Arumugaperumal et al. Annals of Microbiology (2020) 70:3
https://doi.org/10.1186/s13213-020-01549-w

Annals of Microbiology

UNIVERSITÀ DEGLI STUDI DI MILANO

ORIGINAL ARTICLE Open Access

The draft genome of a new *Verminephrobacter eiseniae* strain: a nephridial symbiont of earthworms

Arun Arumugaperumal[1†], Sayan Paul[1†], Saranya Lathakumari[1], Ravindran Balasubramani[2] and Sudhakar Sivasubramaniam[1*]

Abstract

Purpose: *Verminephrobacter* is a genus of symbiotic bacteria that live in the nephridia of earthworms. The bacteria are recruited during the embryonic stage of the worm and transferred from generation to generation in the same manner. The worm provides shelter and food for the bacteria. The bacteria deliver micronutrients to the worm. The present study reports the genome sequence assembly and annotation of a new strain of *Verminephrobacter* called *Verminephrobacter eiseniae* msu.

Methods: We separated the sequences of a new *Verminephrobacter* strain from the whole genome of *Eisenia fetida* using the sequence of *V. eiseniae* EF01-2, and the bacterial genome was assembled using the CLC Workbench. The de novo-assembled genome was annotated and analyzed for the protein domains, functions, and metabolic pathways. Besides, the multigenome comparison was performed to interpret the phylogenomic relationship of the strain with other proteobacteria.

Result: The FastqSifter sifted a total of 593,130 *Verminephrobacter* genomic reads. The de novo assembly of the reads generated 1832 contigs with a total genome size of 4.4 Mb. The Average Nucleotide Identity denoted the bacterium belongs to the species *V. eiseniae*, and the 16S rRNA analysis confirmed it as a new strain of *V. eiseniae*. The AUGUSTUS genome annotation predicted a total of 3809 protein-coding genes; of them, 3805 genes were identified from the homology search.

Conclusion: The bioinformatics analysis confirmed the bacterium is an isolate of *V. eiseniae*, and it was named *Verminephrobacter eiseniae* msu. The whole genome of the bacteria can be utilized as a useful resource to explore the area of symbiosis further.

Keywords: Symbiosis, Bacterial genome, Earthworm, *Verminephrobacter*

Introduction

The symbiosis has been recognized as a central driver for evolutionary innovation (Raina et al. 2018). The symbiotic relationship defines the interaction between the symbiont and host in an intimate association which can be mutualistic, commensalistic, or parasitic (Dimijian 2000). In mutualistic symbiosis, both the interacting partners get the benefit from each other. In commensalism, the symbiont gets benefited in terms of fitness without affecting or harming the host. In parasitism, the symbiotic association harms the host, but the symbiont enjoys the benefits (Fukui 2014). In the recent past, the event of symbiosis in the annelids has been explored to a certain extent. In marine annelids, the chemosynthetic symbionts act as a potential supplier of energy and carbon (Dubilier et al. 2008). In medicinal leech, the symbionts play an essential role in supplying the necessary nutrients and vitamins usually lacking in the blood meal (Lund et al. 2010). In lumbricid earthworms, the symbionts reside in the excretory organ and benefit the species through internal recycling of the

* Correspondence: sudhakar@msuniv.ac.in
†Arun Arumugaperumal and Sayan Paul contributed equally to this work.
[1]Department of Biotechnology, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu 627012, India
Full list of author information is available at the end of the article

nitrogen. Among these symbionts, the well-delineated organism is *Verminephrobacter*.

The *Verminephrobacter* genus comprises bacteria that inhabit the nephridia of earthworms. The bacteria are given shelter and food by earthworms. The bacterium, in turn, provides a reproductive advantage to the earthworm (Lund et al. 2010; Viana et al. 2018). Apart from this bacterial genus, there are also other bacteria that harbor the earthworms' nephridia (Davidson et al. 2013). The nephridia are present in pairs in each segment of earthworms and they interconnect adjacent septa. The coelomic fluid is taken in through the nephrostome in one segment and is circulated through three loops and finally emptied outside through nephridiopore present in the adjacent segment. The symbiotic bacteria are present in the region of the second loop known as the ampulla (Schramm et al. 2003). The bacteria are deposited in the cocoon along with the eggs and sperms. It lives on the albumin present in the cocoon and is selectively recruited through a canal that forms in the embryos (Davidson and Stahl 2008). Using the type IV pili, bacteria reach the bladder and then the bacteria use flagella to reach further to the ampulla (Dulla et al. 2012). The bacteria help for earlier sexual maturity and increase cocoon hatching success rate during nutrition depletion (Lund et al. 2010; Viana et al. 2018) and might also provide vitamin B_2 which is a source of the cofactor of FMN and FAD and pyrroloquinoline quinone which helps for better mitochondria function (Pinel 2009). In total, 191 16S rRNA genes of different *Verminephrobacter* clones are reported in the NCBI nucleotide database, and the whole genome of *V. aporrectodeae* At4 (T) and *V. eiseniae* EF01-2 was sequenced already (Kjeldsen et al. 2012; Pinel et al. 2008).

Recently, we have performed the whole-genome annotation of earthworm *Eisenia fetida*, and among the annotated 29,552 protein-coding genes, 6121 genes were obtained from the bacteria (Paul et al. 2018). This indicates the symbiotic relationship and event of vertical gene transfer between the symbiont and host (Davidson et al. 2014; Paz et al. 2017). Since 61% of these bacterial genes were retrieved from the vastly studied earthworm symbiont *Verminephrobacter eiseniae*, it captivated us to focus on the genetic material of bacterium to identify whether it is the different species of *Verminephrobacter* genus or a new strain of the *V. eiseniae* species. Significantly, the 16S rRNA confirmed that our studied bacterium was a new strain of the *Verminephrobacter eiseniae* species. The present paper deals with the whole-genome sequencing, genome feature annotation, and multigenome comparison of the newly identified strain *Verminephrobacter eiseniae* msu. The genome resource and genome features of the strain unveiled through our study will be helpful to the earthworm research community to analyze the symbiotic relation of the species in depth.

Materials and methods

Retrieval, identification, and separation of the *Verminephrobacter* genome sequence reads from earthworm genome

The raw sequence reads of *Eisenia fetida* whole genome were downloaded from NCBI using the GenBank accession CYRZ000000000 (BioProject: PRJEB10048; BioSample: SAMEA3495318) (Zwarycz et al. 2015). All the forward and reverse read files were concatenated using Linux command line programming. The *Verminephrobacter eiseniae* EF01-2 chromosome and plasmid sequences were downloaded from NCBI using the GenBank accessions CP000542 and CP000543, respectively. The genome sequence reads of earthworm *E. fetida* were aligned to the *Verminephrobacter eiseniae* EF01-2 genome using the Burrows-Wheeler Aligner (BWA) algorithm (Li and Durbin 2009). The aligned reads were sifted using the FastqSifter tool (https://github.com/josephryan/FastqSifter).

Quality control, de novo assembly, and genome completeness evaluation

The FastQC quality control tool version 0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (Andrews 2016) and CLC Genomics Workbench version 11.0.1 (Rathy et al. 2018) were used to analyze the quality of the sifted raw reads and trim the ambiguous low-quality reads. After quality assessment, the filtered reads were subjected to de novo assembly by using the CLC Genomics Workbench. The quality of the assembly and completeness of the genome were assessed by using the gVolante web server with Benchmarking Universal Single-Copy Orthologs (BUSCO) v1 ortholog search pipeline (Nishimura et al. 2017; Waterhouse et al. 2017). The BUSCO v1 analyzed the completeness of the genome based on the single-copy orthologs obtained from OrthoDB v9 (Zdobnov et al. 2016). The identified single-copy orthologs (BUSCOs) were further categorized as complete and single-copy BUSCOs (S), complete and duplicated BUSCOs (D), fragmented BUSCOs (F), and missing BUSCOs (M). The SNPs, InDels, and other structural variations present between the genomes of our *Verminephrobacter* bacteria and previously reported *Verminephrobacter eiseniae* EF01-2 were detected by using the Basic Variant Detection tool and the InDels and Structural Variants tool of the CLC Genomics Workbench.

16S rRNA analysis and identification of the bacteria

The 16S rRNA gene sequence within the bacterial genome was predicted by using the RNAmmer web server

version 1.2 (Lagesen et al. 2007). Subsequently, the predicted 16S rRNA sequence was blasted against the curated 16S ribosomal RNA sequence database residing in NCBI using the BLASTn algorithm and default parameters to identify its closest phylogenetic neighbors based on sequence similarity. The top 50 closest neighboring strains' 16S rRNA sequences were extracted from the BLAST search. The deduced 16S rRNA sequence of *Verminephrobacter eiseniae* msu was aligned to its nearby neighbor strains using the ClustalW multiple sequence alignment method (Thompson et al. 1994) with the following parameters: gap opening penalty, 15; gap extension penalty, 6.66; DNA weight matrix, IUB; and transition weight, 0.5. The phylogeny reconstruction was performed through a maximum likelihood method (Steel and Penny 2000) with bootstrap replicate value 100 and the Kimura 2-parameter substitution model using the MEGA 7 software (Kumar et al. 2016). The best fit substitution model was detected by using the Find Best DNA/Protein Models (ML) tool of MEGA. The option tests the alignment file for the goodness of fit to the popular evolution prediction models using the parameters like frequencies, transition probabilities, and rate variation. The model with the lowest Bayesian information criterion (BIC) score was considered as the best model to describe the substitution pattern (Hall 2013). The taxonomic affiliation of the new *Verminephrobacter* genome was confirmed through the Average Nucleotide Identity (ANI) with the genomes of its closely related taxa using the Orthologous Average Nucleotide Identity Tool (OAT) (https://www.ezbiocloud.net/tools/orthoani) with the species demarcation cutoff value at 95% (Lee et al. 2016).

Genome annotation and visualization of bacterial genome map

The initial genome annotation was performed by using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al. 2016) while submitting the genome sequence to GenBank. Besides, the genome was annotated by using the Rapid Annotation using Subsystem Technology (RAST) version 2.0 (http://rast.nmpdr.org/) (Aziz et al. 2008) and AUGUSTUS ab initio gene prediction server (Hoff and Stanke 2013) using bacteria as reference species. The AUGUSTUS-predicted protein-coding genes were used for functional annotation and pathway analysis. The non-coding RNA present in the genome was predicted using the cmscan option of the Infernal software. The mobile genetic elements such as DNA transposons and retrotransposons were predicted by scanning the *Verminephrobacter eiseniae* msu genome using the tool TEclass (http://www.bioinformatics.uni-muenster.de/tools/teclass/generate/index.pl?lang=en) (Abrusán et al. 2009). Simultaneously, the insertion sequence (IS) families associated with the *Verminephrobacter eiseniae* msu genome were predicted by using the ISfinder tool (https://www-is.biotoul.fr/) using the ISsaga pipeline (Varani et al. 2011). The graphical circular maps of the genome describing the sequence feature, base composition, and sequence similarity plots were created by using the CGView server (Grant and Stothard 2008).

Identification, functional annotation, and pathway analysis of *V. eiseniae* msu protein-coding genes

The AUGUSTUS-predicted *Verminephrobacter eiseniae* msu protein-coding genes were identified by BLAST search against NCBI nr (non-redundant) database using the BLASTx algorithm with *E* value threshold 1E−05. The Gene Ontology (GO) annotation describing the biological processes, molecular functions, and cellular components associated with the *V. eiseniae* msu protein-coding genes was performed by using the BLAST2GO functional annotation software version 5.0 (Ashburner et al. 2000; Conesa et al. 2005). The GO terms and the enzyme commission number (EC number) were assigned based on the parameters like annotation cutoff, 55; GO weight, 5; *E* value hit filter, 1E−6; HSP hit coverage cutoff, nil; and hit filter, 500. The conserved domains, motifs, and functional sites associated with the *V. eiseniae* msu genome were identified by annotating the protein-coding genes against the InterPro database using the InterProScan plug-in of BLAST2GO (Mulder and Apweiler 2008). The orthologous groups related to the *V. eiseniae* msu genes were predicted and classified by using the EggNog tool (evolutionary genealogy of genes) (http://eggnogdb.embl.de/) with the parameters *E* value, 1E−3; filter by similarity, 50%; and Hsp/Hit coverage filter, 0 (Huerta-Cepas et al. 2015). The cellular and metabolic pathways related to the bacterial genome were predicted by annotating the protein-coding genes against the Kyoto Encyclopedia of Genes and Genomes (KEGG) online database using KEGG Automatic Annotation Server (KAAS) web annotation server (Moriya et al. 2007). The KEGG pathways were predicted by assigning the *K* numbers to the *V. eiseniae* msu genes obtained from the bi-directional best hit (BBH) search.

Functional enrichment analysis between *V. eiseniae* msu and *V. eiseniae* EF01-2 genes

The functional enrichment of the Gene Ontology (GO) terms associated with *Verminephrobacter eiseniae* msu genes in comparison with the *Verminephrobacter eiseniae* EF01-2 genome dataset was analyzed by using Fisher's exact test integrated within BLAST2GO version 5.0 (Glass and Girvan 2014). The annotated genes of *Verminephrobacter eiseniae* msu were used as the test set, and the *Verminephrobacter eiseniae* EF01-2 protein-coding gene annotations were used as the reference set

for the analysis. Two-tailed Fisher's test was carried out, and the corrected P value < 0.05 was taken as statistically significant.

Identification of symbiosis associated genes in *V. eiseniae* msu genome

The symbiosis-associated genes present in the genome dataset of *V. eiseniae* msu were identified by annotating the bacterial protein-coding genes against the Symbiosis database integrated within the GIPSy software (Soares et al. 2016) using the BLASTx algorithm with an E value cutoff of 1E−05. The metabolic pathways associated with these genes were identified from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation using the KAAS web server (Moriya et al. 2007). The functional analysis of the identified symbiotic genes was performed through Gene Ontology (GO) annotation using the OmicsBox software version 1.1 (https://www.biobam.com/omicsbox/).

Multi genome alignment, comparison of the orthologous genes, and phylogenomic analysis

The genome sequence of *Verminephrobacter eiseniae* msu was aligned to the genomic dataset of its neighboring proteobacteria species: *Alicycliphilus denitrificans* K601 (NC_015422), *Delftia acidovorans* SPH-1 (NC_010002), *Rhodoferax ferrireducens* T118 (NC_007908), *Verminephrobacter aporrectodeae* subsp. tuberculatae strain At4 (AFAL00000000), and *Verminephrobacter eiseniae* EF01-2 (NC_008786), by using Mauve multiple genome alignment tool (http://darlinglab.org/mauve/mauve.html) (Darling et al. 2004). The Mauve genome alignment provides the complete scenario of the conserved genomic regions across these species and also portrays the events of genomic rearrangement and horizontal gene transfer (Darling et al. 2004). The InDels and structural variations between the *V. eiseniae* msu strain and its neighboring proteobacterial genomes were detected by using the variant detection tools of the CLC Genomics Workbench. Simultaneously, the genome-wide comparisons of the orthologous gene clusters among these proteobacteria were analyzed by using the OrthoVenn web server (http://www.bioinfogenome.net/OrthoVenn/) (Wang et al. 2015). The phylogenomic analysis of the selected proteobacterial genomes was performed by REALPHY phylogeny builder server 1.2 (Bertels et al. 2014), and the phylogenomic tree was reconstructed through a maximum likelihood method by using the PhyML tool version 3.0 (Guindon et al. 2005). The pan and core genome analysis between the above-mentioned six bacterial strains were analyzed through the orthology calling approach of the GET_HOMOLOGUES software package (Contreras-Moreira and Vinuesa 2013). The software clusters the homologous gene families by using the bidirectional best hit (BDBH), COG, and OrthoMCL algorithms and estimates the pan and core genome sizes for the given strains. The genomic and pathogenic islands residing in the genome dataset of *V. eiseniae* msu were predicted by using the GIPSy software (Soares et al. 2016). Simultaneously, the virulence factors and pathogens associated with the *V. eiseniae* msu genome were predicted by annotating the AUGUSTUS-predicted *V. eiseniae* msu protein-coding genes against the Virulence Factor Database (VFDB) (Chen et al. 2005) using the following parameters: E value cutoff of 1E−10 (stringent condition) and minimum sequence homology of 60% and above. The antimicrobial resistance genes residing in the genome of *V. eiseniae* msu were screened by annotating the protein-coding genes (CDS) against the MEGARes database using the local BLASTn search with an E value threshold of 1E−10 (Lakin et al. 2016).

Results and discussion

Extraction of genome sequence reads, quality assessment, and assembly

The whole genome of earthworm *Eisenia fetida* was reported by Zwarycz et al. 2015, and though many of the proteobacteria species have been reported, the whole genome has been sequenced only for *V. eiseniae* EF01-2 and *V. aporrectodeae* subsp. *tuberculatae* At4 (Kjeldsen et al. 2012; Pinel 2009). The genome sequence assembly of *V. eiseniae* EF01-2 was used to fetch out the bacterial sequences from the genome dataset of raw reads sequences of earthworm *Eisenia fetida* (Figure S1). A total of 261,108,322 worm genome reads were concatenated and processed for short read alignment. Among these concatenated reads, a total of 593,130 paired-end reads with an average length of 100 bp and GC content of 65% were aligned to the genome dataset of *V. eiseniae* EF01-2. The aligned reads were sifted and subjected to quality assessment and trimming. After quality assessment and trimming of ambiguous, low-quality reads and adapter sequences, a total of 592,455 filtered reads were obtained with an average length of 89.1 bp (Table S1, Figure S2). The de novo assembly of the filtered reads using the CLC Genomics Workbench version 11.0.1 generated a total of 1832 contigs with a total genome size of 4,422,260 bp (4.4 Mb). The genome assembly statistics obtained from the CLC Genomics Workbench denoted the average length, N50, and GC% of the assembled contigs of a bacterial genome were 2414 bp, 3,593 bp, and 65.5%, respectively (Table S1). The bacterial genome sequence has been deposited in NCBI GenBank under the accession number SDQN00000000. The genome size of the type strain *V. eiseniae* EF01-2 was 5,597,943 bp and that of *V. aporrectodeae* was 4,681,801 bp.

In vertically transmitted, intracellular obligate symbionts, the genome size gets reduced over time. For instance, the genome size of symbiotic beta proteobacteria was around 2 Mbp. *Candidatus zinderia* insecticola has a 2.08-Mbp genome (McCutcheon and Moran 2010) and the genome of *Candidatus tremblaya* princeps is 1.39 Mbp in size (Von Dohlen et al. 2001), but generally, a bacterial genome ranges from about 4 to 6 Mbp. Isolation of pure endosymbionts in a niche is the reason for their smaller genome. An endosymbiotic bacterium has reduced the chance of contact with other bacterial species, and hence, the exchanges of DNA sequences from different bacterial species are rare. It leads to the genome streamlining of the symbiotic bacteria. But *Verminephrobacter* is having an average bacterial genome size, and there is a lack of genome erosion (Kjeldsen et al. 2012). The symbiotic relationship between the genus *Nephrothrix* and lumbricid earthworms happened much later than that between *Verminephrobacter* and earthworm (Kjeldsen et al. 2012). *Nephrothix* was shown to have switched hosts and became species-specific which could have occurred only through horizontal transmission. *Verminephrobacter* did not show any host switching evidence and is vertically transmitted via the cocoon. *Verminephrobacter* is an extracellular symbiont and a triangular association was established between the earthworm, *Verminephrobacter*, and *Nephrothrix* in the nephridia which provides the chance for exchange of genes between the organisms (Lund et al. 2014; Møller et al. 2015). Also, the biparental transmission of the symbiont provides a chance for the exchange of genome fragments (Paz et al. 2017). Simultaneously, the *Verminephrobacter* can incorporate the species-specific foreign DNA from the environment during the natural transformation within the earthworm egg capsule (Davidson et al. 2014). The ability of the symbiont to uptake foreign DNA may play an essential role to maintain their core genome and assist in the inclusion of the foreign genes within the host worm system (Davidson et al. 2014; Treangen et al. 2008). The genome completeness was evaluated by using the gVolante web server with the BUSCO ortholog search pipeline. The BUSCO sets were constructed by the orthologous group of genes, found as single-copy orthologs in at least 90% of the species. In our BUSCO analysis dataset, a total of 40 BUSCO orthologous groups were identified. Among them, the genome dataset of the extracted bacterial genome showed 85% of complete BUSCO orthologs and 7.5% each of the fragmented and missing BUSCOs (Table S2). The variant analysis demonstrated a total of 41,130 variants between the genomes of our *Verminephrobacter* bacteria and reported *Verminephrobacter eiseniae* EF01-2. Among these 41,129 variants, 1140 were multi-nucleotide variants (MNVs), 39,116 were single nucleotide variants (SNVs), and 873 were InDels and structural variants. Of the 873 InDels and structural variants, a total of 451 deletions, 259 insertions, 7 inversions, 150 replacements, and 6 other structural variants were detected between the 2 genomes (Table 1). This indicates that the organism in our study is a new strain of *Verminephrobacter eiseniae*. Hence, the bacterium was named *Verminephrobacter eiseniae* msu. Here, msu refers to Manonmaniam Sundaranar University, where the bacterial genome assembly was done. The list of all the SNVs and MNVs was documented in Table S3.

16S rRNA analysis and taxonomic classification

The generic feature format (GFF) file generated from RNAmmer prediction showed a 16S rRNA gene sequence of the extracted bacterial genome in contig 776 from sequence positions 306 to 1827 with an HMM alignment score of 1842.5. Only 1 16SrRNA sequence was identified in the extracted bacterial genome. The BLAST search of the predicted 16S rRNA sequence against the 16S ribosomal RNA sequence database in NCBI using the BLASTn algorithm showed that extracted bacterial 16S RNA molecule had the closest

Table 1 Summary of InDels and structural variants identified between the *Verminephrobacter eiseniae* msu (SDQN00000000) and *Verminephrobacter eiseniae* EF01-2 (NC_008786) genomes

Variant type	Variant subtype	No. of variants
Deletion	Self-mapped	353
Deletion	Paired breakpoint	59
Deletion	Cross-mapped breakpoints	39
Total (deletion)		**451**
Insertion	Self-mapped	138
Insertion	Paired breakpoint	49
Insertion	Close breakpoints	65
Insertion	Tandem duplication	7
Total (insertion)		**259**
Inversion	Cross-mapped breakpoints	7
Inversion	Paired breakpoint	0
Total (inversion)		**7**
Replacement	Paired breakpoint	150
Total (replacement)		**150**
Translocation	Multiple breakpoints	0
Total (translocation)		**0**
Complex	Cannot resolve sequence	5
Complex	Multiple breakpoints	1
Complex	Cross-mapped breakpoints (invalid orientation)	0
Total (complex)		**6**
Total (InDels and structural variants)		**873**

phylogenetic similarity with the proteobacteria *Verminephrobacter eiseniae* EF01-2 (99.87%), *Verminephrobacter aporrectodeae tuberculatae* strain At4 (96.48%), *Acidovorax delafieldii* strain 133 (95.31%), and *Acidovorax defluvii* strain BSB411 (94.88%). The top 50 16S rRNA sequences from the BLAST search were aligned together using the ClustalW tool. The Kimura 2-parameter substitution model with the discrete gamma distribution (+G) of 5 rate categories and evolutionary invariable sites (+I) was considered as the best model to describe the substitution pattern as it received the lowest BIC score of 15,390.05 (File S1). A maximum-likelihood phylogenetic tree constructed based on the Kimura 2-parameter substitution model using the MEGA 7 software placed the extracted genome along with the proteobacteria from the genus *Verminephrobacter* with a bootstrap confidence value of 100%. The strain showed close evolutionary relatedness and grouped together as a monophyletic clade with the proteobacteria *Verminephrobacter eiseniae* EF01-2 (NR_074705 and NR_043719), *Verminephrobacter aporrectodeae* subsp. tuberculatae strain At4 (NR_116575), and *Verminephrobacter aporrectodeae* subsp. caliginosae strain Ac9 (NR_116576) as they share the same common ancestor (Fig. 1). Simultaneously, the taxonomic affiliation of the new genome was verified using the Average Nucleotide Identity (ANI) with its neighboring taxa using the Orthologous Average Nucleotide Identity Tool (OAT). The ANI comparison of our *Verminephrobacter* bacterial genome demonstrated an OrthoANI score of 98.72% with the genome of *Verminephrobacter eiseniae* EF01-2, 81.92% with *Verminephrobacter aporrectodeae*, and 76.64% with *Variovorax paradoxus* (Figure S3). As the Average Nucleotide Identity between the genome of *Verminephrobacter eiseniae* msu and *Verminephrobacter eiseniae* EF01-2 is above the species demarcation cutoff value (> 95%) (Goris et al. 2007; Richter and Rosselló-Móra 2009), it indicates that the genome of our *Verminephrobacter* bacterial strain belongs to the species *Verminephrobacter eiseniae*. Simultaneously, we aligned the NCBI Prokaryotic Genome Annotation Pipeline (PGAP)-annotated 16S rRNA gene of our *Verminephrobacter* bacteria with the 3 previously reported full-length 16S rRNA genes of *Verminephrobacter eiseniae* EF01-2 (File S2). The 16S rRNA gene sequence alignment had a 0.13% mismatch.

Genome sequence annotation, mobilome analysis, and genome map visualization

A PGAP annotation of the bacterial genome predicted a total of 5895 CDS (4447 protein-coding genes and 1448 pseudogenes) and 40 RNA genes including 3 rRNAs, 34 tRNAs, and 3 other non-coding RNAs. The complete PGAP annotation details were given in the whole-genome shotgun sequencing project (WGS) with the GenBank accession number SDQN00000000. The annotation of the *V. eiseniae* msu genome using the RAST server predicted a total of 5963 protein-coding genes, which were categorized into 302 subsystems (Fig. 2a). The comparison of the RAST-annotated subsystem features between *V. eiseniae* msu and *V. eiseniae* EF01-2 denoted that among the subsystems, "amino acids and derivatives" (500 *V. eiseniae* msu genes and 481 *V. eiseniae* EF01-2 genes); "fatty acids, lipids, and isoprenoids" (183 *V. eiseniae* msu genes and 177 *V. eiseniae* EF01-2 genes); "membrane transport" (141 *V. eiseniae* msu genes and 131 *V. eiseniae* EF01-2 genes); and "nucleosides and nucleotides" (109 *V. eiseniae* msu genes and 79 *V. eiseniae* EF01-2 genes) were highly represented in *V. eiseniae* msu compared to the *V. eiseniae* EF01-2 (Fig. 2b). In contrast, the subsystems like "cofactors, vitamins, prosthetic groups, and pigments"; "cell wall and capsule"; "RNA metabolism"; and "stress response" were observed to be dominant in *V. eiseniae* EF01-2. The 3 RAST annotation terms membrane transport; fatty acid, lipids, and isoprenoids; and amino acids and derivatives were enriched in *V. eiseniae* msu compared to EF01-2. The importance of membrane transport and lipid transfer is well established in the legume-rhizobia symbiosis (Udvardi and Day 1997) and *Solemya velum* symbiosis (Conway and Capuzzo 1991). Earthworms may also get benefited from the supply of lipids and amino acids from *Verminephrobacter*. For instance, the riboflavin (vitamin B_2) acts as a major source for the autofluorescence property of earthworms, and it also supports the regeneration process of the worm upon amputation (Johnson Retnaraj Samuel et al. 2011; Subramanian et al. 2017). The worm lacks the ability to synthesize the riboflavin by itself. The genome dataset of *V. eiseniae* msu denoted the bacterium can successfully synthesize the riboflavin and may act as a potential supplier of the vitamin to the host species. The ability of the *Verminephrobacter* symbionts in supplying the essential vitamins and cofactors to their host worms has been reported previously (Lund et al. 2014). Simultaneously, the enhanced role of membrane transport is needed for nutrient transfer particularly for extracellular symbiosis (Smith et al. 1994). Besides, the annotated subsystem features denoted that 37 genes are associated with the "virulence, disease, and defense" feature including 18 genes for resistance to antibiotics and toxic compounds, 18 genes for invasion and intracellular resistance, and 1 gene for bacteriocins, ribosomally synthesized antibacterial peptides. The RAST genome annotation data were listed in Table S4. The AUGUST US ab initio gene prediction server using bacteria as reference species predicted a total of 3809 *V. eiseniae* msu protein-coding genes. The number of PGAP- and RAST-predicted protein-coding genes in our bacterial strain genome was higher compared to the other

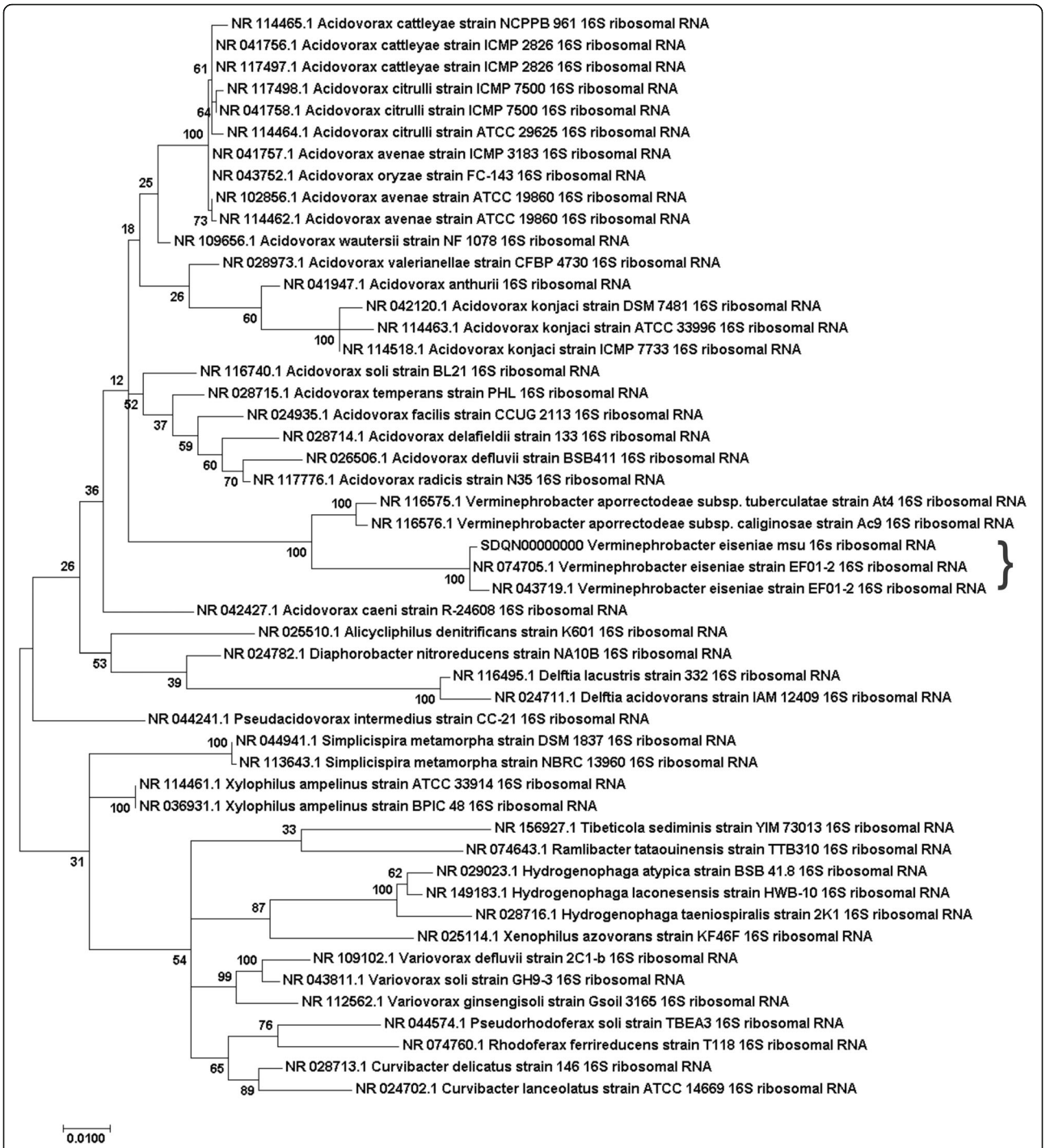

Fig. 1 Phylogenetic tree based on 16S rRNA comparison of *Verminephrobacter eiseniae* msu with its closely related homologs. The tree was constructed using the maximum likelihood method and bootstrapping of 100 replicates

reported *Verminephrobacter* symbionts (Kjeldsen et al. 2012). This may be due to the assembly/annotation error or the presence of redundant contigs in the genome dataset. In contrast, the number of AUGUSTUS-predicted protein-coding genes for *Verminephrobacter eiseniae* msu strain was considerably lower and found close to the predicted coding genes of *Verminephrobacter aporrectodeae* subsp. *tuberculatae* strain At4^{T} (*Vtu*) (Kjeldsen et al. 2012). Notably, the total genome size of our strain was also observed close to the genome size of *Vtu* symbiont. The mobile genetic elements detected by using the TEclass tool denoted the presence of 1613

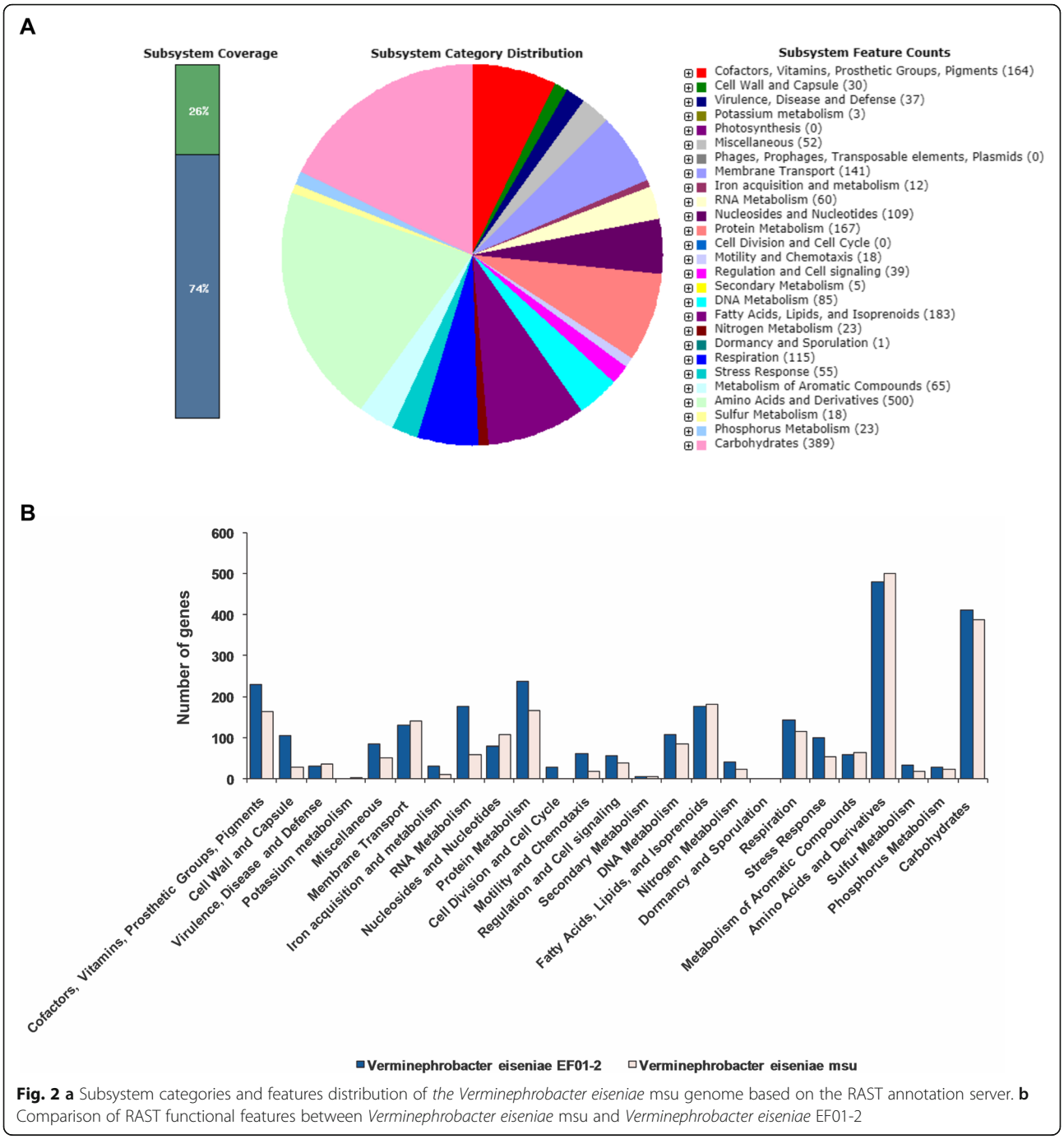

Fig. 2 a Subsystem categories and features distribution of *the Verminephrobacter eiseniae* msu genome based on the RAST annotation server. **b** Comparison of RAST functional features between *Verminephrobacter eiseniae* msu and *Verminephrobacter eiseniae* EF01-2

total transposons, consisting of 481 DNA transposons and 1132 retrotransposons. Of these 1132 retrotransposons, a total of 698 LINEs, 433 LTRs, and 1 SINEs were identified in the genome of *V. eiseniae* msu (Table S5A). Besides, the ISfinder tool using the ISsaga pipeline detected 34 ORFs associated with 14 insertion sequence (IS) families. Among the predicted IS families, ISL3, IS630, and IS21 family transposase were dominant within the genome dataset (Table S5B). The graphical circular genome map of *V. eiseniae* msu along with their genome annotation features is represented in Fig. 3.

Function and pathway analysis of *V. eiseniae* msu protein-coding genes

We have used the AUGUSTUS-predicted genes for functional and pathway analysis of the species. The non-coding RNAs, *cis*-regulatory elements, and other self-splicing RNAs present in *V. eiseniae* msu genome are listed in Table S6. The genome annotation comparison

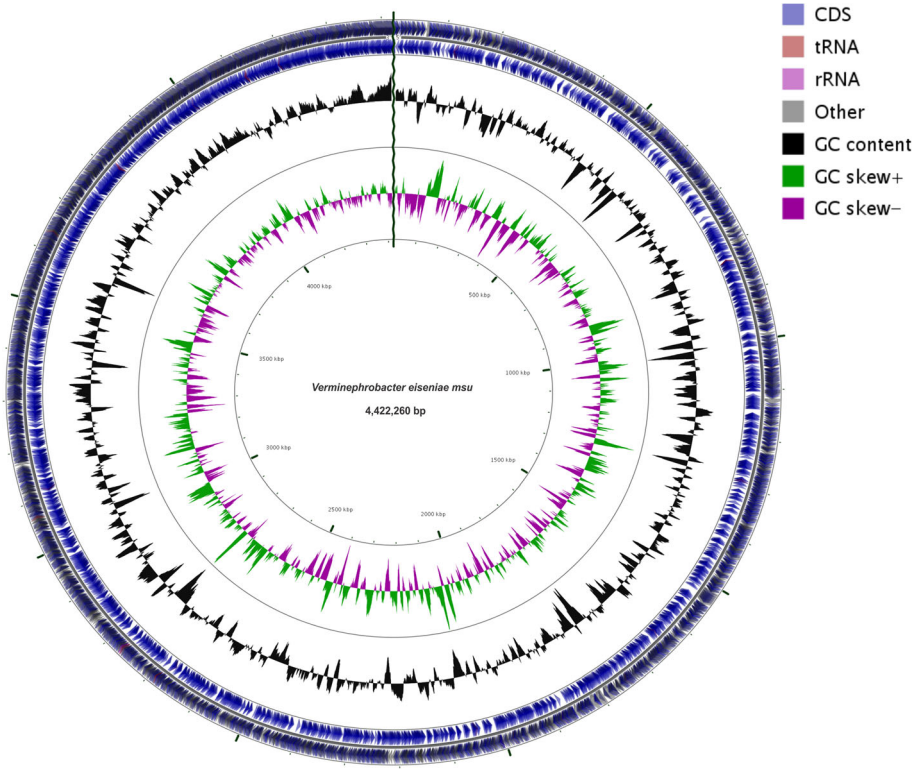

Fig. 3 Circular genome map of *Verminephrobacter eiseniae* msu generated by the CGView server. From outside to the center: rings 1 and 2 demonstrate the protein-coding genes (CDS) on both forward and reverse strand, ring 3 represents the GC content plot, and ring 4 represents both positive and negative GC skew

between *V. eiseniae* EF01-2 and *V. eiseniae* msu strains was demonstrated in Table S7. Out of the total 3809 AUGUSTUS-predicted genes, 3805 genes showed BLAST annotation hits to their homologous sequences in the NCBI nr database (Fig. 4a). The evolutionarily conserved domains, functional sites, and motif signatures associated with the translated protein sequences of *V. eiseniae* msu were annotated against the InterPro domain database using the InterProScan plug-in of BLAST2GO. A total of 3426 *V. eiseniae* msu protein sequences were successfully annotated against the InterPro database and subsequently categorized into 1184 domains. The top 30 InterPro domains/families obtained for the annotated *V. eiseniae* msu proteins were summarized in Fig. 4b. The domain distribution data denoted that the "P-loop containing nucleoside triphosphate hydrolase" (IPR027417) was the most highly represented domain with 280 protein sequences followed by "MetI-like superfamily" (IPR035906) (141 sequences), "NAD(P)-binding domain superfamily" (IPR036291) (125 sequences), and "winged helix-like DNA-binding domain superfamily" (IPR036388) (124 sequences). The P-loop NTPase is a prevalent domain, characterized by the presence of 2 signature motifs called Walker A and Walker B, which binds to the NTPs and Mg2+ cation, respectively (Walker et al. 1982). The domain assists in energy production by NTP hydrolysis and acts as a substrate for nucleotide binding (Ponesakki et al. 2017). The P-loop containing nucleoside triphosphate hydrolase superfamily proteins play a crucial role in determining the host specificity of the endophyte *Streptomyces scabrisporus* NF3 during symbiosis (Ceapă et al. 2018). Besides, the transcriptome analysis of the *Cardinium* strain *cEper1*, in its host parasitic wasps, *Encarsia suzannae* demonstrated the upregulation of the P-loop NTPase domain-containing gene CAHE_0544 upregulated within the male insects (Mann et al. 2017). The *V. eiseniae* msu protein dataset associated with P-loop containing nucleoside triphosphate hydrolase family denoted the presence of AAA family ATPase, ABC transporters, ATP-binding cassette domain-containing proteins, and DEAD/DEAH box helicase. Notably, the ABC transporters (ATP-binding proteins) play a major role in importing essential nutrients and exporting toxic substances (Davidson et al. 2008). Besides, the proteins couple the energy of ATP hydrolysis to facilitate the essential biological phenomena like DNA repair (Goosen and Moolenaar 2001) and translation elongation (Chakraburtty 2001). Previous reports have also suggested that the eukaryotes probably acquire class 1 and

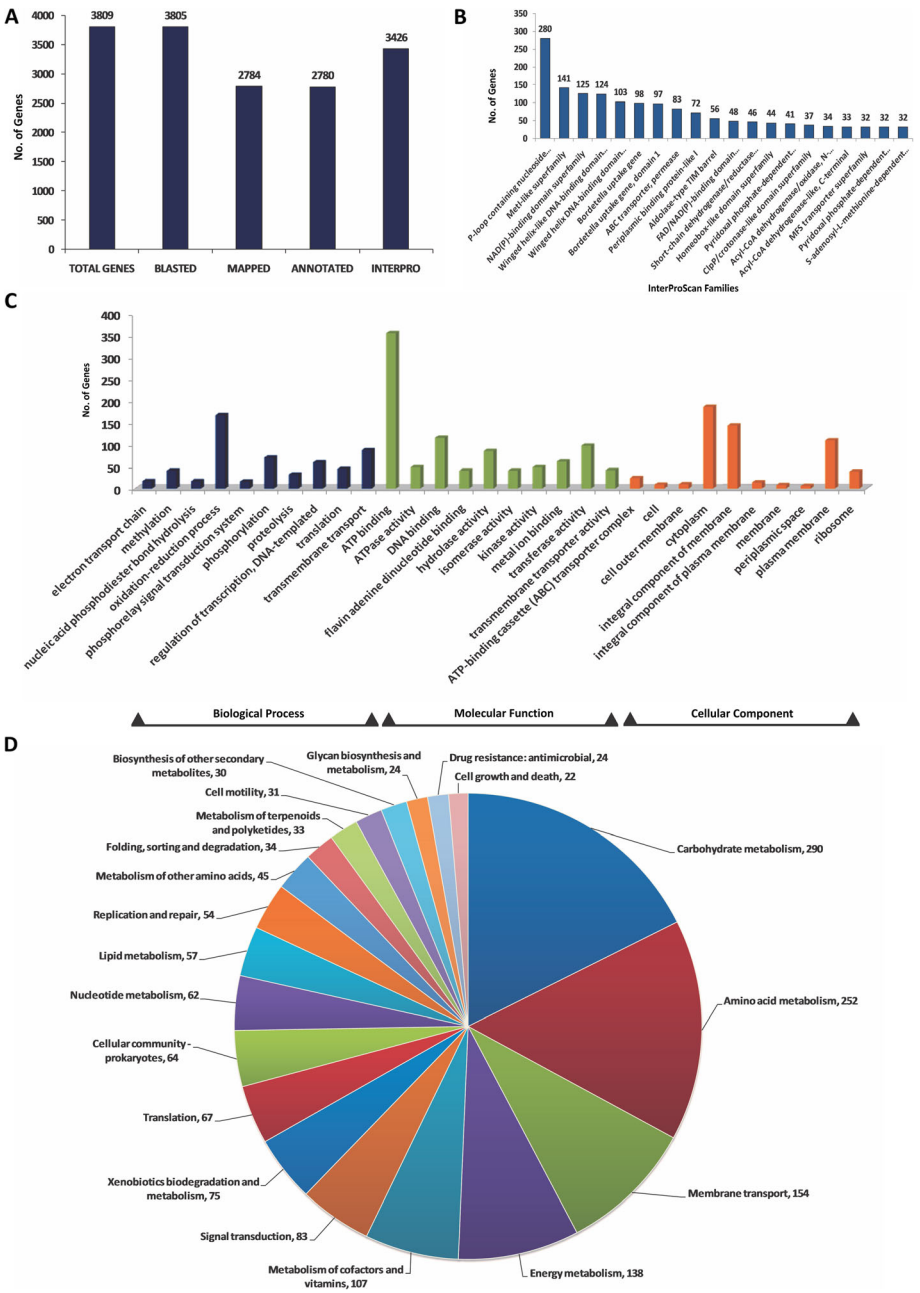

Fig. 4 **a** The data distribution summary generated from BLAST2GO annotation represents the BLAST hits, mapped sequences, Gene Ontology-annotated sequences, and InterProScan-annotated sequences. **b** Histogram of top 20 InterPro domains distribution. **c** Histogram representing the Gene Ontology (GO) distribution of the *Verminephrobacter eiseniae* msu protein-coding genes. The functionally annotated genes were classified into three GO categories: biological process (BP), molecular function (MF), and cellular component (CC). **d** Pie chart representing the distribution of KEGG pathways associated with the protein-coding genes of *Verminephrobacter eiseniae* msu

class 2 ABC transporters from their symbiotic bacterial systems (Davidson et al. 2008).

The Gene Ontology (GO) annotation of *V. eiseniae* msu protein-coding genes was performed through the mapping and annotation steps of BLAST2GO. Out of the 3723 *V. eiseniae* msu genes with nr BLAST hits, a total of 2784 genes were mapped with their associated GO terms, and among them, 2780 gene sequences were annotated to a total of 1414 GO terms. Of these annotated GO terms, 522 GO terms belong to the biological process (BP), 824 GO terms belong to the molecular function (MF), and 68 GO terms belong to the cellular component (CC). The GO distribution of the functionally annotated protein-coding genes (Fig. 4c) denoted

that within the biological process, the dominant subcategories are "oxidation-reduction process" (168 genes), "transmembrane transport" (88 genes), and "phosphorylation" (71 genes); among the molecular function, most of the genes were assigned to the subcategories like "ATP binding" (356 genes), "DNA binding" (116 genes), and "transferase activity" (98 genes); and within the cellular component, the most highly represented GO terms were "cytoplasm" (187 genes), "integral component of membrane" (144 genes), and "plasma membrane" (110 genes).

The Clusters of Orthologous Groups (COGs) database allows the prediction and classifications of the function more accurately and reliably based on the orthologous relation of the gene products. The COG analysis data of the *V. eiseniae* msu genome dataset obtained from the EggNog analysis denoted that a total of 3381 genes were assigned to 20 functional categories. Among the obtained functional groups, the cluster for "function unknown" (596 genes) constitutes the largest functional group (Figure S4). Among the other functional groups, the clusters for "amino acid transport and metabolism" (403 genes), "inorganic ion transport and metabolism" (355 genes), "energy production and conversion" (341 genes), and "transcription" (268 genes) were the highly represented categories. The dominance of amino acid transport and metabolism in the orthologous group dataset indicates the role of the bacteria in supporting the nitrogen recycling process of the host worm (Kjeldsen et al. 2012; Schramm et al. 2003). The RAST genome annotation data of *V. eiseniae* msu suggested that 20 genes of the bacterial strain are associated with ammonia assimilation including the proteins like glutamate synthase, glutamine synthetase, ammonium transporter, glutamate-ammonia-ligase adenylyltransferase, and [protein-PII] uridylyltransferase. The ammonia assimilation-related genes may play a key role in the nitrogen recycling process (Ankrah et al. 2017; Macdonald et al. 2012). Besides, we also observed the presence of denitrification-specific enzymes like nitrate reductase (EC 1.7.99.4) and nitrite reductase (EC 1.7.1.4), which are involved in converting the nitrates into gaseous nitrogen (Moreno-Vivián et al. 1999; Rinaldo and Cutruzzolà 2007; Tiso and Schechter 2015).

A total of 1796 genes were assigned to 39 KEGG pathways (Fig. 4d). Among these retrieved pathways, carbohydrate metabolism (290 genes), amino acid metabolism (252 genes), membrane transport (154 genes), and energy metabolism (138 genes) were the most dominant KEGG pathways observed in the genome dataset of *V. eiseniae* msu. The overall KEGG pathway analysis data suggested that most of the annotated *V. eiseniae* msu genes were assigned to the pathways associated with metabolism category (1113 genes), while a few genes were observed to be mapped with the pathways related to human diseases (113 genes) and organismal systems (46 genes).

Enrichment analysis of functional GO terms

The enrichment analysis of the Gene Ontology terms associated with the annotated *V. eiseniae* msu protein-coding genes in comparison with the *Verminephrobacter eiseniae* EF01-2 genome dataset was performed by using Fisher's two-tailed test with corrected P value < 0.05. Figure 5 denoted all the functionally enriched GO terms associated with the genome of *V. eiseniae* msu. The GO enrichment data denoted that "intracellular" (GO: 0005622) and "carboxylic acid metabolic process" (GO: 0019752), "acyl-CoA dehydrogenase activity" (GO: 0003995), and "carbohydrate binding" (GO:0030246) were the most enriched GO terms in the annotated genome dataset of *V. eiseniae* msu. In contrast, the GO terms like "DNA binding" (GO:0003677), "electron transport chain" (GO:0022900), and "endonuclease activity" (GO:0004519) were found to be enriched in the genome of *Verminephrobacter eiseniae* EF01-2. The ability to metabolize carboxylic acids helps *Verminephrobacter* to assimilate carboxylic acids present in the nephridial excretion (Rich et al. 2015). The enzyme having acetyl-CoA dehydrogenase activity is used in the metabolism of fatty acids. The acyl CoA dehydrogenase affects the first step of β-oxidation of fatty acids (Kurtz et al. 1998). The carbon content obtained through this fatty acid metabolism helps in the growth of *Verminephrobacter*.

Identification of symbiosis-associated genes in *V. eiseniae* msu gene pool

The symbiosis-associated genes of *V. eiseniae* msu were screened by comparing the bacterium genome dataset with 2834 symbiotic protein sequences present in the Symbiosis database of the GIPSy software. The BLAST search identified a total of 586 symbiotic genes in the *V. eiseniae* msu genome (Table S8). The bioluminescent bacteria *Vibrio fischeri* exclusively produce the enzyme 1-acyl-sn-glycerol-3-phosphate acyltransferase during their symbiotic relationship with *Euprymna tasmanica* which plays a major role in metabolism (Jones and Nishiguchi 2006). The same gene was also observed in the symbiotic gene pool of *V. eiseniae* msu. Besides, the bacterial strain contains ABC transporter ATP-binding protein which is shown to play a role in releasing cell signaling molecules in legumes—*Rhizobium* symbiosis (Sugiyama et al. 2007). Similarly, the gene for NADP-dependent malic enzyme, inevitable in nitrogen fixation by symbiotic *Rhizobium meliloti* (Driscoll and Finan 1993), was also observed in *V. eiseniae* msu genome.

The KEGG metabolic pathway annotation of the identified symbiotic genes suggested that most of the genes

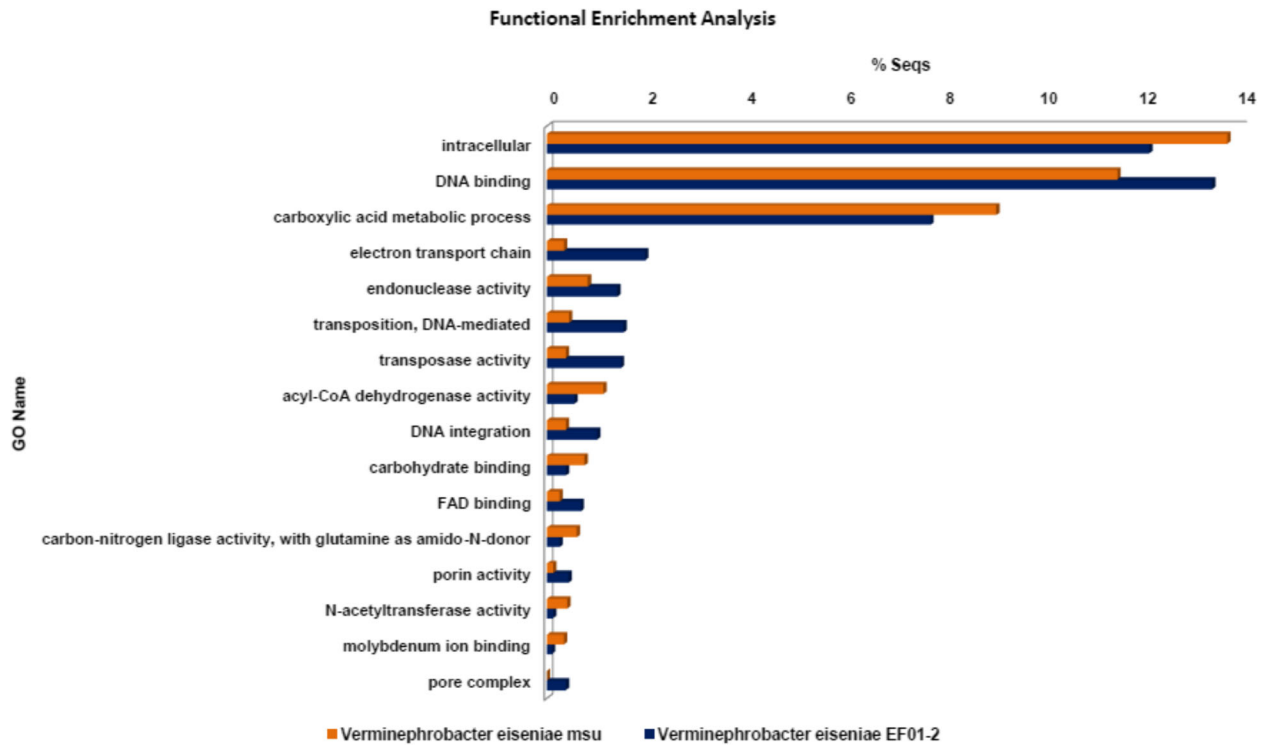

Fig. 5 Bar chart representing the functionally enriched Gene Ontology (GO) terms associated with the *Verminephrobacter eiseniae* msu genome in comparison with the *Verminephrobacter eiseniae* EF01-2 genome dataset

were assigned to the pathway carbohydrate metabolism followed by amino acid and energy metabolism (File S3A, B). The carbohydrates like mannose, fucose, and galactose are present in the glycosylated surfaces of the ampullar epithelium of the host (worm). The symbionts utilize the carbohydrates and sugar residues from the host surface glycans as a potential source of energy for their growth (Pinel et al. 2008; Sonnenburg et al. 2005). Simultaneously, it was observed that the two isolates of *Verminephrobacter aporrectodeae* At4^T and Ac9^T and *Verminephrobacter eiseniae* utilize a broad range of amino acids like alanine, aspartate, and glutamate; sugars like fucose, galactose, glucose, and mannose; and several fatty acids to grow aerobically. The sugar resources present in the earthworm cocoon provides the energy required for the vertical transmission of the symbiont from one host generation to another (Lund et al. 2012). Besides, the symbiont has a beneficial effect on earthworm reproduction as it supplies the essential vitamins and cofactors to the host cocoons and compensates for their nutrient deficiency (Lund et al. 2014). The functional analysis of the identified *V. eiseniae* msu symbiotic genes demonstrated that the oxidation-reduction process, ATP binding, ATPase activity, and integral components of the membrane were the dominant functional categories within the gene pool (File S3C).

Genome sequence alignment, comparison of orthologous gene clusters, and phylogenomic analysis

The RAST annotation of *the V. eiseniae* msu genome denoted the list of 30 closest neighbors of the bacterial species (Table S9). According to the RAST genome sequence comparison data, the top 5 closest neighboring species for *V. eiseniae* msu were identified as *Verminephrobacter eiseniae* EF01-2 (score 549), *Alicycliphilus denitrificans* K601 (score 506), *Acidovorax* sp. JS42 (score 506), *Delftia acidovorans* SPH-1 (score 503), and *Rhodoferax ferrireducens* DSM 15236 (score 487). The Mauve multiple genome alignment tool was used to align the genome sequence of *V. eiseniae* msu to its closest proteobacterial strains *Alicycliphilus denitrificans* K601 (NC_015422), *Delftia acidovorans* SPH-1 (NC_010002), *Rhodoferax ferrireducens* T118 (NC_007908), *Verminephrobacter aporrectodeae* subsp. tuberculatae strain At4 (AFAL00000000), and *Verminephrobacter eiseniae* EF01-2 (NC_008786). The Mauve alignment generated 84, 123, 515, 478, and 771 locally collinear blocks (LCBs) between the genomes of *A. denitrificans* K601-*V. eiseniae* msu, *D. acidovorans* SPH-1-*V. eiseniae* msu, *R. ferrireducens* T118-*V. eiseniae* msu, *V. aporrectodeae* At4(T)-*V. eiseniae* msu, and *V. eiseniae* EF01-2 and *V. eiseniae* msu with minimum LCB weight of 1033, 1546, 198, 497, and 960, respectively (Fig. 6a). The

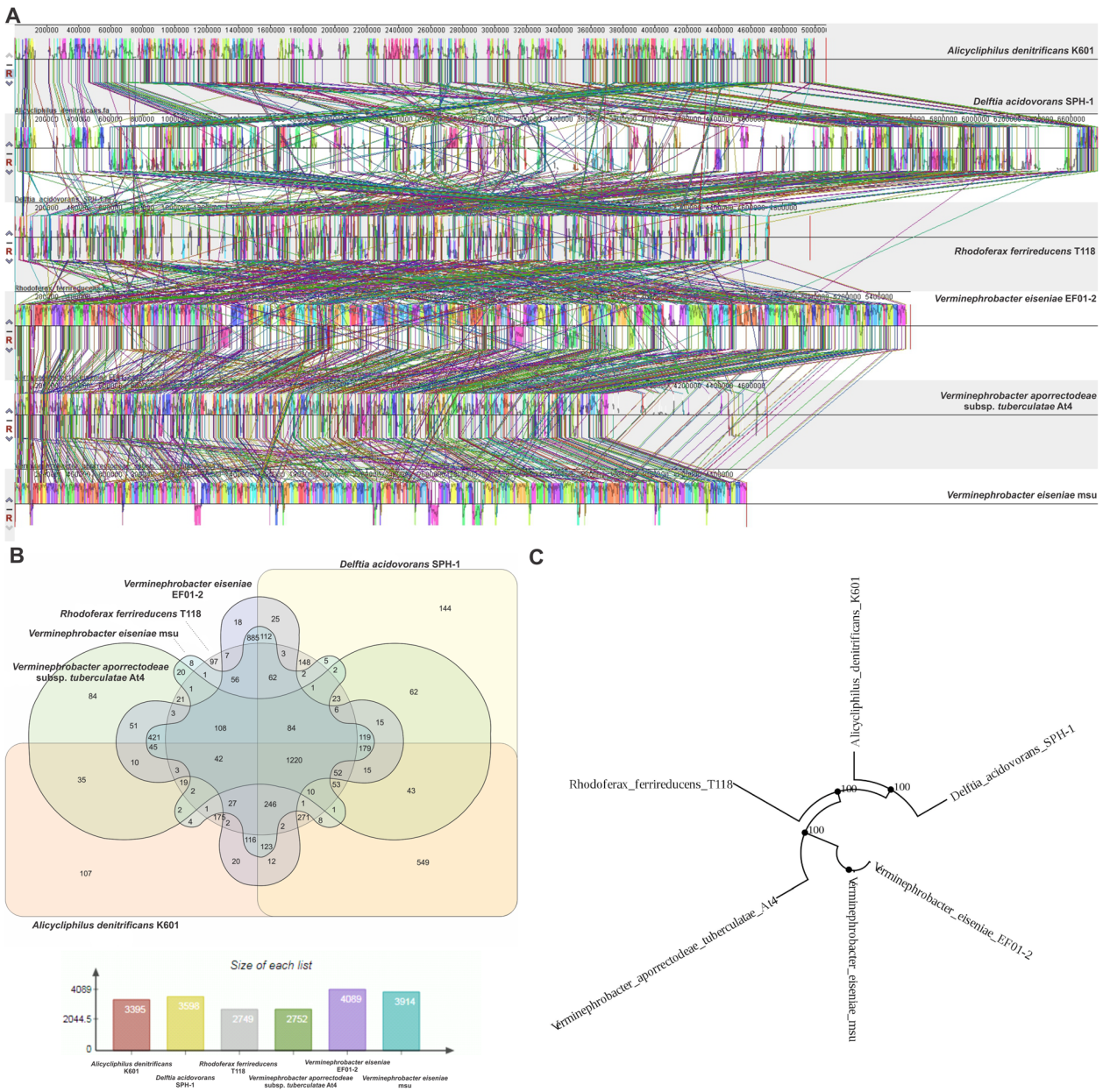

Fig. 6 a Mauve multiple genome sequence alignment between *Verminephrobacter eiseniae* msu and its neighboring proteobacteria. **b** Genome-wide comparison of the orthologous clusters between proteobacteria species: *Alicycliphilus denitrificans* K601, *Delftia acidovorans* SPH-1, *Rhodoferax ferrireducens* T118, *Verminephrobacter aporrectodeae* subsp. tuberculatae strain At4, *Verminephrobacter eiseniae* EF01-2, and *Verminephrobacter eiseniae* msu. **c** Phylogenomic tree constructed based on proteobacterial genome comparison using the PhyML tool

variant analysis demonstrated 1966, 2074, 1659, and 1434 InDels and structural variants between the genomes of our *V. eiseniae* msu strain and its closest proteobacteria *A. denitrificans* K601 (NC_015422), *D. acidovorans* SPH-1 (NC_010002), *R. ferrireducens* T118 (NC_007908), and *V. aporrectodeae* At4 (T) (AFAL00000000), respectively (Table S10A-D). Simultaneously, the comparison of the orthologous gene clusters across the proteome of these selected proteobacterial species was carried out by using the OrthoVenn web tool. A total of 20,497 gene clusters were obtained from OrthoVenn analysis, and among them, 3395, 3598, 2749, 2752, 4089, and 3914 clusters belong to the species *A. denitrificans* K601, *D. acidovorans* SPH-1, *R. ferrireducens* T118, *V. aporrectodeae* At4(T), *V. eiseniae* EF01-2, and *V. eiseniae* msu, respectively (Fig. 6b). A 6-way Edwards' Venn diagram data denoted that among the identified clusters, 1220 gene

clusters were commonly shared by all the 5 species. Besides, 421, 116, 112, 56, and 8 gene clusters were found to be annotated between *V. aporrectodeae* At4(T)-*V. eiseniae* EF01-2-*V. eiseniae* msu, *A. denitrificans*-*V. eiseniae* EF01-2-*V. eiseniae* msu, *D. acidovorans*-*V. eiseniae* EF01-2-*V. eiseniae* msu, *R. ferrireducens*-*V. eiseniae* EF01-2-*V. eiseniae* msu, and *A. denitrificans*- *D. acidovorans*-*V. eiseniae* msu, respectively (Fig. 6b). The phylogenomic analysis based on the genome sequence comparison of these proteobacteria was performed by using the REALPHY phylogeny builder web tool. A maximum likelihood phylogenomic tree data constructed by the PhyML server denoted a strong evolutionary relationship between the genome of *V. eiseniae* msu and *V. eiseniae* EF01-2 and grouped them as a monophyletic clade (Fig. 6c).

To obtain the pan and core genome information, the homologous gene families of the abovementioned 6 bacterial strains were calculated by using the BDBH, OMCL, and COG clustering strategies with minimum pairwise alignment coverage of 75%. The data denoted the presence of 15,424 COG clusters, 15,034 OMCL clusters, and 808 BDBH clusters (File S4A). Among them, 738 clusters were found to be a consensus between the 3 algorithms, and 13,506 clusters were common between the COG and OMCL algorithms. The BDBH strategy was used to estimate the pan and core genome sizes of the strains. We observed the fitted curves for both the pan and core genomes with residual standard errors of 620.08 and 366.03, respectively (File S4B, C). The fitted values used to estimate the pan and core genome sizes were given in File S4E and F. Besides, the software portioned the 13,506 pan-genome matrix clusters (13,506) common between the COG and OMCL into core, soft-core, shell, and cloud compartments. Among these 13,506 gene clusters, 761, 1375, 1255, and 10,876 clusters represented the core (genes conserved in all the genomes considered), soft-core (genes conserved in 95% of the genomes considered), shell (moderately conserved genes present in 3–4 genomes in our study), and cloud (rare genes present in ≤ 2 genomes in our study) genomes, respectively (File S4D). The core gene clusters conserved among the 6 strains were listed in Table S11. The comparison of protein-coding genes between *V. eiseniae* msu and *V. eiseniae* EF01-2 identified 7 *V. eiseniae* msu-specific genes missing in the genome dataset of *V. eiseniae* EF01-2 strain (File S4G). Among the identified 7 genes, we observed DNA damage-inducible protein D, IS5 family transposase, chromosome partitioning protein ParB, and 4 other hypothetical proteins. The *V. eiseniae* EF01-2 genome exhibited the presence of IS4 family transposase (Veis_4381; https://www.uniprot.org/uniprot/A1WR27) whereas the presence of IS5 family transposase is unknown. In contrast, the genome dataset of *V. eiseniae* msu demonstrated both IS4 and IS5 family transposase homologs. The transposases play a pivotal role in maintaining the genome plasticity and host adaptation of the bacteria (Vigil-Stenman et al. 2017). In extracellular luminal symbionts, the expansion of the transposon elements leads to their genome reduction (Hendry et al. 2018). Simultaneously, the previous study has highlighted the ability of the *Verminephrobacter* bacteria to incorporate the free DNA from environmental sources and enable its acquisition within the host worms. The ability to uptake the free DNA from the environment may serve as a potential source of nutrition and facilitate the DNA repair of the symbiont bacteria to maintain their genome stability (Davidson et al. 2014). The DinD (DNA damage-inducible protein D) gene plays a major role in recombinational DNA repair by modulating the Recombinase A (RecA) gene activity (Uranga et al. 2011). Notably, the RecA gene was also identified in the genome dataset of *V. eiseniae* msu.

The genomic and pathogenic island prediction using the GIPSy software identified a total of 9 genomic islands and 12 pathogenic islands throughout the concatenated genome of *V. eiseniae* msu. Among the identified pathogenic islands, pathogenic island 7 exhibited 55% virulence factors and both pathogenic islands 8 and 11 exhibited 53% virulence factors (Table S12A). Besides, the annotation of the *V. eiseniae* msu protein-coding genes against the Virulence Factor Database (VFDB) exhibited a total of 189 virulence factors having a sequence homology of 60% and above with their homologs in the database (Table S12B). Among the annotated virulence factors, the PilB gene codes for the traffic NTPase require for the assembly of the type IV pili (Davidson et al. 2014). The type IV pili play a regulatory role in colonizing the nephridia of the nascent earthworms and assist in the incorporation of the foreign genes within the earthworm egg capsules through natural transformation (Davidson et al. 2014; Dulla et al. 2012). As the virulence activity and pathogenicity of the *Verminephrobacter* species were not investigated to a significant extent, the pathogenic islands and the virulence factors demonstrated in our study can be utilized further to explore the pathogenic nature of the strain and interpret their association in the symbiotic relationship between the bacteria and the worm. Besides, we have identified 25 potential antimicrobial resistance genes in the genome dataset of *V. eiseniae* msu (Table S13). Lund et al. 2014, in their study, demonstrated that the *Verminephrobacter* symbionts have a beneficial effect on the host worm reproduction as it protects the developing embryos from the pathogens (Lund et al. 2014). But the antimicrobial properties of the bacteria are yet to be explored. In this scenario, the annotated

antimicrobial resistance genes identified in the *V. eiseniae* msu can be explored further to investigate the antimicrobial properties of the strain and interpret their impact on the reproduction of the host.

Bacteria were reported to supplement host organisms with a number of metabolites. Some of the bacteria living in the intestine of humans provide short-chain fatty acids like propionate, pyruvate, and acetate, and vitamins like thiamine, vitamin B2, folate, biotin, riboflavin, pantothenic acid, and vitamin K (LeBlanc et al. 2017). The gene pool of *V. eiseniae* msu also shows that the organism has the ability to produce vitamin B_2, thiamine, biotin, propionate, pyruvate, folate, and pantothenate. In zebrafish, bacteria depleted embryos resulted in abnormal neurobehavioral development (Phelps et al. 2017). This study triggers a thought of the possibility of bacterial symbiosis in vertebrates. The virulent genes of *Verminephrobacter* and *Mycobacterium* are closely related. To find a functional correlation, the property of *Verminephrobacter* and *Mycobacterium* was analyzed and found to both take about 15–20 days to form a colony in vitro (Pinel et al. 2008). Several genes like mce1 (mycobacterial cell entry protein) and phthiocerol synthesis polyketide synthase type I (ppsA and ppsB), which are associated with the virulence property of the mycobacteria, are also involved in their slow growth (Beste et al. 2009). Lewin et al. 2005 demonstrated all the highly pathogenic *Mycobacterium* species belong to the risk group 3 exhibit slow-growing property (Lewin and Sharbati-Tehrani 2005). It indicates a strong connection between the pathogenicity and the slow growth of the bacteria. Hence, the slow-growing property is probably due to the 18 *Mycobacterium* virulence operon homologs listed in Table S14. Among them, 4 genes are components of the ribosome, 4 genes are associated with NAD and NADP biosynthesis, and the products of 4 genes are components of the transcription machinery. The ribosome biosynthesis, NAD and NADP biosynthesis, and transcription-specific genes were previously connected with the slow growth rate of prokaryotic organisms like bacteria and yeast (Esquerre et al. 2013; García-Martínez et al. 2015; St John and Goldberg 1978; Szenk et al. 2017). The *Mycobacterium* is a well-known intracellular pathogen, whereas the *Verminephrobacter* acts as an extracellular symbiont. The slow growth also might be the reason for the successful symbiotic association as seen in squids—*Vibrio fischeri* and alfalfa-*Rhizobium meliloti* symbiosis wherein the bacterial doubling times were 5 h and 11 h, respectively (Gage et al. 1996; Lee and Ruby 1994). The earthworm provides the nitrogen source and shelter for the bacteria. However, the involvement of the host worm in supplying the carbohydrate source to the bacterial strain is still unknown. As described earlier, the genome of *V. eiseniae* msu contains glutamate and glutamine biosynthesis-specific enzymes, involved in ammonia assimilation. Besides, the RAST data also denoted the presence of alanine biosynthesis-specific enzymes and proteins associated with the urea cycle like urea carboxylase, urea ABC transporter, and several urease accessory proteins. Besides, the KEGG pathway annotation data demonstrated that among the 290 carbohydrate metabolism-specific genes, 29 genes were associated with glucose metabolism, 14 genes were assigned to fructose and mannose metabolism, and 11 genes were associated with galactose metabolism. The amino acids like glutamate, glutamine, alanine, and nitrogenous compounds like ammonia and urea were reported as the potential nitrogen sources, whereas the sugars like glucose, galactose, fructose, and mannose were the potential carbon sources required for the growth of the *Verminephrobacter* bacteria (Pinel et al. 2008).

Conclusion

Symbiotic organisms that support embryogenesis and normal homeostasis of living organisms are important to focus on the development of a better healthcare system. Our study characterized the genome of a new *Verminephrobacter* strain and highlighted the crucial genes and pathways, playing a pivotal role during the symbiotic relationship of the species with the host worm. The detailed analysis of the bacterial genome shed light on the mobilome, pathogenicity, virulence, and antimicrobial resistance of the strain and investigated the pan and core gene clusters with closest neighbors. In the *Verminephrobacter* genus, the third whole genome reported here will be helpful for understanding the symbiosis which supports embryogenesis and overall fitness of the host worms.

Supplementary information

Supplementary information accompanies this paper athttps://doi.org/10.1186/s13213-020-01549-w.

Additional file 1 Figure S1: Diagrammatic representation of the isolation of *Verminephrobacter eiseniae* msu genome sequence reads from earthworm genome through FastqSifter.

Additional file 2 Figure S2: Histogram representing the (A) length, (B) Phred quality score and (C) GC content distribution of filtered raw reads (after trimming).

Additional file 3 Figure S3: The average nucleotide identity between the genomes of *Verminephrobacter eiseniae* msu and its closely related orthologs. The heatmap generated with OrthoANI values were calculated using the OAT tool.

Additional file 4 Figure S4: Histogram showing Clusters of Orthologous Groups (COG) classification of the *Verminephrobacter eiseniae* msu protein-coding genes obtained from the EggNog database.

Additional file 5 File S1: Maximum Likelihood fits of 24 different nucleotide substitution models obtained from the Find Best DNA/Protein Models (ML) tool of MEGA 7 software.

Additional file 6 File S2: Multiple sequence alignment of the 16S rRNA gene sequence between *Verminephrobacter eiseniae* msu and *Verminephrobacter eiseniae* EF01-2.

Additional file 7 File S3: (A) Distribution of metabolic pathways associated with symbiosis specific genes in *Verminephrobacter eiseniae* msu genome. (B) List of KEGG metabolic subpathways associated with the *Verminephrobacter eiseniae* msu symbiotic genes. (C) The functional annotation summary (biological process, molecular function and cellular components) of the symbiosis related genes in *Verminephrobacter eiseniae* msu.

Additional file 8 File S4: The pan and core genome analysis summary of six proteobacterial strains (A) Venn diagram of the homologous gene families clusters generated by the BDBH, COG, and OMCL strategies. (B) The pan-genome size estimated with the Tettelin fits. (C) The core genome size estimated with the Tettelin fits. (D) Partitioning of the OMCL generated pan genomic matrix into shell, cloud, soft-core, and core compartments. (E) Fitted values used to estimate the pan-genome size. (F) Fitted values used to estimate the core genome size. (G) List of *Verminephrobacter eiseniae* msu specific genes absent in *Verminephrobacter eiseniae* EF01-2.

Additional file 9 Table S1: Summary statistics of de novo assembly of *Verminephrobacter eiseniae* msu genome.

Additional file 10 Table S2: Completeness assessment statistics of *Verminephrobacter eiseniae* msu genome assembly.

Additional file 11 Table S3: List of all the SNVs and MNVs identified between *Verminephrobacter eiseniae* EF01-2 and *Verminephrobacter eiseniae* msu genomes.

Additional file 12 Table S4: RAST annotation details of the *Verminephrobacter eiseniae* msu genome.

Additional file 13 Table S5: (A) Summary statistics of the DNA transposons and Retrotransposons identified in the genome dataset of *Verminephrobacter eiseniae* msu. (B) Summary statistics of the IS (insertion sequence) families identified in the genome dataset of *Verminephrobacter eiseniae* msu.

Additional file 14 Table S6: List of non-coding RNA genes present in *Verminephrobacter eiseniae* msu predicted by Infernal software.

Additional file 15 Table S7: Genome annotation comparison between *Verminephrobacter eiseniae* EF01-2 and *Verminephrobacter eiseniae* msu.

Additional file 16 Table S8: List of symbiosis associated genes present in the genome dataset of *Verminephrobacter eiseniae* msu.

Additional file 17 Table S9: List of closely related bacteria to *V. eiseniae* msu obtained from RAST.

Additional file 18 Table S10: Summary of InDels and Structural Variants identified between (A) *Verminephrobacter eiseniae* msu and *Alicycliphilus denitrificans* K601 genomes, (B) *Verminephrobacter eiseniae* msu and *Delftia acidovorans* SPH-1 genomes, (C) *Verminephrobacter eiseniae* msu and *Rhodoferax ferrireducens* T118 genomes and (D) *Verminephrobacter eiseniae* msu and *V. aporrectodeae* At4 (T) genomes.

Additional file 19 Table S11: List of core genes identified from core genome analysis of proteobacteria species: *Alicycliphilus denitrificans* K601, *Delftia acidovorans* SPH-1, *Rhodoferax ferrireducens* T118, *Verminephrobacter aporrectodeae* subsp. tuberculatae strain At4, *Verminephrobacter eiseniae* EF01-2 and *Verminephrobacter eiseniae* msu.

Additional file 20 Table S12: (A) List of genomic and pathogenic islands detected in the concatenated *Verminephrobacter eiseniae* msu genome. (B) List of virulence genes detected in the *Verminephrobacter eiseniae* msu genome.

Additional file 21 Table S13: List of antimicrobial resistance genes identified in the genome dataset of *Verminephrobacter eiseniae* msu.

Additional file 22 Table S14: List of *Mycobacterium* virulence operon homologs in the genome of *V. eiseniae* msu.

Acknowledgements
AA is thankful to the Department of Biotechnology, India, for SRF (DBT/2015/MSU/447), and all authors thank the DBT Bioinformatics Infrastructure Facility (BT/BI/04/055/2001) at Manonmaniam Sundaranar University for providing the instrument facilities and University Grants Commission (UGC), New Delhi, India [F.No. 43-70/2014 (SR)].

Authors' contributions
AA and SP carried out the bacterial genome assembly, genome completeness evaluation, 16S rRNA analysis, identification of the bacterial strain, genome annotation, functional enrichment analysis, multi genome alignment and phylogenomic analysis. They also participated in the interpretation of the data and manuscript writing. SL participated in genome annotation, prediction of protein coding genes and their function and pathway analysis. RB participated in the design of the study, genome data analysis and interpretation. SS participated in the conception and design of the study, financial support, data analysis and interpretation and manuscript writing. All authors read and approved the final manuscript.

Funding
Department of Biotechnology, India (DBT/2015/MSU/447), DBT Bioinformatics Infrastructure Facility (BT/BI/04/055/2001), University Grants Commission (UGC), New Delhi, India [F.No. 43-70/2014 (SR)].

Ethics approval and consent to participate
N/A

Consent for publication
N/A

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Biotechnology, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu 627012, India. [2]Department of Environmental Energy and Engineering, Kyonggi University, Youngtong-gu, Suwon 16227, South Korea.

Received: 1 August 2019 Accepted: 29 January 2020
Published online: 25 February 2020

References
Abrusán G, Grundmann N, DeMester L, Makalowski W (2009) TEclass—a tool for automated classification of unknown eukaryotic transposable elements. Bioinformatics 25:1329–1330

Andrews S (2016) FastQC: a quality control tool for high throughput sequence data, p 2010

Ankrah NY, Luan J, Douglas AE (2017) Cooperative metabolism in a three-partner insect-bacterial symbiosis revealed by metabolic modeling. J Bacteriol 199: e00872–e00816

Ashburner M et al (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25:25

Aziz RK et al (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75

Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E (2014) Automated reconstruction of whole-genome phylogenies from short-sequence reads. Mol Biol Evol 31:1077–1088

Beste DJ, Espasa M, Bonde B, Kierzek AM, Stewart GR, McFadden J (2009) The genetic requirements for fast and slow growth in mycobacteria. PLoS One 4: e5349

Ceapă CD et al (2018) Genome mining of Streptomyces scabrisporus NF3 reveals symbiotic features including genes related to plant interactions. PloS One 13: e0192618

Chakraburtty K (2001) Translational regulation by ABC systems. Res Microbiol 152: 391–399

Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q (2005) VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33:D325–D328

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Contreras-Moreira B, Vinuesa P (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol 79:7696–7701

Conway N, Capuzzo JM (1991) Incorporation and utilization of bacterial lipids in the Solemya velum symbiosis. Marine Biology 108:277–291
Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403
Davidson AL, Dassa E, Orelle C, Chen J (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol Mol Biol Rev 72:317–364
Davidson SK, Dulla GF, Go RA, Stahl DA, Pinel N (2014) Earthworm symbiont Verminephrobacter eiseniae mediates natural transformation within host egg capsules using type IV pili. Front Microbiol 5:546
Davidson SK, Powell R, James S (2013) A global survey of the bacteria within earthworm nephridia. Mol Phylogenet Evol 67:188–200
Davidson SK, Stahl DA (2008) Selective recruitment of bacteria during embryogenesis of an earthworm. ISME J 2:510–518. https://doi.org/10.1038/ismej.2008.16
Dimijian GG (2000) Evolving together: the biology of symbiosis, part 1. In: Baylor University Medical Center Proceedings, vol 3. Taylor & Francis, pp 217a–2226a
Driscoll BT, Finan TM (1993) NAD+-dependent malic enzyme of Rhizobium meliloti is required for symbiotic nitrogen fixation. Mol Microbiol 7:865–873
Dubilier N, Bergin C, Lott C (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Microbiol 6:725
Dulla GF, Go RA, Stahl DA, Davidson SK (2012) Verminephrobacter eiseniae type IV pili and flagella are required to colonize earthworm nephridia. ISME J 6:1166
Esquerre T, Laguerre S, Turlan C, Carpousis AJ, Girbal L, Cocaign-Bousquet M (2013) Dual role of transcription and transcript stability in the regulation of gene expression in Escherichia coli cells cultured on glucose at different growth rates. Nucleic Acids Res 42:2460–2472
Fukui S (2014) Evolution of symbiosis with resource allocation from fecundity to survival. Naturwissenschaften 101:437–446
Gage DJ, Bobo T, Long SR (1996) Use of green fluorescent protein to visualize the early events of symbiosis between Rhizobium meliloti and alfalfa (Medicago sativa). J Bacteriol 178:7159–7166
García-Martínez J et al (2015) The cellular growth rate controls overall mRNA turnover, and modulates either transcription or degradation rates of particular gene regulons. Nucleic Acids Res 44:3643–3658
Glass K, Girvan M (2014) Annotation enrichment analysis: an alternative method for evaluating the functional properties of gene sets. Scientific Rep 4:4191
Goosen N, Moolenaar GF (2001) Role of ATP hydrolysis by UvrA and UvrB during nucleotide excision repair. Res Microbiol 152:401–409
Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91
Grant JR, Stothard P (2008) The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res 36:W181–W184
Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33:W557–W559
Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol 30:1229–1235
Hendry TA, Freed LL, Fader D, Fenolio D, Sutton TT, Lopez JV (2018) Ongoing transposon-mediated genome reduction in the luminous bacterial symbionts of deep-sea ceratioid anglerfishes. mBio 9:e01033–e01018
Hoff KJ, Stanke M (2013) WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res 41:W123–W128
Huerta-Cepas J et al (2015) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293
Johnson Retnaraj Samuel SC et al (2011) Autofluorescence in BrdU-positive cells and augmentation of regeneration kinetics by riboflavin. Stem Cells Dev 21:2071–2083
Jones B, Nishiguchi M (2006) Differentially expressed genes reveal adaptations between free-living and symbiotic niches of Vibrio fischeri in a fully established mutualism. Can J Microbiol 52:1218–1227
Kjeldsen KU et al (2012) Purifying selection and molecular adaptation in the genome of Verminephrobacter, the heritable symbiotic bacteria of earthworms. Genome Biol Evol 4:307–315
Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
Kurtz DM et al (1998) Targeted disruption of mouse long-chain acyl-CoA dehydrogenase gene reveals crucial roles for fatty acid oxidation. Proc Nat Acad Sci 95:15592–15597
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 35:3100–3108
Lakin SM et al (2016) MEGARes: an antimicrobial resistance database for high throughput sequencing. Nucleic Acids Res 45:D574–D580
LeBlanc JG, Chain F, Martín R, Bermúdez-Humarán LG, Courau S, Langella P (2017) Beneficial effects on host energy metabolism of short-chain fatty acids and vitamins produced by commensal and probiotic bacteria. Microbial Cell Factories 16:79
Lee I, Ouk Kim Y, Park S-C, Chun J (2016) OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. https://doi.org/10.1099/ijsem.0.000760
Lee K-H, Ruby EG (1994) Effect of the squid host on the abundance and distribution of symbiotic Vibrio fischeri in nature. Appl Environ Microbiol 60:1565–1571
Lewin A, Sharbati-Tehrani S (2005) Das langsame Wachstum von Mykobakterien Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz. 48:1390–1399
Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760
Lund MB, Holmstrup M, Lomstein BA, Damgaard C, Schramm A (2010) Beneficial effect of Verminephrobacter nephridial symbionts on the fitness of the earthworm Aporrectodea tuberculata. Appl Environ Microbiol 76:4738–4743
Lund MB, Kjeldsen KU, Schramm A (2014) The earthworm—Verminephrobacter symbiosis: an emerging experimental system to study extracellular symbiosis. Front Microbiol 5:128
Lund MB, Schätzle S, Schramm A, Kjeldsen KU (2012) Verminephrobacter aporrectodeae sp. nov. subsp. tuberculatae and subsp. caliginosae, the specific nephridial symbionts of the earthworms Aporrectodea tuberculata and A. caliginosa. Antonie Van Leeuwenhoek 101:507–514
Macdonald SJ, Lin GG, Russell CW, Thomas GH, Douglas AE (2012) The central role of the host cell in symbiotic nitrogen metabolism. Proc Royal Soc B Biol Sci 279:2965–2973
Mann E, Stouthamer CM, Kelly SE, Dzieciol M, Hunter MS, Schmitz-Esser S (2017) Transcriptome sequencing reveals novel candidate genes for Cardinium hertigii-caused cytoplasmic incompatibility and host-cell interaction. MSystems 2:e00141–e00117
McCutcheon JP, Moran NA (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol Evol 2:708–718
Møller P, Lund MB, Schramm A (2015) Evolution of the tripartite symbiosis between earthworms, Verminephrobacter and Flexibacter-like bacteria. Front Microbiol 6:529
Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F (1999) Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. J Bacteriol 181:6573–6584
Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35:W182–W185
Mulder NJ, Apweiler R (2008) The InterPro database and tools for protein domain analysis. Curr Protocols Bioinform 21:2.7. 1–2.7. 18
Nishimura O, Hara Y, Kuraku S (2017) gVolante for standardizing completeness assessment of genome and transcriptome assemblies. Bioinformatics 33: 3635–3637
Paul S, Arumugaperumal A, Rathy R, Ponesakki V, Arunachalam P, Sivasubramaniam S (2018) Data on genome annotation and analysis of earthworm Eisenia fetida. Data Brief 20:525–534
Paz L-C, Schramm A, Lund MB (2017) Biparental transmission of Verminephrobacter symbionts in the earthworm Aporrectodea tuberculata (Lumbricidae). FEMS Microbiol Ecol 93
Phelps D et al (2017) Microbial colonization is required for normal neurobehavioral development in zebrafish. Sci Rep 7:11244
Pinel N (2009) Physiological and genomic insight into the biology of Verminephrobacter eiseniae, a bacterial symbiont of the earthworm Eisenia fetida. PhD Thesis University of Washington, Seattle
Pinel N, Davidson SK, Stahl DA (2008) Verminephrobacter eiseniae gen. nov., sp. nov., a nephridial symbiont of the earthworm Eisenia foetida (Savigny). Int J Syst Evol Microbiol 58:2147–2157
Ponesakki V, Paul S, Mani DKS, Rajendiran V, Kanniah P, Sivasubramaniam S (2017) Annotation of nerve cord transcriptome in earthworm Eisenia fetida. Genomics Data 14:91–105
Raina J-B, Eme L, Pollock FJ, Spang A, Archibald JM, Williams TA (2018) Symbiosis in the microbial world: from ecology to genome evolution. Biology Open 7: bio032524
Rathy R et al (2018) Data on genome sequencing, analysis and annotation of a pathogenic Bacillus cereus 062011msu. Data Brief 17:15–23

Rich CD, Blaine AC, Hundal L, Higgins CP (2015) Bioaccumulation of perfluoroalkyl acids by earthworms (Eisenia fetida) exposed to contaminated soils. Environ Sci Technol 49:881–888
Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Nat Acad Sci 106:19126–19131
Rinaldo S, Cutruzzolà F (2007) Nitrite reductases in denitrification. In: Biology of the nitrogen cycle. Elsevier, pp 37–55
Schramm A, Davidson SK, Dodsworth JA, Drake HL, Stahl DA, Dubilier N (2003) Acidovorax-like symbionts in the nephridia of earthworms. Environ Microbiol 5:804–809
Smith S, Gianinazzi-Pearson V, Koide R, Cairney J (1994) Nutrient transport in mycorrhizas: structure, physiology and consequences for efficiency of the symbiosis. Plant Soil 159:103
Soares SC et al (2016) GIPSy: genomic island prediction software. J Biotechnol 232:2–11
Sonnenburg JL et al (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science 307:1955–1959
St John A, Goldberg AL (1978) Effects of reduced energy production on protein degradation, guanosine tetraphosphate, and RNA synthesis in Escherichia coli. J Biol Chem 253:2705–2711
Steel M, Penny D (2000) Parsimony, likelihood, and the role of models in molecular phylogenetics. Mol Biol Evol 17:839–850
Subramanian ER, Sudalaimani D, Christyraj JRSS, Christyraj JDS, Renganathan K, Krishnan S, Sivasubramaniam S (2017) Studies on organogenesis during regeneration in the earthworm, Eudrilus eugeniae, in support of symbiotic association with Bacillus endophyticus. Turkish J Biol 41:113–126
Sugiyama A, Shitan N, Yazaki K (2007) Involvement of a soybean ATP-binding cassette-type transporter in the secretion of genistein, a signal flavonoid in legume-Rhizobium symbiosis. Plant Physiol 144:2000–2008
Szenk M, Dill KA, de Graff AM (2017) Why do fast-growing bacteria enter overflow metabolism? Testing the membrane real estate hypothesis. Cell Syst 5:95–104
Tatusova T et al (2016) NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624
Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
Tiso M, Schechter AN (2015) Nitrate reduction to nitrite, nitric oxide and ammonia by gut bacteria under physiological conditions. PloS One 10: e0119712
Treangen TJ, Ambur OH, Tonjum T, Rocha EP (2008) The impact of the neisserial DNA uptake sequences on genome evolution and stability. Genome Biol 9:R60
Udvardi MK, Day DA (1997) Metabolite transport across symbiotic membranes of legume nodules. Annu Rev Plant Biol 48:493–523
Uranga LA, Balise VD, Benally CV, Grey A, Lusetti SL (2011) The Escherichia coli DinD protein modulates RecA activity by inhibiting postsynaptic RecA filaments. J Biol Chem 286:29480–29491
Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M (2011) ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol 12:R30
Viana F, Paz LC, Methling K, Damgaard CF, Lalk M, Schramm A, Lund MB (2018) Distinct effects of the nephridial symbionts Verminephrobacter and Candidatus Nephrothrix on reproduction and maturation of its earthworm host Eisenia andrei. FEMS Microbiol Ecol 94. https://doi.org/10.1093/femsec/fix178
Vigil-Stenman T, Ininbergs K, Bergman B, Ekman M (2017) High abundance and expression of transposases in bacteria from the Baltic Sea. ISME J 11:2611
Von Dohlen CD, Kohler S, Alsop ST, McManus WR (2001) Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. Nature 412:433
Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha-and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1:945–951
Wang Y, Coleman-Derr D, Chen G, Gu YQ (2015) OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res 43:W78–W84
Waterhouse RM et al (2017) BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35:543–548
Zdobnov EM et al (2016) OrthoDB v9. 1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res 45:D744–D749
Zwarycz AS, Nossa CW, Putnam NH, Ryan JF (2015) Timing and scope of genomic expansion within Annelida: evidence from homeoboxes in the genome of the earthworm Eisenia fetida. Genome Biol Evol 8:271–281

Publisher's Note