of the endobacteria was 43% (BUSCO, bacteria_odb10), which was higher than that of the other four MRE genomes (Supplementary Table 15). Nevertheless, all five genomes were assembled in fragments. Genomes of *B. rhizoxinica*, *Ca. G. gigasporarum*, endobacteria of *Mortierella elongata* FMR23-6 had a better assembly results (BUSCO > 85%). Because *B. rhizoxinica* can be cultured artificially (Partida-Martinez and Hertweck, 2005). While *Ca. G. gigasporarum* and endobacteria of *Mortierella elongata* FMR23-6 were closely related, both belonging to the family Burkholderiaceae (Sharmin et al., 2018), and their cell size was larger than that of other endobacteria. Further, the genome of insect *Wolbachia* endosymbiont assembled the completion map (PRJNA544530). Only endobacteria *Ca. G. gigasporarum* has all four rod shape determining gene *ftsZ*, *ispA*, *mreB*, *rodA*, which

TABLE 1 | Presence or absence of conserved genes in endobacteria of fungi (8) and insect (1).

Assembly Accession	Species	16S rRNA	ftsZ	atpD	recA	ispA	mreB	rodA
project ID: CNP0002484	<i>Candidatus</i> Phyllobacterium (alphaproteobacteria) living in <i>E. vermicola</i> .	none	2	1	2	1	none	none
GCA_002888635.1	MRE living in the AMF <i>Dentiscutata heterogama</i> (mycoplasmatales)	none	none	none	none	none	none	none
GCA_001017705.1	Mycoplasmataceae bacterium RC_NB112A (mycoplasmatales)	none	none	none	1	none	none	none
GCA_001017675.1	Mycoplasmataceae bacterium CE_OT135 (mycoplasmatales)	none	none	none	1	none	none	none
GCA_001017685.1	Mycoplasmataceae bacterium RV_VA103A (mycoplasmatales)	none	none	none	1	none	none	none
GCA_000775775.1	Bacterium endosymbiont of <i>Mortierella elongata</i> FMR23-6 (betaproteobacterial)	1	1	none	1	none	1	1
GCA_000227585.1	<i>Candidatus</i> Glomeribacter gigasporarum (Betaproteobacteria)	1	1	none	1	1	1	1
GCA_000198775.1	<i>Mycetohabitans rhizoxinica</i> living in <i>Rhizopus microsporus</i> (Burkholderiaceae)	none	1	none	1	none	2	1
GCF_013458815.1	<i>Wolbachia</i> endosymbiont of <i>Diaphorina citri</i> (a-proteobacteria; alphaproteobacteria)	1	1	1	1	none	none	1

ftsZ, ispA, mreB, rodA: rod shape determining gene.

is consistent with its rod-shaped morphological characteristics (Jargeat et al., 2004).

No α - and β -Pinene Monoterpenoid Synthase-Encoding Genes in This Version Genome

By phylogenetic analysis of terpene synthase from bacteria and fungi, it imply horizontal gene transfer from bacteria to fungi, of which the majority are *Metarhizium* species (Jia, 2016). Our previous results showed that the presence of pine wood nematode but not contact with *E. vermicola* would induce the fungus to produce more lunate spores that prey on nematode (Wang et al., unpublished). Meanwhile, *E. vermicola* can stimulate activity of pine wood nematode under untouched conditions (Wang et al., unpublished). So *E. vermicola* and pine wood nematode communicate with each other by volatile compounds. *Esteya vermicola* attracts pine wood nematode through volatile monoterpene compounds, α -pinene, β -pinene, and camphor (Lin et al., 2013). However, α - and β -pinene monoterpene synthase-encoding genes were absent in *E. vermicola* genome (almost completely assembled, 42 scaffolds without gap), so it was inferred that endobacteria produced the monoterpene that attracted pine wood nematode (Wang et al., 2018). Simultaneously, no monoterpene synthase gene was identified in the assembled version of the endobacterium genome, which was highly possible due to the low integrity of the assembled genome and the annotation. Assuming that the endobacterium produces volatile compounds that attract pine wood nematodes, the endobacteria probably mobilize host fungi to kill nematodes through volatile compounds, although different from free-living *Stenotrophomonas maltophilia*, which used urea metabolite ammonia to function as a signal molecule in *Arthrobotrys oligospora* to initiate the lifestyle switch to form trap structures against nematodes (Wang et al., 2014).

Possible Significance of Highly Repetitive Sequences

In comparison to closed related species, the endobacteria had more repeats sequence. Bacteria can use tandem repeats to adapt to changing environments by reversibly shutting down or modulating the function of specific genes. Tandem repeats

variations have been shown to have a role of in bacterial adaptation strategies, such as immune evasion, tissue tropism and the modulation of environmental stress tolerance, and tandem repeats and their variability provide regulatory strategy for adaptability (Zhou et al., 2014). In general, variable tandem repeats tend to localize in flexible genes involved in pathogens invasion (Moxon et al., 2006; Guo and Mrázek, 2008). Based on the above, we inferred that the repetitive sequences of endosymbiotic bacteria may be used to regulate genes about pathogens invasion.

The Endobacteria Possibly Cooperate With Host to Kill Nematodes

Subtilase can degrade nematode and insect body walls (Lu et al., 2020). According to the phylogenetic tree, the diversity of subtilase was higher in bacteria than it was in fungi. The secretory subtilase in symbiotic bacteria and in their host fungi may work together to infiltrate the cuticle of pine wood nematode, enhancing the efficiency of killing the nematodes. A better understanding of its endobacteria may aid in finding solutions for use of the nematophagous fungi. We point out the probable synergistic function of endobacteria and its fungal host, in order to guide future research toward a deeper analysis of the importance of these endobacteria for the fungal life cycle and killing pine wood nematode.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The sequencing data/the raw sequencing reads in this article have been deposited in the CNSA database, [https://db.cngb.org/\(project ID: CNP0002484\)](https://db.cngb.org/(project ID: CNP0002484)).

AUTHOR CONTRIBUTIONS

RW designed research and wrote the manuscript. LD analyzed data and modified the manuscript. YC and SW performed research. LQ designed research and modified the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Natural Science Foundation of China (31800547) and Fundamental Research Funds of Chinese Academy of Forestry (CAFYBB2021ZG001).

REFERENCES

- Ahmed, N., Dobrindt, U., Hacker, J., and Hasnain, S. E. (2008). Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. *Nat. Rev. Microbiol.* 6, 387–394. doi: 10.1038/nrmicro1889
- Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Archibald, J. (2014). *One Plus One Equals One: Symbiosis and the Evolution of Complex Life*. Oxford: Oxford University Press.
- Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Carroll, L. M., Gaballa, A., Guldimann, C., Sullivan, G., Henderson, L. O., and Wiedmann, M. (2019). Identification of novel mobilized colistin resistance gene *mcr-9* in a multidrug-resistant, colistin-susceptible *salmonella enterica* serotype typhimurium isolate. *MBio* 10, e00853–19. doi: 10.1128/mBio.00853-19
- Chaisson, M. J., and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 13:238. doi: 10.1186/1471-2105-13-238
- Chen, F. (2006). OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.* 34, D363–D368. doi: 10.1093/nar/gkj123
- Chen, J., and Xie, J. (2011). Role and regulation of bacterial LuxR-like regulators. *J. Cell. Biochem.* 112, 2694–2702. doi: 10.1002/jcb.23219
- Clarke, M. B., Hughes, D. T., Zhu, C., Boedeker, E. C., and Sperandio, V. (2006). The QseC sensor kinase: a bacterial adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10420–10425. doi: 10.1073/pnas.0604343103
- Cole, St, Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544. doi: 10.1038/31159
- Craven, S. H., and Neidle, E. L. (2007). Double trouble: medical implications of genetic duplication and amplification in bacteria. *Future Microbiol.* 2, 309–321. doi: 10.2217/17460913.2.3.309
- Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23, 673–679. doi: 10.1093/bioinformatics/btm009
- Douglas, A. E. (2014). Symbiosis as a general principle in eukaryotic evolution. *Cold Spring Harb. Perspect. Biol.* 6:a016113. doi: 10.1101/cshperspect.a016113
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678
- Ferrari, J., and Vavre, F. (2011). Bacterial symbionts in insects or the story of communities affecting communities. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1389–1400. doi: 10.1098/rstb.2010.0226
- Flores-Félix, J.-D., Carro, L., Velázquez, E., Valverde, Á., Cerda-Castillo, E., García-Fraile, P., et al. (2013). *Phyllobacterium endophyticum* sp. nov., isolated from nodules of *Phaseolus vulgaris*. *Int. J. Syst. Evol. Microbiol.* 63, 821–826. doi: 10.1099/ijs.0.038497-0
- Fujimura, R., Nishimura, A., Ohshima, S., Sato, Y., Nishizawa, T., Oshima, K., et al. (2014). Draft genome sequence of the betaproteobacterial endosymbiont associated with the fungus *Mortierella elongata* FMR23-6. *Genome Announc.* 2, e1272–e1214. doi: 10.1128/genomeA.01272-14

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.842684/full#supplementary-material>

- Futai, K. (2013). Pine Wood Nematode, *Bursaphelenchus xylophilus*. *Annu. Rev. Phytopathol.* 51, 61–83. doi: 10.1146/annurev-phyto-081211-172910
- Ghignone, S., Salvioli, A., Anca, I., Lumini, E., Ortu, G., Petiti, L., et al. (2012). The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. *ISME J.* 6, 136–145. doi: 10.1038/ismej.2011.110
- Guo, X., and Mrázek, J. (2008). Long simple sequence repeats in host-adapted pathogens localize near genes encoding antigens, housekeeping genes, and pseudogenes. *J. Mol. Evol.* 67, 497–509. doi: 10.1007/s00239-008-9166-5
- Henderson, I. R., and Nataro, J. P. (2001). Virulence functions of autotransporter proteins. *Infect. Immun.* 69, 1231–1243. doi: 10.1128/IAI.69.3.1231-1243.2001
- Jargeat, P., Cosseau, C., Oláh, B., Jauneau, A., Bonfante, P., Batut, J., et al. (2004). Isolation, free-living capacities, and genome structure of “candidatus *Glomeribacter gigasporarum*,” the endocellular bacterium of the mycorrhizal fungus *Gigaspora margarita*. *J. Bacteriol.* 186, 6876–6884. doi: 10.1128/JB.186.20.6876-6884.2004
- Jia, Q. (2016). *Computational Identification of Terpene Synthase Genes and Their Evolutionary Analysis* Doctoral Dissertations. Knoxville: University of Tennessee.
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116
- Kumar, S., Stecher, G., Li, M., Nknyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Kuo, C. H. (2015). Scrambled and not-so-tiny genomes of fungal endosymbionts. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7622–7623. doi: 10.1073/pnas.1508360112
- Lackner, G., Moebius, N., Partida-Martinez, L. P., Boland, S., and Hertweck, C. (2011). Evolution of an endofungal lifestyle: deductions from the *Burkholderia rhizoxinica* Genome. *BMC Genomics* 12:210. doi: 10.1186/1471-2164-12-210
- Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160
- Lin, F., Ye, J., Wang, H., Zhang, A., and Zhao, B. (2013). Host deception: predaceous fungus, *Esteya vermicola*, entices pine wood nematode by mimicking the scent of pine tree for nutrient. *PLoS One* 8:e71676. doi: 10.1371/journal.pone.0071676
- Liou, J. Y., Shih, J. Y., and Tzean, S. S. (1999). *Esteya*, a new nematophagous genus from Taiwan, attacking the pinewood nematode (*Bursaphelenchus xylophilus*). *Mycol. Res.* 103, 242–248. doi: 10.1017/S0953756298006984
- Lowe, T. M., and Eddy, S. R. (1996). TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.0955
- Lu, X., Liang, Y., Deng, X., Yin, D., and Song, R. (2020). Comprehensive analysis of subtilase gene family and their responses to nematode in trichoderma harzianum. [Preprint] doi: 10.20944/preprints202010.0403.v1
- Mantelin, S., Saux, M. F.-L., Zakhia, F., Béna, G., Bonneau, S., Jeder, H., et al. (2006). Emended description of the genus *Phyllobacterium* and description of four novel species associated with plant roots: *Phyllobacterium bourgognense* sp. nov., *Phyllobacterium ifriqiense* sp. nov., *Phyllobacterium leguminum* sp. nov. and *Phyllobacterium brassic*. *Int. J. Syst. Evol. Microbiol.* 56, 827–839. doi: 10.1099/ijs.0.63911-0
- Moxon, R., Bayliss, C., and Hood, D. (2006). Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* 40, 307–333. doi: 10.1146/annurev.genet.40.110405.090442

- Naito, M., Morton, J. B., and Pawlowska, T. E. (2015). Minimal genomes of mycoplasma-related endobacteria are plastic and contain host-derived genes for sustained life within Glomeromycota. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7791–7796. doi: 10.1073/pnas.1501676112
- Nielsen, H. (2017). “Predicting secretory proteins with SignalP,” in *Protein Function Prediction*, ed. D. Kihara (Berlin: Springer), 59–73. doi: 10.1007/978-1-4939-7015-5_6
- Nishimura, K., Tajima, N., Yoon, Y.-H., Park, S.-Y., and Tame, J. R. H. (2010). Autotransporter passenger proteins: virulence factors with common structural themes. *J. Mol. Med.* 88, 451–458. doi: 10.1007/s00109-010-0600-y
- Pallen, M. J., and Wren, B. W. (2007). Bacterial pathogenomics. *Nature* 449, 835–842.
- Partida-Martinez, L. P., and Hertweck, C. (2005). Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* 437, 884–888. doi: 10.1038/nature03997
- Powell, A. J., Conant, G. C., Brown, D. E., Carbone, I., and Dean, R. A. (2008). Altered patterns of gene duplication and differential gene gain and loss in fungal pathogens. *BMC Genomics* 9:147. doi: 10.1186/1471-2164-9-147
- Sánchez, M., Ramírez-Bahena, M.-H., Peix, A., Lorite, M. J., Sanjuán, J., Velázquez, E., et al. (2014). *Phyllobacterium loti* sp. nov. isolated from nodules of *Lotus corniculatus*. *Int. J. Syst. Evol. Microbiol.* 64, 781–786. doi: 10.1099/ijms.0.052993-0
- Sawada, H., Fujikawa, T., Tsuji, M., and Satou, M. (2021). *Pseudomonas allii* sp. nov., a pathogen causing soft rot of onion in Japan. *Int. J. Syst. Evol. Microbiol.* 71, doi: 10.1099/ijsem.0.004582
- Sharma, M. P., Grover, M., Chourasiya, D., Bharti, A., Agnihotri, R., Maheshwari, H. S., et al. (2020). Deciphering the role of trehalose in tripartite symbiosis among rhizobia, arbuscular mycorrhizal fungi, and legumes for enhancing abiotic stress tolerance in crop plants. *Front. Microbiol.* 11:509919. doi: 10.3389/fmicb.2020.509919
- Sharmin, D., Guo, Y., Nishizawa, T., Ohshima, S., Sato, Y., Takashima, Y., et al. (2018). Comparative genomic insights into endofungal lifestyles of two bacterial endosymbionts, *Mycovaidus cysteinexigens* and *Burkholderia rhizoxinica*. *Microbes Environ.* 33, 66–76. doi: 10.1264/jmsme2.ME17138
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Torres-Cortés, G., Ghignone, S., Bonfante, P., and Schüßler, A. (2015). Mosaic genome of endobacteria in arbuscular mycorrhizal fungi: transkingdom gene transfer in an ancient mycoplasma-fungus association. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7785–7790. doi: 10.1073/pnas.1501540112
- Vicente, C., Espada, M., Vieira, P., and Mota, M. (2012). Pine Wilt Disease: a threat to European forestry. *Eur. J. Plant Pathol.* 133, 89–99. doi: 10.1007/s10658-011-9924-x
- Wang, C. Y., Fang, Z. M., Wang, Z., Zhang, D. L., Gu, L. J., Lee, M. R., et al. (2011). Biological control of the pinewood nematode *Bursaphelenchus xylophilus* by application of the endoparasitic fungus *Esteya vermicola*. *BioControl* 56, 91–100. doi: 10.1007/s10526-010-9302-1
- Wang, R., Dong, L., Chen, Y., Qu, L., Wang, Q., and Zhang, Y. (2017). *Esteya Vermicola*, a nematophagous fungus attacking the pine wood nematode, harbors a bacterial endosymbiont affiliated with Gammaproteobacteria. *Microbes Environ.* 32, 201–209. doi: 10.1264/jmsme2.ME16167
- Wang, R., Dong, L., He, R., Wang, Q., Chen, Y., Qu, L., et al. (2018). Comparative genomic analyses reveal the features for adaptation to nematodes in fungi. *DNA Res.* 25, 245–256. doi: 10.1093/dnares/dsx053
- Wang, X., Li, G. H., Zou, C. G., Ji, X. L., Liu, T., Zhao, P. J., et al. (2014). Bacteria can mobilize nematode-trapping fungi to kill nematodes. *Nat. Commun.* 5:5776. doi: 10.1038/ncomms6776
- Zhou, K., Aertsen, A., and Michiels, C. W. (2014). The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiol. Rev.* 38, 119–141. doi: 10.1111/1574-6976.12036

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

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Dong, Chen, Wang and Qu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.