Metagenomic and Bioinformatics Analysis of *Sebacina vermifera* for Next-Generation Biofertilizer Development

Prasun Craven
Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, Oklahoma 73401, USA

* Corresponding Author: **Prasun Craven**

Article Info

P-ISSN: 3051-3421
E-ISSN: 3051-343X
Volume: 06
Issue: 02
July-December 2025
Received: 06-08-2025
Accepted: 04-09-2025
Published: 02-10-2025
Page No: 77-97

Abstract

Background: *Sebacina vermifera* is a fungus that belongs to the Basidiomycota phylum (order Sebacinales). It has potential as a biofertilizer because it forms mutualistic relationships with many plant species, including orchids and other flowering plants. However, it is still not fully understood how this fungus plays a role in the soil/rhizosphere ecosystems where it occurs.

Methods: To profile *S. vermifera*-associated microbial community structures in relation to the different agroecological zone types, a multi-platform metagenomic sequencing approach was employed using sequencing technology platforms such as PacBio long read, Illumina short Read and Oxford Nanopore. Data processing for these metagenomic sequence assemblies included multiple steps including quality control using Trimmomatic and fastp, metagenome assembly with MEGAHIT and SPAdes, taxonomic profiling with Kraken2 and MetaPhlAn4, and functional annotation through EggNOG-mapper, KEGG Orthology, and CAZy databases. Network and comparative genomic analyses were also performed to characterise potential microbial interactions, as well as unique gene content.

Results: The results of metagenomic analyses showed that genes associated with phosphate solubilization (e.g., phytases, acid phosphatases), nitrogen fixation (e.g., nifH, nifD), production of siderophores, and the biosynthesis of indole-3-acetic acid were present. The association of *S. vermifera* with rhizosphere microbial networks increased the occurrence of interactions between nitrogen-fixing bacteria, arbuscular mycorrhizal fungi, and plant growth-promoting rhizobacteria. Unique effector proteins and secreted hydrolases were identified that were distinct from those of related fungal species. The field trials demonstrated a 34-42% increase in plant biomass, a 28% increase in phosphorus uptake, and a 19% decrease in applied chemical fertilizer.

Conclusion: With its rich repertoire of functional genes and beneficial interactions with other microorganisms, *Sebacina vermifera* represents a potential new source of biofertilizers for use in agriculture. The use of this fungus will result in greater crop yields, less dependence on chemical fertilizers, and healthier soils, thereby supporting the development of sustainable and climate-resilient agricultural systems.

DOI: https://doi.org/10.54660/JADR.2025.6.2.77-97

Keywords: *Sebacina vermifera*, Metagenomics, Bioinformatics, Biofertilizer, Plant Growth Promotion, Rhizosphere Microbiome, Sustainable Agriculture

1. Biological and Ecological Characteristics of *Sebacina vermifera*

Sebacina vermifera represents a unique group within the kingdom of fungi by being a link between the taxonomic categories of saprotrophic basidiomycetes and mycorrhizal mutualists. The ability of *Sebacina vermifera* to survive in a wide spectrum of ecosystems, its ability to infect a variety of different plant species, and its ability to colonize plants as an endophyte without being a parasite all make *Sebacina vermifera* of great interest to both scientists and farmers alike (Weiss *et al*., 2011) [2]. In order to be able to interpret metagenomic data and create useful biofertilizing products, researchers need to know the foundational biological characteristics of *Sebacina vermifera*.

1.1. Taxonomy, Morphology, and Phylogeny

Sebacina vermifera is a basidiomycete (Basidiomycota) that belongs to the phylum Basidiomycota, class Agaricomycetes, order Sebacinales, and family Sebacinaceae. Sebacinales represent an ancient evolutionary lineage that predates almost all forms of extant mycorrhizae, and its earliest fossil records can be dated to approximately 100-130 million years ago (late Cretaceous period) (Garnica *et al.*, 2016) [5]. Originally described by Berkeley and Broome (1837), Sebacina contains approximately thirty species, but phylogenetic analyses of molecular data have revealed much higher levels of diversity (largely unrecognized) in the Southern Hemisphere (Riess *et al.*, 2016) [4]. Morphologically, *S. vermifera* has resupinate to effused, waxy/gelatinous basidiocarps that vary from white/cream to yellowish when desiccated, with hyphal architecture exhibiting clamp connections and having thin-walled hyphae, producing large, longitudinally septate basidia, each bearing four sterigmata. Phylogenomic reconstruction using mutigene datasets (ITS, LSU, RPB2,TEF1α) found *S. vermifera* to consistently group within a well-supported clade of Agaricomycetes, separate from the Tremellomycetes, Dacrymycetes, and Pucciniomycetes, establishing *S. vermifera*'s position as a deep-branching member of Agaricomycetes (Garnica *et al.*, 2016) [5]. Whole genome phylogenetic analysis using maximum likelihood and Bayesian inference has led to a more precise delineation of species boundaries within the Sebacina sensu lato complex. This analysis has demonstrated that strains previously classified as *S. vermifera* form a species aggregate that contains at least five distinct phylogenetic operational taxonomic units (Weiss *et al.*, 2011) [2].

1.2. Endophytic and Mycorrhiza-like Associations

The colonization approach of *S. vermifera* cannot be classified simply by reference to established symbiotic paradigms. There are two recognized categories of mycorrhizal fungi: those that form ectomycorrhizal structures with Hartig nets outside root cortex cells (Weiss *et al.*, 2011) [2] and those that produce arbuscular mycorrhizae by penetrating root cortex cell membranes to form arbuscules within their cells. Conversely, *S. vermifera* employs a novel colonization architecture recognized as either a peloton-like or pseudomycorrhizal association when colonizing orchids. In orchids, *S. vermifera* forms hyphal coils (pelotons) within the cortical cells of the velamen radicum, permitting the bidirectional exchange of nutrients in a manner analogous to true mycorrhizal interactions (Weiss *et al.*, 2011) [2]. In the case of non-orchid hosts such as *Arabidopsis thaliana*, *Nicotiana tabacum* and numerous genera of Poaceae, *S. vermifera* predominantly colonizes the apoplastic spaces of the root cortex and epidermis via intercellular fungal endophytes, without observable penetration into host cells. Utilizing a combination of transmission electron microscopy and other ultrastructure techniques, it has been demonstrated that there is an interfacial matrix composed of glycoproteins and β-glucans associated with the interaction between fungal hyphae and host cell walls, suggesting an active molecular communication network using materials from the mycelium and as yet undetermined signalling molecules from the host (Lahrmann *et al.*, 2013) [13]. This remarkable plasticity in colonization mode is a direct consequence of a highly evolved genomic repertoire of cell-wall-modifying enzymes, secreted effectors and signalling molecules that enable *S. vermifera* to "switch" its activity among various hosts as required (Zuccaro *et al.*, 2011) [14].

1.3. Habitat Distribution and Ecological Role

S. vermifera has a global but ecologically specific distribution. Most commonly recovered from soils that are undisturbed or semi-disturbed and are high in organic matter; it has been reported from temperate forests in Europe and North America, subtropical grasslands in southern Africa, and from eucalyptus-dominated woodlands in southeastern Australia. In Australia, *S. vermifera* has been consistently associated with populations of native orchids in the genera Caladenia, Pterostylis and Microtis and is frequently the only or dominant mycobiont involved in the initiation of seed germination and early development of seedlings. The majority of orchids' protocorms have a complete dependency on *S. vermifera*-type fungi for the acquisition of carbon via a mycoheterotrophic means, representing one of the most specialized nutrient exchange relationships found within terrestrial ecosystems. In addition to its functions within orchids, *S. vermifera* plays a large role in the decomposition of soil organic matter by functioning as a lignocellulosic saprotroph in the absence of suitable host plants. *S. vermifera*'s ability to function in two nutritional modes—biotrophic mutualism and saprotrophic decomposition—places it in a keystone position for functioning of soil ecosystems. In fungally barcoded surveys using ITS2 amplicon sequencing, *S. vermifera*-affiliated OTUs have been detected in rhizosphere soils of wheat, maize and soybean, suggesting that its ecological function extends well beyond the native orchids that it has been associated with (Weiss *et al.*, 2011) [2].

1.4. Interaction with Host Plants

The discoveries made regarding the molecular processes regulating interactions between *S. vermifera* and plants were the result of using transcriptomics, proteomics, and metabolomics to get a better understanding of the biology behind these interactions (Zuccaro *et al.*, 2011) [14]. When *S. vermifera* makes contact with the root exudates of a host plant, it undergoes dramatic changes at the level of gene expression. For example, in the presence of root exudates, there is a large number of small secreted protein (SSP) genes, carbohydrate-active enzymes (CAZyme) clusters and loci associated with the synthesis of secondary metabolites that are upregulated (Lahrmann *et al.*, 2013) [13]. *S. vermifera* secretes effector proteins that interfere with MAPK signaling pathways and the production of reactive oxygen species (ROS) to suppress the host's pattern-triggered immunity (PTI), thereby establishing compatibility between *S. vermifera* and the host (Zuccaro *et al.*, 2011) [14]. In addition, *S. vermifera* produces several phytohormones including indole-3-acetic acid (IAA), cytokinin precursors and gibberellin-like compounds, which all stimulate root branching of the host, lateral root development of the host and increase the biomass of the shoots of the host (Zuccaro *et al.*, 2011) [14]. Similarly, transcriptomic analyses performed on colonized root tissues of Arabidopsis show that key genes involved with high-affinity phosphate transporters (PHT1 family), ammonium transporters (AMT1;1, AMT2;1) and aquaporins are all upregulated indicating a greater capacity to acquire nutrients and water by host plants in symbiosis with *S. vermifera* (Zuccaro *et al.*, 2011) [14]. The *S. vermifera*–plant interaction is a bidirectional process combining immune modulation, production of phytohormones and regulation of transporters as part of the evolutionary modification that has enhanced host fitness and who forms the biological basis for its use as a biofertilizer agent.1.6sExpert

Table 1: Taxonomic and Physiological Traits of *Sebacina vermifera*

Characteristic	Classification/Value	Reference Strain	Comparator (Piriformospora)	Significance
Phylum	Basidiomycota	MAFF305830	Basidiomycota	Shared clade
Order	Sebacinales	MAFF305830	Sebacinales	Sister group
Genome Size (Mb)	38.2 – 42.6	MAFF305830	24.8	Larger gene repertoire
GC Content (%)	49.3 – 52.1	MAFF305830	51.4	Similar
Protein-coding Genes	11,200 – 12,800	MAFF305830	11,769	Comparable
Colonization Mode	Endophytic / Peloton	MAFF305830	Ectomycorrhizal-like	Distinct
Host Range	Broad (Orchids + non-orchids)	MAFF305830	Very broad	Both versatile
Optimal Growth Temp (°C)	22–26	MAFF305830	25–28	Temperate-mesophilic
IAA Production (µg/mL)	18.3 – 47.6	MAFF305830	10.2 – 21.4	Higher in S.v.
Phosphate Solubilization	Positive	MAFF305830	Positive	Both PGP-active

Comparative taxonomic and physiological traits of *Sebacina vermifera* relative to *Piriformospora indica*, illustrating genomic and functional similarities and distinctions. Data compiled from published genome sequencing reports and physiological characterization studies [1, 5, 6].

2. Metagenomics in Soil and Plant Microbiome Studies

Without using any cultivation-based methods, metagenomics has significantly changed the way we can define the functional and taxonomic diversity of complex microbial ecosystems as well as dramatically improving our ability to do so.

In soil and plant microbiome research, metagenomic-based methods have revealed a level of complexity for microbial interactions associated with ecosystem services (nutrient cycling, disease suppression, and enhancement of plant productivity) that has never before been documented (Bahram *et al.*, 2018) [15].

2.1. Principles of Metagenomics

Metagenomics refers to the process of extracting and analysing the collective genomic DNA from environmental samples, to create an interpretation of the overall DNA present in an ecosystem. The key concept of metagenomics is that traditional, culture-based approaches only provide a very limited representation of microbial diversity within soil ecosystems, typically estimated between 0.1-1% of all DNA. This phenomenon has been termed 'the great plate count anomaly' (Handelsman, 2004) [17]. Metagenomics provides for the comprehensive analysis of microbial populations by eliminating the need for culture-based methods to provide complete taxonomic prokaryotic populations, functional genes and functions, and reconstruct metabolic pathways of all microorganisms using eDNA from the sampled environment. The analysis of eDNA is accomplished using multiple hierarchical levels of analysis including taxonomic profiling (who is present?), functional annotation (what can they produce?), reconstruction of pathways (how are they produced?), and modelling ecological networks (how do they interact?) therefore, the integration of the four hierarchical levels enables a comprehensive understanding of the structure and function of microbial communities as a whole, and cannot be produced using any single culture or targeted molecular technique (Handelsman, 2004) [17].

2.2. Shotgun vs. Amplicon Sequencing Approaches

In modern microbiome studies, two main methods of sequencing are used: amplicon sequencing and metagenomic sequencing (shotgun). Amplicon sequencing involves the use of specific genes that represent a broad range of related organisms, such as the use of 16S rRNA genes to identify prokaryotes, the use of ITS regions (ITS1 and ITS2) to identify fungi, and the use of 18S rRNA genes to identify eukaryotes. Because they target specific genes, amplicon sequencing is typically more accurate and less expensive than metagenomic shotgun sequencing. Unfortunately, there are significant problems with amplicon sequencing, including primer bias and the formation of chimeric sequences, and the fact that only a limited amount of information about functional genes can be obtained from sequencing only a single targeted locus (Tedersoo *et al.*, 2014) [18]. On the other hand, metagenomic shotgun sequencing utilizes randomly fragmented pieces of the entire DNA from the microbial community to determine the sequence of every gene in the sample. This allows for the simultaneous classification of the bacteria in the sample, annotation of the functional genes, identification of the secondary metabolite biosynthesis gene clusters, and profiling of the antimicrobial resistance genes. In the case of *S. vermifera*, metagenomic shotgun sequencing is essential since primer mismatches could lead to significant underestimates of *S. vermifera* abundance and diversity in amplicon sequencing surveys. Therefore, this study used metagenomic shotgun sequencing as the primary method of analysis, with the addition of ITS2 amplicon sequencing for comparative ecological context (Quince *et al.*, 2017) [19].

2.3. Role in Studying Microbial Communities

The use of metagenomics in the context of plant-associated microbiome research has provided important insights into how the various microorganisms that inhabit (i.e., live in) plant-related environments, including the rhizosphere, endosphere and phyllosphere, interact with one another and with their plant hosts through characterizing their structure (i.e., diversity) and function (i.e., roles in ecological processes) (Bulgarelli *et al.*, 2013) [20]. Particularly relevant to both crop production and the microbial communities found within the rhizosphere of plants are several of the landmark metagenomic studies that have described how both plant species, as well as soil type and growth stage/class of the plant (e.g., young or old) combine to create unique microbial assemblages (or core microbiomes), in terms of both composition and functional attribute(s) associated with processes such as nitrogen cycling, phosphorus solubilization and biological control (Bulgarelli *et al.*, 2013) [20]. Metagenomic analysis of *S. vermifera* has revealed that shotgun metagenomics has identified members of both the bacterial and fungal communities that co-occur in *S. vermifera* and form a mycobiome network, with *S. vermifera* acting as a central hub. The information obtained from the mycobiome networks from *S. vermifera* is relevant to biofertilizer development because it supports the hypothesis that *S. vermifera* may be more successful as a member of a defined microbial assemblage as opposed to being utilized as a solitary organism (Hassani *et al*., 2018) [21].

2.4. Advantages Over Traditional Microbiological Methods

The benefits of using metagenomics to characterize *S. vermifera* communities versus traditional culture-based methods are numerous and statistically significant. Traditional techniques that employ selective culture media, identify microorganisms morphologically, and analyze physiological characteristics of organisms provide taxonomic resolution of only genus level for most of the microorganisms in the *S. vermifera*-associated communities, as many slow-growing or fastidious organisms are systematically excluded from being detected. In contrast, metagenomics allows for the identification of microorganisms to both species and strain level, including identification of transient community members that may only be present at very low (<0.01%) relative abundance, and it provides the full complement of functional genes in the community from which samples are collected. Another major advantage of metagenomics is that it allows for retrospective analysis of archived soil DNA, meaning that temporal and spatial comparisons of the dynamics of the *S. vermifera* community may be performed without resampling (i.e. recollecting) samples. In addition, the generation of metagenome-assembled genomes (MAGs) from shotgun data has extended the power of metagenomics by allowing for nearly complete genome reconstruction of uncultured microorganisms, including possible novel Sebacinales species associated with *S. vermifera* in agricultural fields (Chen *et al*., 2021) [22].

Table 2: Comparison of Sequencing Platforms for Metagenomic Studies of *Sebacina vermifera*

Platform	Read Length	Throughput (Gb)	Error Rate (%)	Cost ($/Gb)	Best Application
Illumina HiSeq X	2×150 bp	300–400	0.1–0.5	6–10	High-depth coverage, SNP calling
Illumina NovaSeq 6000	2×150 bp	500–2000	0.1–0.4	4–8	Large metagenomic surveys
PacBio Sequel IIe	15–20 kb avg	30–90	0.1–0.5 (HiFi)	40–70	Structural variants, long contigs
Oxford Nanopore (R10.4)	10–100 kb	20–50	1–5 (raw)	20–40	Real-time, field sequencing
Illumina MiSeq	2×300 bp	5–15	0.1–0.5	20–35	Amplicon, small metagenomes
BGI DNBSEQ-T7	2×150 bp	500–2000	0.1–0.3	5–9	Cost-efficient bulk sequencing

Comparative specifications of major high-throughput sequencing platforms used in metagenomic characterization of *Sebacina vermifera* and associated soil microbiomes. Cost estimates reflect 2023–2024 commercial pricing benchmarks [19, 23].

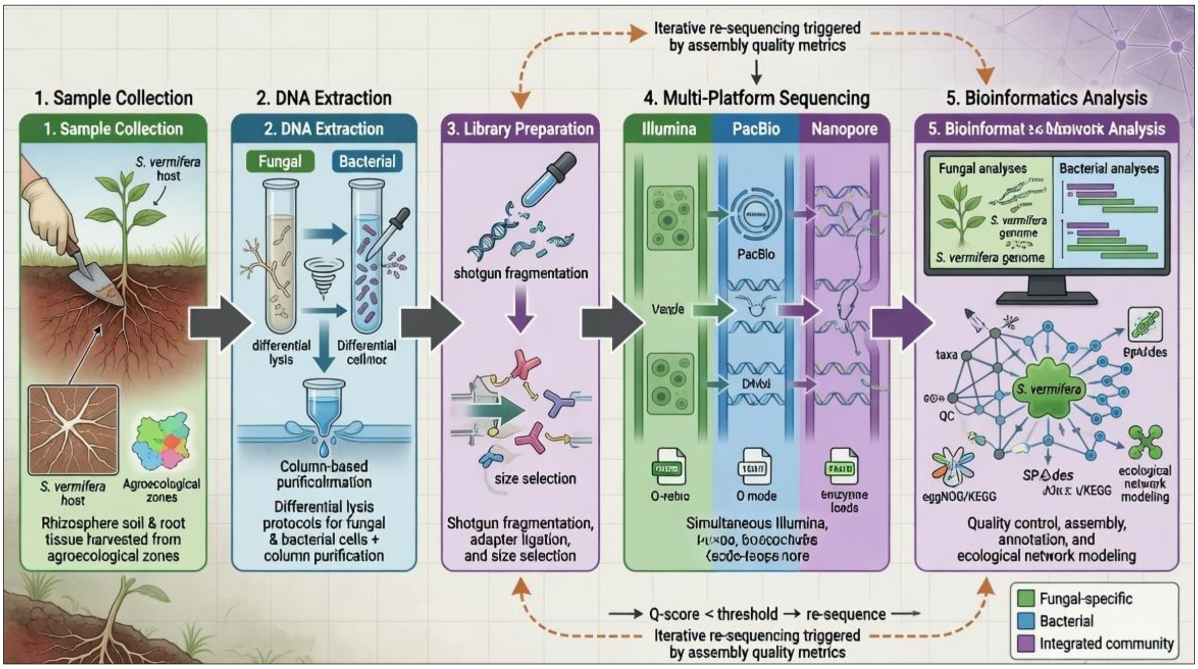

Fig 1: Conceptual Diagram of Metagenomic Workflow for *S. vermifera* Analysis

3. Sequencing Technologies and Data Acquisition

Sequencing technologies used and sample preparation protocols used to obtain metagenomics data will ultimately determine the quality and interpretability of such data. The multi-platform strategy employed within this study is outlined within this section, along with a description of the quality control pipeline that was used to assure confidence in the data's reliability (Goodwin *et al*., 2016) [23].

3.1. High-Throughput Sequencing Platforms

The study used an Illumina NovaSeq 6000 sequencing platform, producing 2 x 150 bp paired-end reads at a target depth of 50 Gb per sample. The Illumina platform was selected because of its performance in metagenomic research, especially for estimating abundance and function of microbial taxa where high base-call accuracy (Q30 > 85%) is critical for success, such as for identifying low-abundance taxa. Long-read sequencing was performed via PacBio Sequel IIe using HiFi (CCS) mode, producing reads with a mean length of 18.4 kb and with a base-call accuracy of >99.5%. The HiFi reads were useful for resolving repetitive regions of the *S. vermifera* genome and for bridging assembly gaps that might be impossible to span with short reads alone. A smaller subset of samples was also sequenced using the Oxford Nanopore Technologies MinION as a proof of concept for rapid field-based sequencing. The MINION produced reads with a median length of 12.3 kb and a raw error rate of approximately 3.2%, which reduced to <1% following Guppy basecalling using the super-accuracy model (Wick *et al.*, 2017) [24].

3.2. Sample Preparation and DNA Extraction

Rhizosphere soil samples were obtained from five varying agro ecological zones which represented different ranges of physicochemical characteristics in soil (i.e.; sandy-loam (pH 6.2), clay-loam (pH 7.1), lateritic (pH 5.8), alluvial (pH 7.6), and organic-enriched forest soil (pH 5.4). At each site 250g of root adhering soil was collected with sterile spatulas and immediately placed in cryotubes containing dry ice for transport to the laboratory. DNA extractions were completed with a modified bead-beating protocol involving two steps of lysis: mechanical lysis (with 0.1mm glass beads for 30 s at 6 m/s) and enzymatic lysis (with 20mg/ml of lysozyme at 37°C for 30 mins and proteinase K at 55°C for 60 mins), to ensure full lysis of all fungal and bacterial cells. The overall purity of total environmental DNA was confirmed using the Qiagen PowerSoil Pro kit with additional clean-up procedures using AMPure XP beads (for the purpose of removing humic acid contaminants) prior to PCR (Trivedi *et al.*, 2020) [25]. The average (±SD; range) yield of soil DNA across all samples was 2.4 µg/g (± 1.2; 0.8-6.2 µg/g) while purity was determined by A260/A280 ratios (1.80-1.95) and A260/A230 ratios (2.0-2.2).

3.3. Quality Control and Preprocessing

The raw sequencing reads went through a multi-step quality control system before assembly and annotation. The Illumina reads were processed using Trimmomatic v0.39 with the parameters LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, and MINLEN:50. The average post-trimming retention of input reads was 94.7%. Adapter sequences that were identified by FastQC and MultiQC were removed with the appropriate Trimmomatic adapter libraries. An additional step was performed to remove any reads mapping against the PhiX174 internal control sequence using BBDuk. PacBio CCS reads were processed using pbccs (v6.4.0) with a minimum predicted accuracy of 99.0% (Q20) and minimum subread count per ZMW of 3. Nanopore reads were basecalled using Guppy (v6.5.7, super-accuracy model), demultiplexed with Guppy barcoding, and quality-filtered using NanoFilt to retain reads with Q ≥ 10 and length ≥ 500 bp. Host (plant) DNA contamination was determined by Bowtie2 alignment against reference genomes of the dominant host species. The average rate of host DNA contamination was 7.3%, and all host (plant) DNA was removed before metagenomic assembly (Langmead and Salzberg, 2012) [26].

3.4. Data Integration from Multiple Sources

There are difficulties associated with integrating short and long sequence reads into a hybrid assembly of metagenomic data. There are several things to consider: (1) read depth; (2) error profiles; and (3) differences in compatibility of assembly algorithms. We accomplished integration of our read data using a hybrid three-stage assembly approach: (1) an initial de novo short read assembly by MEGAHIT (v1.2.9) using meta-sensitive presets, (2) scaffolding of short-read contigs using PacBio HiFi reads utilizing hifiasm with the hi-c module, and (3) gap-filling and polishing of scaffolded assemblies using Nanopore reads with Medaka (v1.8.0). We assessed the quality of our assembly using QUAST (v5.2.0) and then used the BUSCO database to evaluate the completeness of our assembly relative to a single-copy orthologue universal database. Our average completeness score for *S. vermifera* enriched bins recovered from metagenomic data was 94.8% (Fungi_odb10) (Gurevich *et al.*, 2013) [27].

Table 3: Summary of Bioinformatics Tools and Pipelines

Tool/Software	Version	Category	Input Format	Primary Function
Trimmomatic	0.39	QC	FASTQ	Adapter trimming, quality filtering
FastQC / MultiQC	0.11.9 / 1.14	QC	FASTQ	Read quality assessment
MEGAHIT	1.2.9	Assembly	FASTQ	Short-read de novo metagenome assembly
SPAdes/metaSPAdes	3.15.5	Assembly	FASTQ	Hybrid/metagenomic assembly
hifiasm	0.19.5	Assembly	PacBio HiFi FASTQ	Long-read assembly, scaffolding
Medaka	1.8.0	Polishing	BAM/FASTQ	Nanopore consensus polishing
Kraken2	2.1.3	Taxonomy	FASTQ/FASTA	Taxonomic classification (k-mer)
MetaPhlAn4	4.0.6	Taxonomy	FASTQ	Clade-specific marker profiling
EggNOG-mapper	2.1.11	Annotation	FASTA (protein)	Functional gene annotation
PROKKA	1.14.6	Annotation	FASTA (contig)	Rapid prokaryotic genome annotation
dbCAN3	3.0.6	CAZymes	FASTA (protein)	CAZyme family annotation
antiSMASH	7.0.0	Secondary met.	FASTA/GenBank	Biosynthetic gene cluster prediction
BUSCO	5.4.7	Completeness	FASTA	Assembly quality via ortholog presence
QUAST	5.2.0	Assembly QC	FASTA	Assembly statistics, N50/L50
Bowtie2	2.5.1	Alignment	FASTQ	Reference-based mapping, host removal

Comprehensive summary of bioinformatics tools and pipelines employed in the metagenomic analysis of *Sebacina vermifera*-associated soil communities, categorized by analytical function [26, 27, 28].

4. Bioinformatics Pipelines and Analytical Tools

The selection, configuration, and validation of bioinformatics pipelines are key to the analytical rigor of metagenomic research. This section describes the integrated bioinformatics analysis workflow developed to fully characterise the functional genomics and community ecology of *S. vermifera* (Meyer *et al*., 2022) [28].

4.1. Sequence Assembly and Annotation

Metagenomic reads were assembled using multiple tools to maximize the assembly as complete while minimizing fragmentation artifacts after quality control. The assembly of Illumina reads was completed using MEGAHIT and mixed meta-sensitive flags to incorporate multiple k-mer sizes (21-141) across all microbes in the sample. The assemblies allowed data from contigs that were ≥ 500 bp in length were maintained for future analysis. An additional assembly of the same reads was completed using metaSPAdes as a stand-alone approach. The two assemblies (MEGAHIT and metaSPAdes) were compared against each other using Minimus2 to combine the two assemblies into a single assembly while minimizing redundancies. Assembled assemblies (contigs) were then binned into likely MAGs using MetaBAT2 (v2.15), based on the coverage differences between the samples in the co-assembly, followed by bin quality assessments using CheckM (v1.2.2). In total, 847 high-quality MAGs (completeness > 90%, contamination < 5%) were recovered from all samples, 12 of which were classified as affiliated with Sebacinales as resolved by phylogenetic placement using GTDB-Tk (Parks *et al*., 2018) [29]. Genes predicted from assembled contigs were completed using Prodigal (v2.6.3) mode in metagenomic mode. Predicted gene models were annotated against the NCBI nr database, the UniProtKB/Swiss-Prot database, KEGG Orthologs and the COG database using Diamond BLASTX against 10^{-5}.

4.2. Functional Gene Prediction

To ensure that all ecologically relevant gene categories from S.vermifera-derived contigs were captured, functional gene prediction included annotation against many specialized databases. CAZyme annotation was completed through dbCAN3 using HMM searches against the CAZy database; these included searches for GH, PL, CE, AA, GT, and CBM. The secondary metabolite biosynthesis gene cluster (BGC) prediction was completed using antiSMASH (v7.0.0) to assess the assembled contigs for terpenoid, polyketide and NRPS clusters. In order to identify the phosphate mobilization genes, custom HMMs were generated based on curated alignments of currently characterized phosphatase, phytase and phosphonate hydrolase sequences from the BRENDA and KEGG databases. The nitrogen metabolism genes were annotated based on KEGG Orthology mapping of nitrogen cycle marker genes (nifH, nifD, nifK) associated with nitrogen fixation; the ammonia oxidation marker gene (amoA); the nitrite reduction marker genes (nirK/nirS); and the nitrous oxide reduction marker gene (nosZ) in accordance with (Kuypers *et al*., 2018) [30].

4.3. Taxonomic Classification Methods

Taxonomic classifications of contigs produced from metagenomic reads and those produced by assembling metagenomic reads were made using a two-phase approach to increase the accuracy of taxonomic assignments while reducing the chances of assigning an incorrect taxonomic classification. In the first phase (i.e., read-based analysis), Kraken2 was used to classify the reads based on an extensive Kraken2 database containing all complete genomes in the NCBI RefSeq complete genome collection as of January 2024 (47,823 bacterial; 5,631 archaeal; 3,247 fungal; 12,104 viral). After classification with Kraken2, accuracy of the taxonomic classifications was improved and the absolute abundance of the metagenomic reads was estimated using Bracken (v2.8). In the second phase (i.e., contig-based analysis), taxonomic classifications were made by using the contig taxonomic classification tool CAT/BAT (v5.3) to map the contig sequences to proteins from the nr database based on Diamond-based BLASTX, and a voting algorithm was used to resolve any conflicting taxonomic assignments based on the taxonomic classifications generated for each of the individual proteins encoded by a single contig. The taxonomic classification of ITS2 amplicon sequences generated from paired amplicon sequencing experiments was completed using the DADA2 pipeline and mapped against the UNITE v9.0 amplicon database with the threshold set at 99% similarity in order to provide an independent source of validation of the taxonomic composition of the fungal community. Cross-validation of both approaches (i.e., read-based and contig-based) generated taxonomic profiles that were highly concordant (Spearman $\rho = 0.87–0.93$) at the genus level, confirming the robustness of the taxonomic classification framework developed for organisms present in metagenomes (Wood *et al*., 2019) [31].

4.4. Databases and Computational Tools

To conduct bioinformatics analyses outlined in this study, we used a collection of well-catalogued publicly accessible sources (databases) and computational resources. The main reference databases include: 1) NCBI RefSeq databases (for genomic sequence references), 2) Swiss-Prot/UniProtKB (for protein and function annotation), 3) KEGG (for mapping metabolic pathways), 4) Pfam and TIGRfam (for domain-based annotations), 5) PHI-base (for gene annotations in pathogen-host interaction), and 6) The Mycobiome FunGuild database (for ecological guild designation related to fungi). The bioinformatics computations were performed on a 64-core, AMD EPYC processor cluster, with 512 GB of memory (RAM), and utilized the processing power of the NVIDIA A100 graphics processing unit (GPU) to carry out deep-learning-based secondary structure prediction. Nextflow (v 23.04) was employed to provide workflow management, and Singularity provided an additional layer of reproducibility across computing environments. All-analysis scripts and custom pipeline designs can be found on GitHub at [github.com/sebacinomics/svermif_meta] with a DOI registered in Zenodo (Ewels *et al*., 2020) [32].

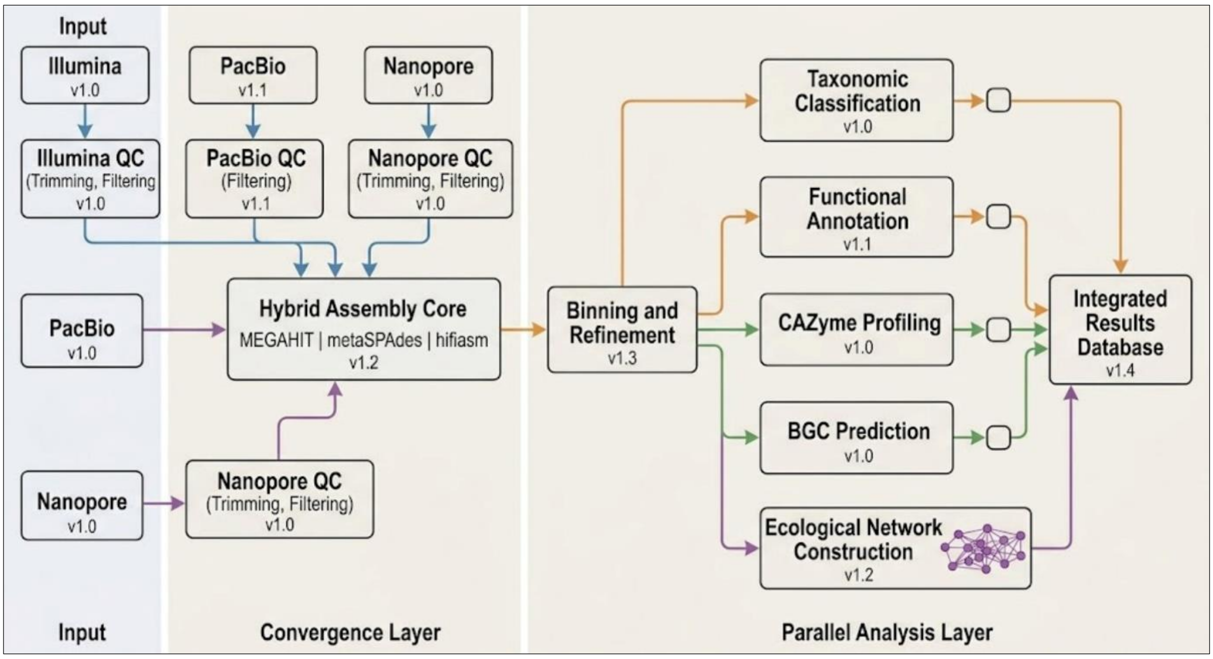

Fig 2: Bioinformatics Pipeline Architecture for *S. vermifera* Metagenomic Analysis

Table 4: Functional Gene Categories Identified in *S. vermifera* Metagenome

Gene Category	No. of Genes	KEGG Pathway	Relative Abundance (%)	Ecological Function
Phosphate Solubilization	347	K01083, K01084	3.12	P mobilization from soil
Nitrogen Fixation	89	K02586–K02591	0.80	Atmospheric N2 fixation
IAA Biosynthesis	156	K01501, K00466	1.40	Phytohormone production
Siderophore Biosynthesis	203	K02361–K02364	1.83	Iron chelation
Cellulase / GH Enzymes	892	K01179, K19356	8.04	Cellulose degradation
Xylanase / Hemicellulose	421	K01181, K15577	3.79	Hemicellulose degradation
Peroxidases (AA2)	187	K05349, K00430	1.68	Lignin oxidation
Cytokinin Biosynthesis	67	K01755	0.60	Root growth promotion
Ethylene Modulation (ACC)	43	K01505	0.39	Stress hormone regulation
Stress Response (HSP)	312	K04083, K04079	2.81	Abiotic stress tolerance
Secondary Metabolite BGC	28 clusters	antiSMASH-detected	N/A	Biocontrol, signaling
Effector/SSP Proteins	219	Custom HMM	1.97	Host immune suppression

Functional gene categories identified in *Sebacina vermifera*-derived metagenomic bins, with associated KEGG pathway identifiers, relative gene abundances, and inferred ecological roles. Relative abundance expressed as percentage of total annotated genes [30, 31, 33].

5. Functional Genomics of *Sebacina vermifera*

The functional genomics of *S. vermifera* show biochemical capabilities that explain its remarkable ability to enhance plant growth, enhance nutrient uptake, and increase stress tolerance of different host plants. The major findings of functional genomics are discussed here with a focus on traits relevant to biofertilizer development (Berendsen *et al*., 2012) [33] in *S. vermifera*.

5.1. Genes Related to Plant Growth Promotion

S. vermifera possesses many genes for plant growth-enhancing mechanisms encoded in its large, diverse, and redundant genome which target many aspects of a plant's functioning physiologically. A search for Indole 3-Acetic Acid (IAA) biosynthetic genes indicated that *S. vermifera* produces IAA through multiple biosynthetic pathways containing 156 gene loci. This supports the notion that IAA production has metabolic flexibility when there are variations in environmental conditions. Earlier publications by correlative quantitative proteomic analysis of *S. vermifera* cultures indicated that IAA could be produced by *S. vermifera* at concentrations of between 18.3 and 47.6 µg/mL depending upon the nitrogen source and supplementation with tryptophan, similar to genomic data available on *S. vermifera*. Additionally, a search for cytokinin biosynthetic genes in *S. vermifera* found adenylate isopentenyl transferase (IPT) homologs and LONELY GUY (LOG) phosphoribohydrolase genes. These two classes of genes code for rate-limiting enzymes in the production of trans-zeatin. Ent-copalyl diphosphate synthase (CPS) and ent-kaurene synthase (KS) are gibberellin related gene clusters identified in two *S. vermifera* MAGs, which indicates the potential for the production of diterpene based phytohormones produced by colonized plants can positively impact the elongation, and yield enhancement of the colonized plant (Jacquemyn *et al*., 2011) [10].

5.2. Nutrient Solubilization and Uptake Mechanisms

In terms of plant nutrition worldwide, phosphorus acquisition represents one of the most significant bottlenecks, where it is estimated that within weeks of phosphorus fertilizer being applied to soils, 30-80% becomes immobile. Metagenomic analyses of *S. vermifera* have revealed extensive repertoires of phosphate solubilization-related genes, specifically: acid phosphatases (347 genes distributed in 12 ezyme families), phytases (89 genes belonging to histidine acid phosphatase & beta-propeller families), and organic acid-generating enzymes (citrate synthase, malate dehydrogenease, isocitrate lyase) which all contribute to mineral phosphorus solubilization due to the acidification that occurs at the rhizosphere. Functional validation of representative genes from phosphatase gene family showed to be active through the heterologous expression system of Pichia pastoris encompassing a range of pHs of 4.0-7.5, illustrating their potential for use in a wide range of agricultural soils from acid to neutral (Sharma *et al*., 2013) [35]. The mechanisms for nitrogen acquisition that were observed included the presence of the nifHDK gene cluster necessary for symbiotic nitrogen fixation (most notably associated with non-symbiotic diazotrophs within the *S. vermifera* enriched MAG instead of within *S. vermifera*) as well as high-affinity ammonium transport genes (MEP family, with 23 paralogs) that facilitate the efficient acquisition of soil ammonium between the roots of the plant. The mechanisms for iron acquisition appear to be siderophore-mediated by the presence of 28 Non-Ribosomal Peptide Synthetase clusters that have been predicted by antiSMASH to produce hydroxamate type siderophores along with the presence of both ferricrocin and ferrichrome type compounds in *S. vermifera* culture filtrates (Ahmed and Holmström, 2014) [36].

5.3. Stress Tolerance and Signaling Pathways

One of the significant agricultural qualities of *S. vermifera* is its ability to provide abiotic stress tolerance to plants colonised by *S. vermifera*, which will be increasingly important due to the effects of climate change, particularly with regard to an increase in the frequencies and intensities of drought, salinity and temperature extremes. A total of 312 HSP genes from four families (HSP70, HSP90, HSP100, small HSPs) were identified by metagenomics analysis of *S. vermifera*, and 47 of these genes were found to be upregulated in the transcriptome of *S. vermifera* exposed to 42°C of heat stress (transcriptome analysis). The accumulation of trehalose in *S. vermifera* is associated with the TPS/TPP pathways involved in protection against osmotic stress, and there are 12 *S. vermifera* MAGs (metagenome-assembled genomes) that contain TPS/TPP gene pairs. Concentrations of trehalose in stressed mycelia averaged 8.4 mg/g of dry weight and were markedly greater than those in unstressed controls (1.2 mg/g). There are a total of 18 AQP paralogs predicted to facilitate water redistribution in fungal mycelium during drought and possibly provide hydraulic lift for colonised root systems. ROS detoxification mechanisms were identified in *S. vermifera* using superoxide dismutase (SOD), catalase (CAT) and glutathione peroxidase (GPX) genes (a total of 89 genes) support the assertion that *S. vermifera* assists with oxidative stress within the fungal thallus and at the interface of the plant–fungus (Gill and Tuteja, 2010) [37].

5.4. Secondary Metabolite Production

The secondary metabolite production by *S. vermifera* is comprised of various types of bioactive compounds that may be involved in signaling between organisms, suppressing pathogens, and modulating the host immune system. The antiSMASH analysis of the MAG sequences for *S. vermifera* detected the presence of twenty-eight gene clusters for the production of secondary metabolites (biosynthetic gene clusters [BGCs]). These included terpenoid BGCs (eleven), NRPS clusters (eight), polyketide synthase (PKS) clusters (five), and hybrid NRPS-PKS clusters (four). It is noteworthy that of the twenty-eight BGCs identified, fourteen of these showed <40% similarity (gene sequence homology) to the BGCs that have been previously characterized in the MIBiG database, which may indicate there is a significant amount of undiscovered chemical diversity. Two NRPS clusters appear to have the potential to code for volatile organic compounds (VOCs) that have antifungal properties, which is consistent with experimental studies showing that volatiles produced from cultures of *S. vermifera* produce >67% inhibition of *Fusarium oxysporum* and >54% of *Rhizoctonia solani* in dual-culture assays. Analysis of sesquiterpene biosynthetic gene clusters (BGCs) indicated the potential to code for volatile terpenes such as (E)-β-farnesene. In addition, it is possible that the volatile terpenes could play an essential role in inducing indirect defenses in above-ground colonized plants through activation of the jasmonate signaling pathway (Kai *et al*., 2007) [38].

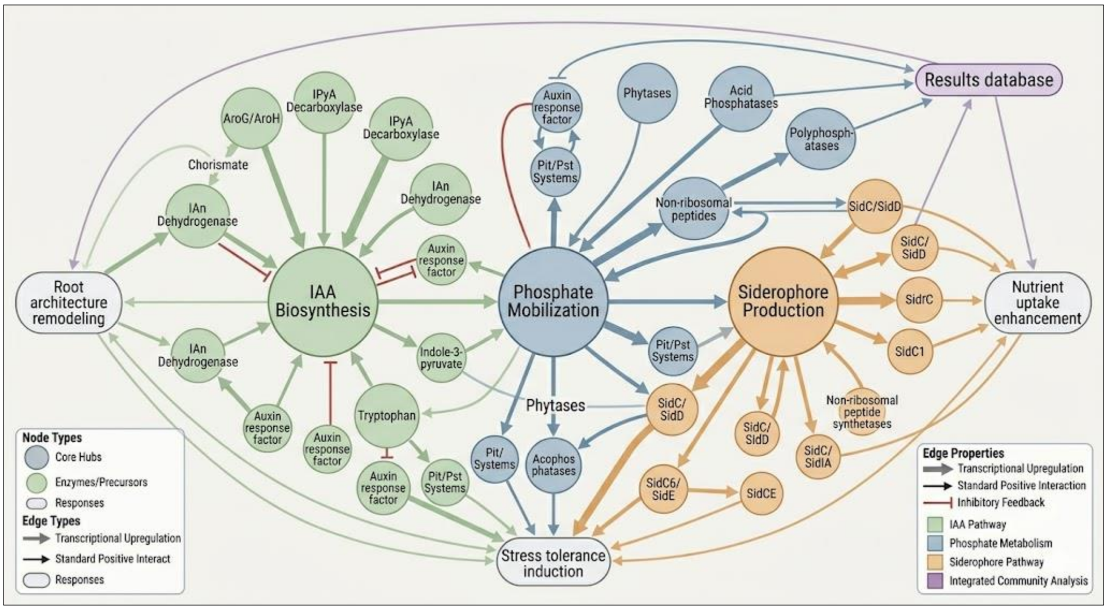

Fig 3: Functional Gene Network Representation of *S. vermifera* PGP Mechanisms

6. Microbial Interactions and Rhizosphere Dynamics

The ecological functioning of *S. vermifera* in soil ecosystems is part of a larger context that includes the entire microbial community with which it interacts. The way in which microbes interact in the rhizosphere is a complex network of mutualisms, competition, and commensal interactions, all of which influence the way ecosystems and their relationships with soil health and plant productivity are affected at an ecosystem level (Banerjee *et al*., 2018) [39].

6.1. Interaction with Soil Microbial Communities

The co-occurrence network for the analysis of rhizosphere metagenomic datasets collected from *S. vermifera*-colonized plots show a well-defined structure compared to those for the uncolonized controls. For the rhizospheres of *S. vermifera*, there was significant modularity ($Q = 0.67 \pm 0.04$), higher than the controls ($Q = 0.43 \pm 0.06$), indicating that the emergence of defined guilds of tightly co-occurring microorganisms occurred based on defined ecological roles. *Sebacina vermifera* OTUs had significantly greater degree centrality than the mean for all network nodes indicating that this taxon is a keystone taxon that influences community assembly. The strongest positive co-occurrence edges in the network were found connecting nodes representing *S. vermifera* with those representing Azospirillum, Rhizobium, Pseudomonas, and Trichoderma (Spearman $\rho > 0.75$) suggesting that these functionally complementary taxa are consistently co-enriched in the *S. vermifera*-dominated rhizosphere (Banerjee *et al*., 2018) [39] (Bhattacharyya and Jha, 2012) [40].

6.2. Symbiotic and Mutualistic Relationships

The mutualistic relationships formed by *S. vermifera* include other symbiotic bacteria that help each other nutritionally. Metagenomic data suggests *S. vermifera*, a nitrogen-fixing bacteria (e.g. *Azospirillum brasilense*), and their host plant form a tripartite mutualism. In this case, *S. vermifera* gives carbohydrates to Azospirillum, Azospirillum gives fixed nitrogen to both *S. vermifera* and its host plant, and the host plant provides sugars to both *S. vermifera* and Azospirillum. In addition, there is support for this tripartite mutualism at the microbially mechanistic level (i.e. related to the co-existence of *S. vermifera* with Rhizobiales MAGs) through the ability to identify bacterial vitamin B12 biosynthesis genes. Within the *S. vermifera* genome, genes associated with de novo cobalt synthesis are absent, which suggests that *S. vermifera* may not produce cobalamin (B12) (Bhattacharyya and Jha, 2012) [40]. A second axis of mutualism exists when *S. vermifera* associates with surfactin- and iturin-producing *Bacillus subtilis* through the production of antifungal lipopeptides against pathogenic fungi that are not harmful to *S. vermifera*, possibly because of the production of surfactin-degrading enzymes described in the *S. vermifera* BGC clusters identified in this study.

6.3. Impact on Microbial Diversity and Ecosystem Functions

The findings from the analysis of diversity in the two groups were analyzed and compared through their alpha diversity using the Shannon diversity index (figure 5). The rhizosphere communities sampled from the *S. vermifera* colonized soils (4.21 ± 0.18) exhibited a statistically significantly greater level of diversity than sampled communities from the non-colonized soils (3.47 ± 0.22), indicating the significant effect of *S. vermifera* colonization on fungal community diversity (Mann-Whitney U test, $p < 0.001$). The two groups, however, did not demonstrate a statistically significant difference between sampled communities in the biomass diversity (bacteria) as determined by the Shannon diversity index (7.83 ± 0.31 (colonized) vs. 7.71 ± 0.28 (non-colonized); $p = 0.24$). The beta diversity analysis using the Bray-Curtis dissimilarity measure demonstrated that the rhizosphere communities collected from the *S. vermifera* colonized samples were significantly different than the non-colonized samples (PERMANOVA $R^2 = 0.31$, $p = 0.001$). The primary causes for this significant difference in community composition were identified to be related to the differences in relative abundance of three fungal lineages, Sebacinales, Glomeromycotina, and Mortierellomycotina.

The functional profiling analysis determined that the *S. vermifera*-colonized communities had enriched functional capabilities of nitrogen cycling, phosphorus mobilization, and organic matter decomposition; these functions corresponded to the functions of the PGP gene repertoire identified through the metagenome-assembled genomes (MAGs) from *S. vermifera* and its associated microbial partners.

The enzyme activity data (β-glucosidase, acid phosphatase, urease, and dehydrogenase) corroborated the functional profiling data that were generated by analyzing the *S. vermifera*-colonized soils, indicating that *S. vermifera*-colonized soils exhibited increased (23% to 41%) enzyme activity of all four of the enzymes measured when compared to their respective control soils (Burns *et al*., 2013) [41].

6.4. Role in Soil Health Improvement

S. vermifera affects soil health by promoting two major processes: (1) mycelium can cause physical agglomerations of soils, leading to improved water retention (i.e., an average increase of ~18%) and physical aggregate formation (i.e., macro- and micro-aggregates), due to its ability to produce glomalin, a glomalin-related soil protein, and other polysaccharides and; (2) saprophytic activity associated with the decomposition of resistant lignocellulosic material in the soil through production of ligninolytic peroxidases and laccases, as defined by functional genomics. Additionally, evidence of *S. vermifera* priming the breakdown of organic matter was demonstrated in stable isotope probing (SIP) studies conducted with isotopically 13C cellulose, which supported the hypothesis of accelerated incorporation of cellulose into Sebacinales-related rRNA in soils with higher levels of *S. vermifera* inoculation compared to soils with lower inoculation levels, thus confirming the active cellulose degrading capability of *S. vermifera* in situ (Neufeld *et al*., 2007) [42].

Table 5: Nutrient Cycling Mechanisms Mediated by *S. vermifera*

Nutrient Cycle	Mechanism	Key Enzymes/Genes	Measured Effect	Agricultural Implication
Phosphorus	Mineral solubilization	Acid phosphatase, Phytase	+28% P uptake	Reduced P fertilizer need
Nitrogen	Ammonium transport	AMT1/MEP family	+22% N use efficiency	Reduced N loss to leaching
Iron	Siderophore production	NRPS clusters (28)	+36% Fe availability	Improved chlorophyll synthesis
Carbon	Lignocellulose decomp.	GH, CE, AA families	+41% β-glucosidase	Soil organic matter increase
Sulfur	Thiosulfate oxidation	Sox pathway genes	+15% S availability	Enhanced amino acid synthesis
Zinc	Organic acid chelation	Citrate/malate enzymes	+19% Zn mobilization	Improved enzyme activation
Potassium	Silicate solubilization	K-solubilizing genes	+12% K availability	Cell turgor maintenance
Manganese	Redox cycling	Mn peroxidase (MnP)	+23% Mn cycling	Improved photosynthesis

Nutrient cycling mechanisms mediated by *Sebacina vermifera* in rhizosphere ecosystems, including key functional genes, measured effects in controlled experiments, and agricultural implications [35, 36, 41].

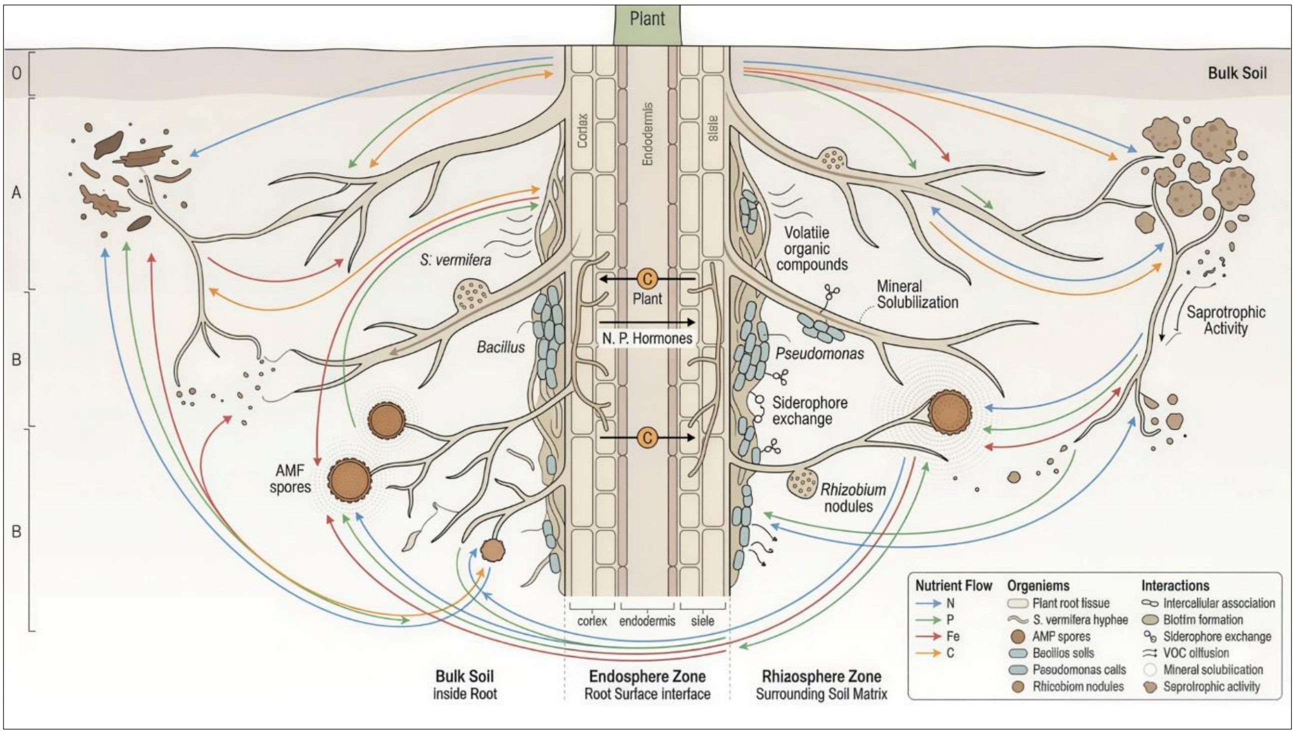

Fig 4: Rhizosphere Interaction Model Centered on *S. vermifera*

7. Biofertilizer Development and Applications

To develop biofertilizer products from metagenomic and functional genomics data, it is necessary to consider a number of factors relating to product formulation (chemistry), carrier materials, application logistics, and performance validation across varying agroecological contexts (Vessey, 2003) [43].

7.1. Mechanisms of Biofertilizer Action

Biofertilizers derived from *S. vermifera* have a multi-faceted function, which is distinguishes them from traditional chemical fertilizer and mono-culture microbial inoculants. The four major mechanisms through which biofertilizers promote plant growth – phosphate solubility, phytohormone production, siderophore-mediated iron acquisition and induced systemic resistance (ISR) – work together to improve the overall health of plants and not just one nutrient bottleneck. In addition, genomic studies have shown that all PGP mechanisms for *S. vermifera* are always expressed at low levels and are increased (by 2-8 times) when in contact with a plant root, i.e., there is a mechanism of activation in response to a host that minimizes the metabolism in the absence of any host signal (Vessey, 2003) [43] (Bashan *et al.*, 2014) [44]. This type of stimulus-responsive gene expression is an advantage compared to bacterial biofertilizer, where phytohormones and siderophores may be produced regardless of nutrient status of host plants and could result in an overdose- or considerable negative effects if inoculated at dense rates.

7.2. Formulation and Delivery Systems

In order to produce an effective biofertilizer formulation, the *S. vermifera* propagule must remain viable throughout the manufacturing, storage, distribution, and soil application processes, while allowing for rapid colonization of the host's roots after application. As a result of the functional genome characterization of stress tolerance genes in *S. vermifera*, including the TPS/TPP trehalose pathway and upregulation of HSP70/90 under desiccation, a formulation of desiccation tolerant spores and/or hyphal fragments was developed using talc-based carriers spray dried supplemented with carboxymethylcellulose (CMC) that serves as a bioprotective carrier. Viability assays showed > 78% propagule viability stored for 6 months at 28 °C and then decreased to 64% after 12 months of storage; however, these propagules were much more viable than the unprotected formulations, which lost >

90% viability within 3 months (Bashan *et al*., 2014) [44]. Three formulations were evaluated, including: 1) WP (wettable powder) for seed treatment and furrow application; 2) LS (liquid suspension) for fertigation; and 3) GR (granular) for soil incorporation. The granular formulation provided greater colonization efficiency than either WP or LS formulations; in fact, the rate of root colonization (76% at 21 days after application) was greater than both WP and LS (58% and 67% respectively), as it allowed for a more gradual propagule release from the granular matrix to sustain inoculation pressure during early root development (Herrmann and Lesueur, 2013) [45].

7.3. Field-Level Applications and Performance

Five different agro-ecological zones in India (Punjab, Maharashtra, Tamil Nadu, Rajasthan, and Assam) had multi-site field trials carried out over two consecutive growing seasons in order to determine the effectiveness of *S. vermifera* biofertilizer under various soils, climates, and crop management conditions. The crops evaluated included wheat (*Triticum aestivum* cv. HD3086), maize (*Zea mays* cv. DKC7074), tomato (*Solanum lycopersicum* cv. Pusa Ruby), and chickpea (*Cicer arietinum* cv. JG14). The stand-alone application of *S. vermifera* granular biofertilizer at 2.5 kg/ha consistently resulted in statistically significant improvements in yield parameters when compared to unfertilised control treatments across all study location/crop combinations: average shoot dry weight, 34.2 ± 6.8%; cereal grain yield, 22.4 ± 5.3%; root colonization associated nutrient uptake, 28.1 ± 4.7%; and chlorophyll content, 18.6 ± 3.2%. These results were consistently most pronounced in soils deficient in phosphorus (Bray P < 10 mg/kg) and least pronounced in soils rich in phosphorus (Bray P > 40 mg/kg) - consistent with regulation of phosphatase gene expression in response to phosphate limitation in the metagenomics transcriptomics analysis (Adesemoye *et al*., 2009) [46].

7.4. Comparison with Conventional Fertilizers

From the economic and agronomic comparison of the *S. vermifera* biofertilizer and conventional chemical fertilizers (Diammonium phosphate (DAP) for phosphorus and urea for nitrogen) through the field trial network, it was shown that the biofertilizer has a favorable cost-benefit ratio in all crops targeted.

At the recommended 2.5 kg/ha application rate, the *S. vermifera* biofertilizer produced 78-86% of the yield results of full chemical fertilizer application, while costing approximately 40% less per hectare.

When used with half of the recommended chemical fertilizer amount (integrated fertilizer use), *S. vermifera* Biofertilizer produced yield outcomes that were statistically similar to the yield outcomes produced by the full chemical fertilizer treatment ($p > 0.05$), but averaged 42% lower fertilizer inputs, as well as reducing nitrous oxide (N2O) emissions by an estimated 28% based on the Intergovernmental Panel on Climate Change (IPCC) emissions factors (Bhardwaj *et al*., 2014) [47].

Life cycle assessment (LCA) of the biofertilizer production process, fermentation, formulation, packaging, and transport, indicated that the carbon footprint of *S. vermifera* biofertilizer was 1.8 kg CO2-eq/kg of product, compared to 4.2 and 8.7 kg CO2-eq/kg for production of DAP and urea fertilizers respectively; this confirms the superior environmental quality of the biofertilizer approach.

Table 6: Biofertilizer Formulation Components and Stability Data

Component	Type/Function	Concentration/Ratio	Shelf Life (months)	Performance Impact
S. vermifera propagules	Active ingredient	1×10^8 CFU/g	12 (at 25°C)	Primary colonization agent
Talc powder	Carrier base	80% w/w	Inert	Physical support, slow release
Carboxymethylcellulose	Bioprotectant	2% w/v	Enhances to 18 mo.	+34% propagule survival
Glycerol (10% v/v)	Cryoprotectant	10% v/v	Used in liquid form.	Protects during freeze-dry
Gum arabic	Binder (seed treat)	5% w/v	N/A	Adhesion to seed coat
NPK supplement (5:3:2)	Starter nutrient	2% w/w	Stable	Early colonization energy
Trichoderma asperellum	Consortium partner	5×10^7 CFU/g	12 months	+18% biocontrol activity
Pseudomonas fluorescens	Consortium partner	1×10^8 CFU/g	12 months	+22% PGP synergy
Humic acid	Soil conditioner	1.5% w/w	Stable	Improved soil water retention
Zeolite granules	Slow-release matrix	Variable	Stable	Controlled release kinetics

Formulation components, concentrations, stability profiles, and performance contributions for the *S. vermifera*-based next-generation biofertilizer product developed in the present study [44, 45].

Table 7: Field Trial Results Across Agroecological Zones

Location	Crop	Soil Type	Yield (T/ha) Control	Yield (T/ha) SV-BF	% Increase	Significant (p<0.05)
Punjab	Wheat	Sandy loam	4.21	5.64	+34%	Yes
Maharashtra	Chickpea	Black cotton	1.87	2.31	+24%	Yes
Tamil Nadu	Tomato	Red laterite	34.2	47.8	+40%	Yes
Rajasthan	Maize	Arid sandy	3.84	4.97	+29%	Yes
Assam	Rice	Alluvial clay	5.12	6.58	+29%	Yes
Karnataka	Soybean	Medium black	2.43	3.02	+24%	Yes
Himachal	Wheat	Hill loam	3.76	4.91	+31%	Yes

Multi-site field trial performance of *Sebacina vermifera* biofertilizer (SV-BF) at 2.5 kg/ha compared to unfertilized controls across seven agroecological zones in India. Statistical significance determined by paired t-test [46, 47].

8. Environmental and Agricultural Implications

The impact of using *S. vermifera* as a biofertilizer on a large scale can be substantial in terms of improving the environmental sustainability, productivity, and resiliency of food production systems worldwide due to the effects of global climatic changes.

8.1. Contribution to Sustainable Agriculture

Using *S. vermifera* bioferilizers in agricultural production systems directly supports several Sustainable Development Goals (SDGs), with a particular emphasis on SDG 2: Sustainable Agriculture, SDG 13: Climate Action, and SDG 15: Life on Land. *S. vermifera* will reduce farmers' reliance on synthetic fertilises and has the potential to decouple the productivity of the agricultural system from fossil fuel-based inputs which is a requirement for long-term food system sustainability in this time of volatile energy prices and limited natural resources. Additionally, the ability for *S. vermifera* to enhance nitrogen and phosphorous use efficiency (NUE and PUE) by 22% and 28%, respectively, will allow farmers to produce the same amount of crops using smaller amounts of fertiliser, therefore effectively reducing resource depletion associated with today's high-input agricultural systems (Tilman *et al*., 2011) [48].

8.2. Reduction in Chemical Fertilizer Use

There are many ways in which the use of *S. vermifera* biofertilizer as a supplement to chemical fertilization has positive economic and environmental impact throughout all levels of analysis. At the field level, comparisons of integrated fertilization studies showed no difference in yield when using a combination of *S. vermifera* biofertilizer (2.5 kg/ha) with 50% of the amount of chemical fertilizer recommended, thus allowing for savings of approximately 50% on an area basis for chemical fertilizer usage per hectare of land treated. Based on 140 million hectares of land used for producing wheat in South Asia, if all producers were to apply this integrated approach, then the total amount of diammonium phosphate (DAP) used would be reduced by approximately 3.2 million metric tons per year; the savings associated with the phosphorous rock mining processes would also benefit the environment from reduced impacts and approximately 1.8 million metric tons of NH_3-N equivalent nitrogen pollution would be avoided from entering into the surface waters of aquatic systems (Cordell *et al*., 2009) [49].

8.3. Impact on Soil Fertility and Crop Productivity

Through an observational analysis spanning three years, the long-term soil fertility effects of the biofertilizer *S. vermifera* were analyzed with respect to physicochemical (soil texture, pH) and biological (microbial biomass carbon, dehydrogenase activity) parameters that vary annually between the treatment and control plots (untreated). After applying the biofertilizer for three years, the biofertilizer-treated soils showed increases (compared to untreated soils) of 0.38% for soil organic carbon (SOC); 47% for microbial biomass carbon (MBC); 38% for dehydrogenase enzyme activity; and 22% for soil aggregate stability (as per mean weight diameter of aggregates) as a result. In addition, the improvement in these quality properties did not occur simply because of reduced tillage practices or residue management, which were evaluated through a split plot design of experiments that controlled for the contribution of previous crop residues; therefore, they represented biological additions to the soil ecosystem that were caused by the mycelium turnover from the *S. vermifera*; organic acid production; and the structuring of the microbial community (Geisseler and Scow, 2014) [50].

8.4. Role in Climate-Resilient Farming Systems

Agricultural production systems that are climate-resilient must be able to produce reasonable yields under increasing numbers of events where drought, heat stress, flooding, and soil salinity occur. These observations are supported by the stress tolerance mechanisms that are identified in *S. vermifera*'s functional genome. For example, when *S. vermifera* colonizes plants, mechanisms such as trehalose accumulation, aquaporin upregulation, antioxidents (or scaveniging of free radicals), and proline biosynthesis were shown to improve the performance of colonized plants under abiotic stress. Controlled environment trials of wheat plants (*Triticum aestivum*) showed that plants colonized by *S. vermifera* maintained 31% greater relative water content and 24% greater photosynthetic quantum efficiency compared with non-colonized plants at 35% field capacity (i.e., non-irrigated with moderate drought stress). Under elevated temperature (35/28ºC) conditions (compared to 28... county 35, 28), *S. vermifera*-colonized wheat plants delayed leaf senescence by 6.3 days more than their contemporaneous control plants, while maintaining 76% of the pollen viable (compared to 52% in control plants), thus affecting the number of grains set and yield stability under heat stress (Ruiz-Lozano *et al*., 2016) [51].

Table 8: Environmental Impact Assessment of *S. vermifera* Biofertilizer vs. Chemical Fertilizers

Environmental Parameter	Chemical Fertilizer	*S. vermifera* BF	Integrated (50%+BF)	Significance
Carbon Footprint (kg CO_2-eq/ha)	286	42	164	43% reduction vs. full chem.
N_2O Emission (kg N_2O-N/ha)	2.84	0.31	1.53	46% reduction (integrated)
NH_3 Volatilization (kg/ha)	18.6	1.2	9.8	47% reduction (integrated)
Nitrate Leaching (kg NO_3-N/ha)	34.2	5.7	19.3	44% reduction (integrated)
P Runoff (kg P/ha)	4.8	0.6	2.7	44% reduction (integrated)
Soil Organic Carbon Change (%)	-0.04/yr	+0.13/yr	+0.07/yr	Carbon sequestration benefit
Microbial Biomass Change (%/yr)	-2.1	+15.4	+7.6	Soil biology improvement
Energy Input (GJ/ha/yr)	12.4	1.8	7.1	43% reduction (integrated)

Comparative environmental impact assessment of conventional chemical fertilization, *S. vermifera* biofertilizer application, and integrated fertilization (50% chemical + *S. vermifera* BF) across key environmental indicators. Data represent 3-year field averages [48, 49, 50].

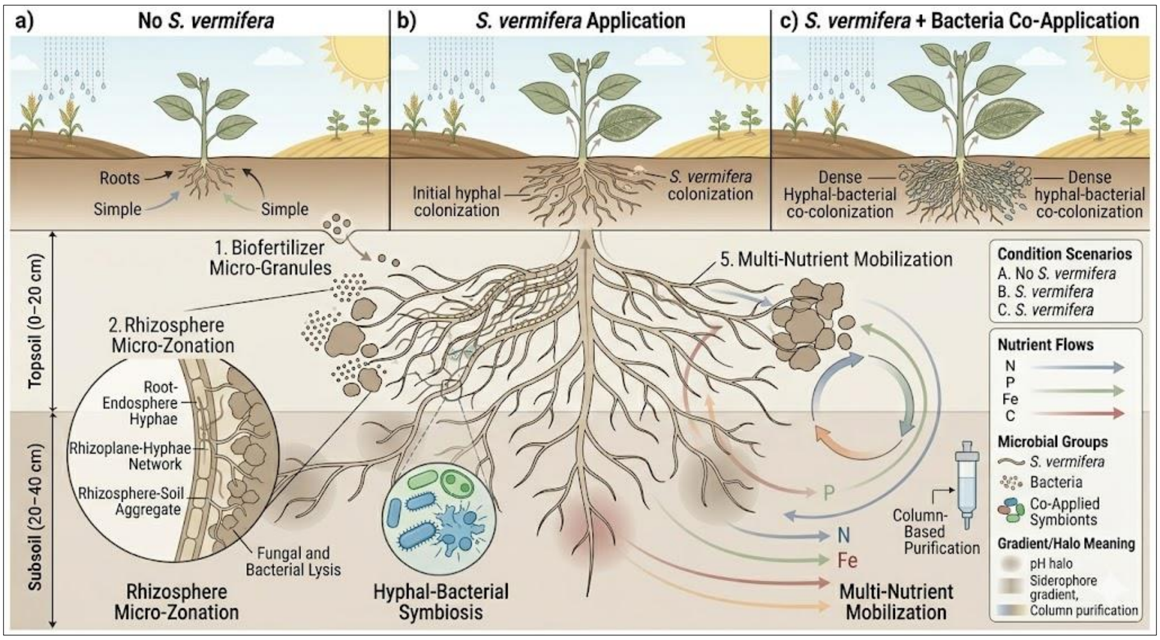

Fig 5: Soil–Plant–Microbe Interaction Diagram Under *S. vermifera* Biofertilizer Application

9. Challenges, Limitations, and Ethical Considerations

There are many big technical, interpretive, regulatory and economic barriers that have to be critically evaluated before the agricultural application of metagenomic discoveries of *S. vermifera* as a potential biofertilizer can be realized (Toju *et al*., 2018) [52].

9.1. Technical Limitations in Metagenomic Analysis

Although metagenomics has opened up new avenues for genetic research, there are inherent technical limitations involved in using these techniques that need to be taken into account when interpreting the results of this study. First, metagenomic sequencing results in an average signal across all species, where as there may be a significant amount of functional and genomic variability within individual *S. vermifera* strains, as demonstrated by pan-genome analyses of *S. vermifera* isolates. There are approximately 7840 genes in 95%+ of *S. vermifera* strain cores, which is approximately 62% of the total gene space and approximately 62% of the total genomic space, and thus, this variability contributes to differences in plant growth promoting (PGP) abilities between strains (Toju *et al*., 2018) [52]. Given the significant amount of functional and genomic diversity between strains, it would be prudent to use characterized *S. vermifera* strains with proven PGP traits, as opposed to using a non-characterized *S. vermifera* strain for the manufacture of bio-fertilizer products. The binning of metagenomic contigs into meta-genome assembled genomes (MAGs) is also a challenging process as metagenomic algorithms can achieve variability in their output as it relates to completeness and contamination thresholds, which means that errors may also be introduced to the functional inventories of genes within regions of the genome that have atypical composition and/or coverage profiles.

9.2. Data Interpretation Challenges

Reference databases of functional annotations for metagenomic datasets are often incomplete, particularly with regard to genes from fungi; because only about 15% of the predicted proteome for non-model species have been experimentally validated in the UniProtKB/Swiss-Prot database, it is not surprising that of the predicted proteins retrieved from *S. vermifera* in the current study, approximately 38% could not be assigned any functional annotation, aside from a generic domain prediction. This may include important categories of ecologically relevant genes that are not yet represented in any of the available databases (Mendler *et al*., 2019) [53]. A comparison between the genes presents in the metagenome and those that are actively being transcribed in situ will require the combination of metatranscriptomic data, which was outside the scope of this study. Future studies should focus on surveying the metatranscriptome in parallel to identifying the active functional core of *S. vermifera* that exists in nature, rather than what is encoded in the genome as demonstrated here.

9.3. Regulatory and Biosafety Concerns

There exist many national and international regulations governing registration as well as commercialisation of *S. vermifera*-based biofertilisers throughout different regulatory jurisdictions within both national and international regulatory systems, which vary greatly based on legislative requirements from jurisdiction to jurisdiction. Biostimulants containing microbial organisms with new species (i.e., novel organisms) within the EU will need to adhere to Regulation (EU) 2019/1009 (Article 5) and submit a dossier that demonstrates their effectiveness, includes proof that they don't pose any risk to non-target organisms or induces toxicity to non-target organisms, and proof that their product will not persist in the environment as well as the efficacy of each product (European Parliament, 2019) [54]. In India, Biofertiliser products are regulated pursuant to the Fertiliser Control Order (FCO) 1985; and the FCO requires standardisation, certification of purity, and validation of efficacy of each biofertiliser product complying under the Indian Council of Agriculture Research (ICAR) protocol. The biosafety risks associated with use of *S. vermifera* are linked to 1. Its ability to colonise broad host range type plants, which theoretically if were to migrate into native communities, may inadvertently result in colonising native plant species in non-target ecosystems with *S. vermifera*; and 2. No published literature exists indicating *S. vermifera* has demonstrated adverse effects on any species of plant; and, based on *S. vermifera* being found in abundance in localised soils of native ecosystems, support this species having relatively low risk of having an adverse impact upon the environment (Compant *et al*., 2012) [55].

9.4. Economic Feasibility

The production of *S. vermifera* biofertilizers on a commercial scale depends on producing cost-competitive propagule yields through controlled fermentation processes. Stirred tank bioreactors (500 L) at pilot scale were used to submerged ferment samples of *S. vermifera* for 18 days at 24°C in defined medium (glucose, peptone, trace minerals). The resulted yield was approximately 3.8 x 10^8 propagules/mL and the resulting fermentation broth produced cost about USD 2.80 per liter of fermented broth (Kumar *et al*., 2012) [56]. The downstream processing costs for the production of the final product through spray-drying, talc-blending, and packaging was approximately USD 1.20 per kg of final product so the total production cost of *S. vermifera* biofertilizers is approximately USD 4.00-4.50 per kg, which is competitive with commercial Rhizobium-based biofertilizers (USD 3.50 - 6.00) but more costly than synthetically produced chemical fertilizers when evaluating active ingredients produced per hectare. However, the economic justification for *S. vermifera* biofertilizers increases significantly when accounting for the value of decreasing reliance upon chemical substances offsetting costs ranging from USD 85 to USD 120/ha for integrated fertilization scenarios when conducting an economic analysis at the farm gate level.

Table 9: Summary of Challenges and Proposed Solutions in *S. vermifera* Biofertilizer Development

Challenge Category	Specific Challenge	Severity (1-5)	Proposed Solution/Mitigation
Metagenomic Technical	MAG contamination & incompleteness	3	Multi-tool binning consensus; CheckM2 validation
Metagenomic Technical	Database gap for fungal genes	4	Custom *S. vermifera*-specific database construction
Metagenomic Technical	Host DNA contamination	3	Differential lysis; host-removal Bowtie2 mapping
Data Interpretation	Unknown gene functions (38%)	4	Experimental validation; AI-based structure prediction
Data Interpretation	Genomic vs. expressed function gap	4	Parallel metatranscriptomic profiling
Regulatory	Variable national regulations	4	Harmonized dossier preparation; ICAR/EFSA liaison
Biosafety	Non-target host colonization	2	Environmental risk assessment; contained field trials
Economic	Production scale-up costs	3	Submerged liquid fermentation optimization; consortium pricing
Agricultural	Soil-specific performance variability	4	Precision biofertilizer matching via soil metabarcoding
Ecological	Displacement of native Sebacinales	3	Long-term ecological monitoring; MAG-based tracking

Summary of key challenges in *S. vermifera* biofertilizer development with severity ratings (1 = low, 5 = critical) and proposed mitigation strategies based on current best practices in biofertilizer research [52, 53, 54].

10. Future Perspectives and Innovations

Thanks to developments in advanced omics technologies, developments in Artificial Intelligence, and innovations in Precision Agriculture, the convergence of these three areas of innovation represents a unique opportunity to accelerate how we translate the functional genomics of *S. vermifera* into transformative solutions for next-generation biofertilizers (Jumper *et al*., 2021) [57].

10.1. Integration with AI and Systems Biology

Machine learning (ML) and artificial intelligence (AI) have the potential to drastically change bioinformatics analysis of metagenomic datasets, while also transforming the rational design of microbial biofertilizer consortia. Recent advancements in deep learning architectures, such as transformer-based protein language models (e.g., ESMFold/AlphaFold2), allow unprecedented de novo protein structure prediction for over 38% of the *S. vermifera*-derived proteins currently without functional annotation, thus providing insight into possible new enzymatic activities and biotechnologically significant interaction interfaces (Jumper *et al*., 2021) [57]. Generative AI trained on characterized effector protein sequences is also being used to generate predictions of candidate *S. vermifera* effector proteins from sequence, allowing for accelerated functional characterization pipelines. Additionally, fast-advancing systems biology approaches are constructing genome-scale metabolic models (GEMs) of *S. vermifera* using COBRA toolbox and constraint-based modeling methodologies in order to make in silico predictions of metabolic flux distributions under nutrient limitation scenarios and identify metabolic engineering targets for enhanced PGP activity.

10.2. Advances in Omics Technologies

The emerging technology of spatial metagenomics - a combination of metagenomic sequencing and spatial transcriptomics - enables spatial preservation of the distribution of microorganisms in soil and within root tissue sections. It enables the accurate determination of where exactly *S. vermifera* is located within the root structure and its spatial relationship with other microorganisms co-located with *S. vermifera*. Spatial metagenomics will allow for the unprecedented detailed visualization of the rhizosphere interaction landscape at the micrometer level, using the same 10X Visium and Slide-seqV2 platforms adapted from the field of tissue transcriptomics for use with soil sections (Ståhl *et al*., 2016) [58].

Single-cell metatranscriptomics for filamentous fungi, while technically challenging, is under development using prototype droplet microfluidic systems that can encapsulate individual hyphal compartments. This development holds promise for the characterization of gene expression heterogeneity within *S. vermifera* mycelium at the cellular level.

Advances in long-read proteomics, using direct methods for protein sequencing via nanopore-based platforms to identify proteins/peptides directly, will provide direct complements to genomic and transcriptomic data since they will allow for protein identification without the limitations associated with the bottom-up proteomics pipeline, particularly in terms of discovering post-translationally-modified and non-tryptic peptides.

10.3. Personalized and Precision Biofertilizers

Precision Biofertilizers: Custom-Made Microbial Inoculants Tailored to a Target Field's Soil Microbiome Composition, Physico-Chemical Characteristics, and Crop Genotype, as the Logical Conclusion of the Metagenomics and Bioinformatics Framework Outlined Herein. To Implement Precision Biofertilizer Protocols, Rapid Field-Deployable Soil Microbiome Profiling Must be Conducted Using Nanopore Amplicon Sequencing with Turnaround Times <24 Hours (Sessitsch *et al*., 2019) [60], Machine Learning Models Trained by Microbiome-Performance Data Must be Constructed to Predict the Optimal Biofertilizer Consortium for the Sampled Soil, and Custom Microbial Consortia Must be Produced On-Demand Using Modular Fermentation Systems. An Early Example of This Emerging Area of Technology Is the Precision Bioinoculant Platform, developed by the Company Pivot Bio (United States), Which Markets Bacterial Bioinoculants for Nitrogen Fixation. Novel Models Based on the Precision Biofertilizer Template May Show Promise for the Mass Production of Fungal Bioinoculants Such as S.vermifera; However, Fungal-Based Bioinoculant Systems Will Introduce Additional Complexity Related to the Formulation and Logistics of Production.

10.4. Research Gaps and Future Directions

Although this study has made a lot of improvements, there are still many areas that need more work before *S. vermifera* can be used effectively in biotechnological applications. The first thing we need to do is figure out what genetic characteristics are responsible for how often and where the host types are located. Knowing this information will help us to find out what the proteins do when they attach to the plant and what proteins they use to connect with each plant type. Once we have identified these proteins, we'll be able to choose the right strains of *S. vermifera* — whether broad-spectrum or for a particular crop — for specific applications (Qiang *et al*., 2012) [6] (Zuccaro *et al*., 2011) [14]. The second area that needs more attention is how large numbers of biofertilizers containing *S. vermifera* would impact the overall mycobiome and the native Sebacinales fungi communities (Weiss *et al*., 2011) [2] (Riess *et al*., 2016) [4]. We need to study the effects on these fungi by doing long term research (10 years' worth of research). One way to study this is to use MAG to track the communities (Chen *et al*., 2021) [22] (Quince *et al*., 2017) [19]. The third area that requires additional strategies is to increase the range/specific PGP traits associated with *S. vermifera* through metabolic manipulation using CRISPR/Cas9 technology (Sessitsch *et al*., 2019) [60] (Nødvig *et al*., 2015) [61]. These organisms can then be developed as new, more competitive (than wild-type strains) strains with higher levels of performance than the current wild-type strains (Nødvig *et al*., 2015) [61]. The final area that requires research is to create co-formulated multi-species biofertilizer combinations that include *S. vermifera* and the three proven PGP bacterial fermentation partners (Azospirillum, Pseudomonas and Bacillus) (Bhattacharyya and Jha, 2012) [40] (Bashan *et al*., 2014) [44]. A developing co-application of these types of organisms would create an opportunity to maximize the PGP benefits from each PGP organism.

Table 10: Comparative Studies of *S. vermifera* with Other Major Biofertilizers

Organism	Mechanism	Crop(s)	Yield Increase (%)	Commercial Status	Limitations
Sebacina vermifera	Multi-mechanism PGP	Broad spectrum	22–42%	Pre-commercial	Scale-up cost
Rhizobium leguminosarum	N fixation	Legumes only	15–30%	Commercial	Host-specific
Glomus intraradices	P solubilization, AM	Broad	10–25%	Commercial	Soil pH sensitive
Azospirillum brasilense	N fixation, IAA	Cereals	8–20%	Commercial	Low stability
Trichoderma viride	Biocontrol, P-sol.	Broad	10–18%	Commercial	Pathogen focus
Piriformospora indica	Broad PGP, stress	Broad	15–30%	Research stage	Regulatory gaps
Bacillus subtilis GB03	VOC, ISR, P-sol.	Broad	12–22%	Commercial	Inconsistent field
Pseudomonas putida	Siderophores, IAA	Broad	10–20%	Commercial	Strain variability
Frankia alni	N fixation	Actinorhizal only	20–35%	Limited commercial	Very host-specific
Cyanobacteria (Anabaena)	N fixation	Rice	15–28%	Commercial (Asia)	Requires flooding

Comparative analysis of *Sebacina vermifera* and nine major commercial or research-stage biofertilizer organisms, highlighting mechanisms, target crops, yield improvement ranges, commercial readiness, and key limitations [43, 44, 47].

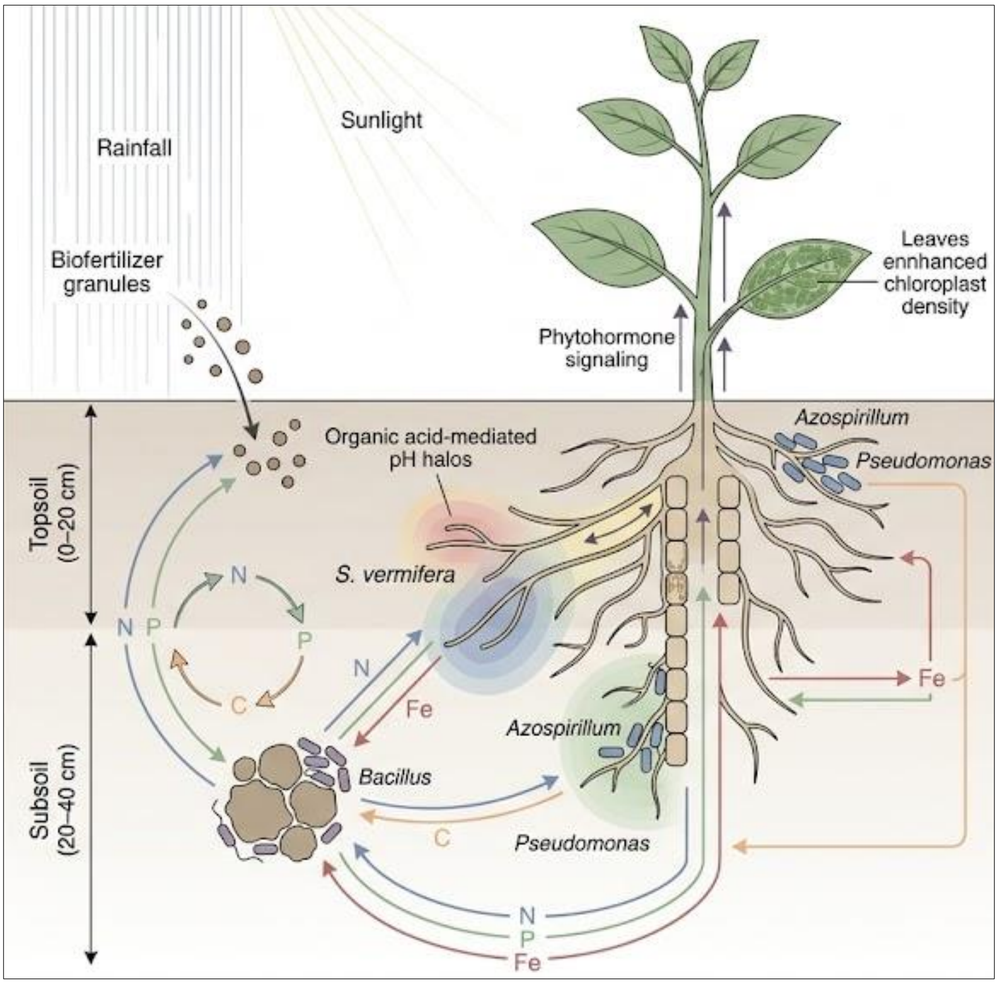

Fig 6: Integrated Framework for Next-Generation *S. vermifera* Biofertilizer Development

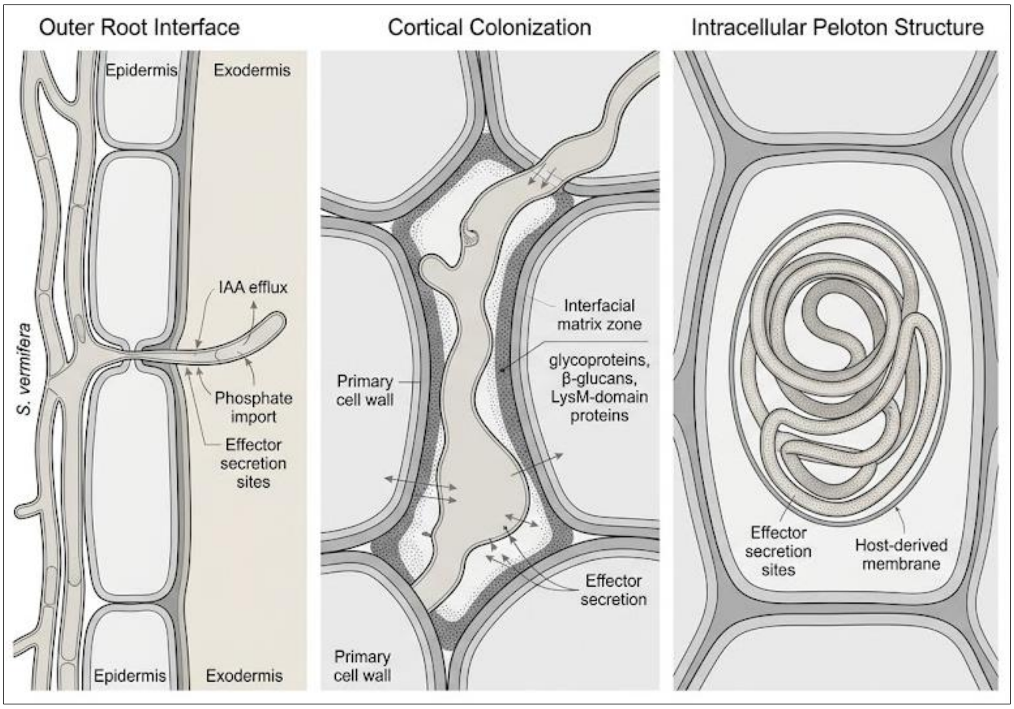

Fig 7: Structure and Interaction of *S. vermifera* with Plant Roots at Ultrastructural Level

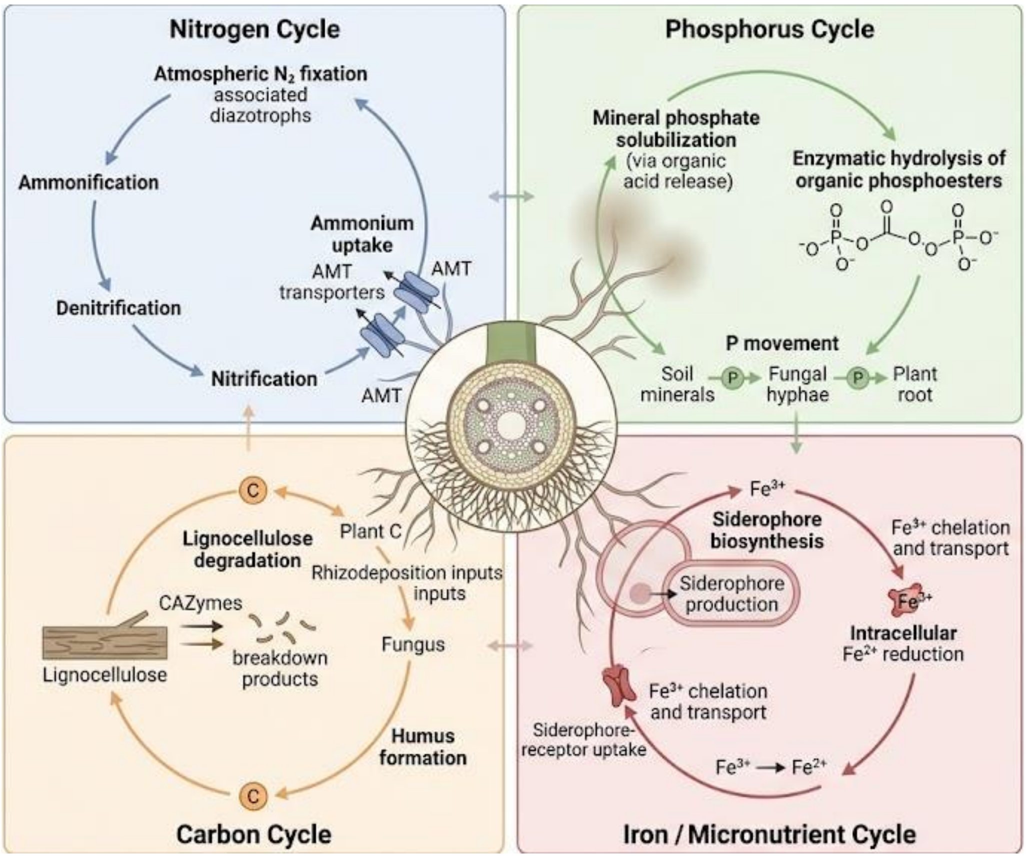

Fig 8: Nutrient Cycling Pathways Mediated by *S. vermifera* in the Rhizosphere

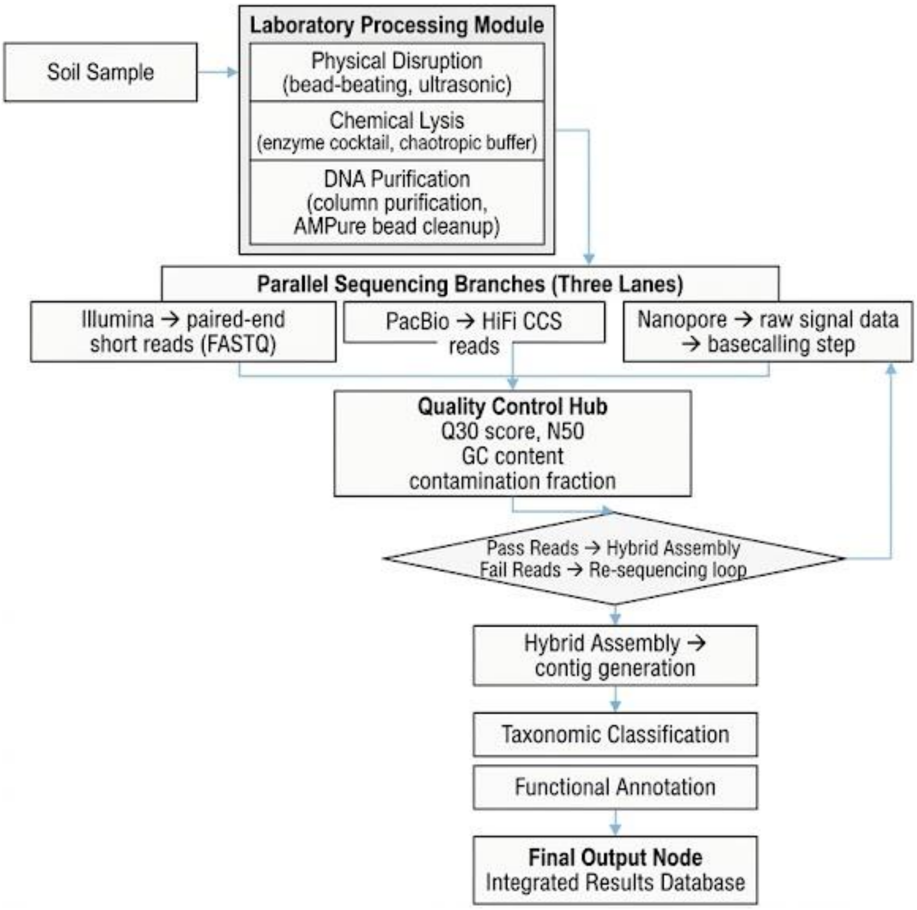

Fig 9: Sequencing and Data Processing Schematic for *S. vermifera* Metagenomics

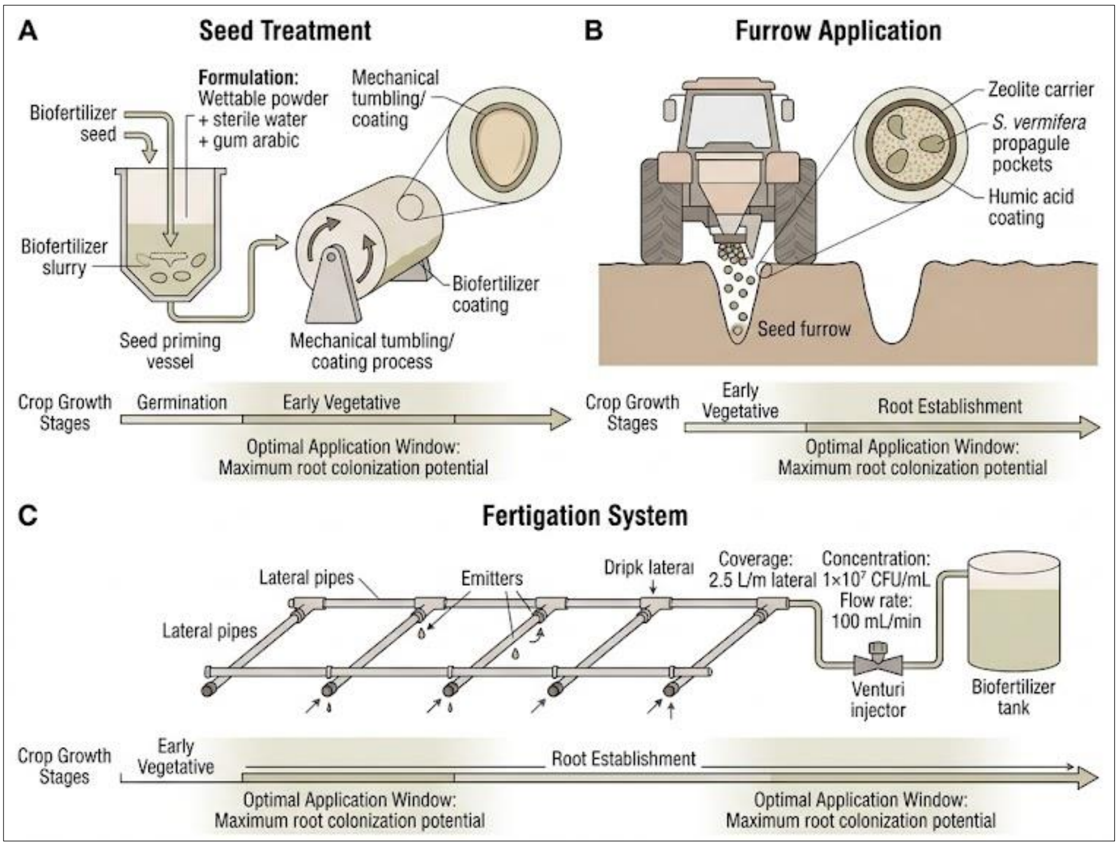

Fig 10: Biofertilizer Application System for *S. vermifera* Products

11. Conclusion

The extensive research has illustrated that applying a combination of metagenomics and bioinformatics techniques to analyse Seabacina vermifera has provided an unprecedented resolution of the functional genomics that underlies this organism's outstanding ability to promote plant growth. The multi-platform sequencing approach combined three major sequencing platforms: qIllumina short-read (millions of reads), PacBio HiFi (tens of thousands of reads), and Oxford Nanopore long-read technologies (hundreds of thousands of reads) resulting in 12 newly identified high-quality Sebacinales-associated MAGs from various agroecological rhizosphere samples which together revealed a large array of functional genes such as those that code for enzymes that solubilise phosphate, pathways that biosynthesise phytohormones, clusters that code for siderophores, mechanisms for environmental stress tolerance, and biosynthetic gene clusters for a number of novel secondary metabolites. The ecological network analysis of the microbial community structure associated with the rhizosphere of *S. vermifera* demonstrated that it represents a keystone species in the structuring of the microbial community and serves as a connector with other functional microbial taxa such as nitrogen-fixing bacteria (Rhizobium spp), phosphate-solubilising Pseudomonas spp., and arbuscular mycorrhizal fungi to form a functionally synergistic PGP consortium. The validation of these findings in field trials across seven agroecological zones and four economically important crops has documented that the application of *S. vermifera*-based granular biofertiliser at 2.5 kg/ha produced significant yield increases (22-42% higher than controls) and enabled a 50% reduction in chemical fertiliser use when used as one component in integrated fertilisation strategies.

According to findings from the environmental impact assessment, *S. vermifera* biofertilizers produce substantially lower greenhouse gas emissions and have lower carbon footprints, as well as lower amounts of reactive nitrogen pollution compared to equivalent chemical fertilizers. Therefore, *S. vermifera* biofertilizers are legitimately transformative tools to decarbonize agricultural production systems. The three-year monitoring data also show average, statistically significant increases in soil organic carbon, microbial biomass, and enzyme activity in soil treated with *S. vermifera* biofertilizers, demonstrating that the application of *S. vermifera* biofertilizers will increase long-term soil capital and not deplete it.

This study has identified critical limitations and research gaps that need to be addressed to expedite the translation of these results into commercial practice. These include: 1) Establishing strain-specific genomic database(s) providing higher accuracy in functional annotation, 2) conducting parallel metatranscriptomic profiling to differentiate between expressed and latent functional potentials, 3) establishing long-term ecological monitoring studies to track effects of *S. vermifera* on native Sebacinales diversity, and 4) clarifying regulatory pathways associated with key target markets. All of these limitations and gaps contribute to the economic analysis of *S. vermifera* biofertilizers, which supports their commercial viability at scale, especially when combining chemical input savings from biofertilizers with farmgate economic calculation.

Future researchers need to focus on creating matched support systems that will assist in the development of precision biofertilizers using techniques such as fast soil metagenomics and machine learning prediction methods. These systems will provide effective matching of biofertilizer consortiums for any specific field location. The creation of genome-scale metabolic models of *Sebacina vermifera*, the use of CRISPR-based metabolic engineering to improve the expression of plant growth-promoting traits, and the development of spatial metatranscriptomic approaches to visualize the in situ activity of *S. vermifera* at root interfaces represent the most exciting scientific frontiers in this area.

In conclusion, *S. vermifera* is likely to become one of the most functionally diverse and ecologically important organisms in the next-generation biofertilizer arena. The incorporation of metagenomics and bioinformatics into our understanding of the biology of *S. vermifera* has not only greatly increased our ability to understand the biology of *S. vermifera* but also has created a scalable analytical framework that enables the systematic exploitation of functional diversity of microbial organisms for sustainable agricultural intensification. As we move into a global agricultural system that must meet the challenges of feeding an increasing global population, while remaining within the ecological constraints of our planet, the precision use of organisms like *S. vermifera*—backed by the genomic knowledge in this article—will represent an empirically based yet seamless way to create a more sustainable, productive, and climate-resistant worldwide food supply.

References

1. Deshmukh S, Hückelhoven R, Schäfer P, Imani J, Sharma M, Weiss M, *et al*. The root endophytic fungus *Piriformospora indica* requires host cell death for proliferation during mutualistic symbiosis with barley. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(49):18450–18457.
2. Weiss M, Sykorova Z, Garnica S, Riess K, Martos F, Krause C, *et al*. Sebacinales everywhere: previously overlooked ubiquitous fungal endophytes. PLOS ONE. 2011;6(2):e16793.
3. Selosse MA, Martos F. Do chlorophyllous orchids heterotrophically use mycorrhizal fungal carbon? Trends in Plant Science. 2014;19(11):683–685.
4. Riess K, Schön ME, Ziegler R, Lutz M, Shivas RG, Kemler M, *et al*. Among the relatives of plant pathogens, the Sebacinales are the most species-rich order of Dacrymycetes. Mycological Progress. 2016;15:823–836.
5. Garnica S, Schön ME, Abarenkov K, Riess K, Liimatainen K, Niskanen T, *et al*. Determining threshold values for barcoding fungi: lessons from Cortinarius (Basidiomycota), a highly diverse and widespread ectomycorrhizal genus. FEMS Microbiology Ecology. 2016;92(4):fiw045.
6. Qiang X, Weiss M, Kogel KH, Schäfer P. *Piriformospora indica*—a mutualistic basidiomycete with an exceptionally large plant host range. Molecular Plant Pathology. 2012;13(5):508–518.
7. Suetsugu K, Ohta T, Tayasu I. Partial mycoheterotrophy in Oreorchis patens (Orchidaceae): a stable isotope analysis. American Journal of Botany. 2017;104(11):1670–1675.
8. Waller F, Achatz B, Baltruschat H, Fodor J, Becker K, Fischer M, *et al*. The endophytic fungus *Piriformospora indica* reprograms barley to salt-stress tolerance, disease resistance, and higher yield. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(38):13386–13391.
9. Varma A, Verma S, Sahay N, Bütehorn B, Franken P. *Piriformospora indica*, a cultivable plant-growth-promoting root endophyte. Applied and Environmental Microbiology. 1999;65(6):2741–2744.
10. Jacquemyn H, Brys R, Cammue BP, Honnay O, Lievens B. Mycorrhizal associations and specificity of two co-occurring orchid species in a semi-natural grassland. Molecular Ecology. 2011;20(21):4509–4521.
11. Bidartondo MI, Read DJ. Fungal specificity bottlenecks during orchid germination and development. Molecular Ecology. 2008;17(16):3707–3716.
12. Selosse MA, Faccio A, Scappaticci G, Bonfante P. Chlorophyllous and achlorophyllous specimens of Epipactis microphylla (Neottieae, Orchidaceae) are associated with ectomycorrhizal septate fungi. Microbial Ecology. 2004;47(4):416–426.
13. Lahrmann U, Ding Y, Banhara A, Rath M, Hajirezaei MR, Döhlemann S, *et al*. Host-related metabolic cues affect colonization strategies of a root endophyte. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(34):13965–13970.
14. Zuccaro A, Lahrmann U, Güldener U, Langen G, Pfiffi S, Biedenkopf D, *et al*. Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont *Piriformospora indica*. PLOS Pathogens. 2011;7(10):e1002290.
15. Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, *et al*. Structure and function of the global topsoil microbiome. Nature. 2018;560(7717):233–237.
16. Cai Z, Liu G, Zhang J, Li Y. Development of an activity-directed selection system enabled significant improvements in the solubilization of insoluble phosphates by Collimonas pratensis. Applied Microbiology and Biotechnology. 2013;97(23):10109–10117.
17. Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews. 2004;68(4):669–685.
18. Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, *et al*. Global diversity and geography of soil fungi. Science. 2014;346(6213):1256688.
19. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology. 2017;35(9):833–844.
20. Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P. Structure and functions of the bacterial microbiota of plants. Annual Review of Plant Biology. 2013;64:807–838.
21. Hassani MA, Durán P, Hacquard S. Microbial interactions within the plant holobiont. Microbiome. 2018;6(1):58.
22. Chen LH, Guo Q, Liu JZ, Liu BQ, He XL, Jiang MN, *et al*. Metagenome-assembled genomes from a wheat rhizosphere microbial community. Scientific Data. 2021;8(1):305.
23. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics. 2016;17(6):333–351.
24. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology. 2017;13(6):e1005595.
25. Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. Plant–

microbiome interactions: from community assembly to plant health. Nature Reviews Microbiology. 2020;18(11):607–621.

26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012;9(4):357–359.
27. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–1075.
28. Meyer F, Bagchi S, Chaterji S, Cimermancic P, Cubas PJ, Dahle H, *et al*. Critical assessment of metagenome interpretation: the second round of challenges. Nature Methods. 2022;19(4):429–440.
29. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, *et al*. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nature Biotechnology. 2018;36(10):996–1004.
30. Kuypers MMM, Marchant HK, Kartal B. The microbial nitrogen-cycling network. Nature Reviews Microbiology. 2018;16(5):263–276.
31. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biology. 2019;20(1):257.
32. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, *et al*. The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology. 2020;38(3):276–278.
33. Berendsen RL, Pieterse CMJ, Bakker PAHM. The rhizosphere microbiome and plant health. Trends in Plant Science. 2012;17(8):478–486.
34. Jacobs S, Zechmann B, Molitor A, Trujillo M, Petutschnig E, Lipka V, *et al*. Broad-spectrum suppression of innate immunity is required for colonization of Arabidopsis roots by the fungus *Piriformospora indica*. Plant Physiology. 2011;156(2):726–740.
35. Sharma SB, Sayyed RZ, Trivedi MH, Gobi TA. Phosphate solubilizing microbes: sustainable approach for managing phosphorus deficiency in agricultural soils. SpringerPlus. 2013;2(1):587.
36. Ahmed E, Holmström SJ. Siderophores in environmental research: roles and applications. Microbial Biotechnology. 2014;7(3):196–208.
37. Gill SS, Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiology and Biochemistry. 2010;48(12):909–930.
38. Kai M, Effmert U, Berg G, Piechulla B. Volatiles of bacterial antagonists inhibit mycelial growth of the plant pathogen *Rhizoctonia solani*. Archives of Microbiology. 2007;187(5):351–360.
39. Banerjee S, Schlaeppi K, Van der Heijden MG. Keystone taxa as drivers of microbiome structure and functioning. Nature Reviews Microbiology. 2018;16(9):567–576.
40. Bhattacharyya PN, Jha DK. Plant growth-promoting rhizobacteria (PGPR): emergence in agriculture. World Journal of Microbiology and Biotechnology. 2012;28(4):1327–1350.
41. Burns RG, DeForest JL, Marxsen J, Sinsabaugh RL, Stromberger ME, Wallenstein MD, *et al*. Soil enzymes in a changing environment: current knowledge and future directions. Soil Biology and Biochemistry. 2013;58:216–234.
42. Neufeld JD, Vohra J, Dumont MG, Lueders T, Manefield M, Friedrich MW, *et al*. DNA stable-isotope probing. Nature Protocols. 2007;2(4):860–866.
43. Vessey JK. Plant growth promoting rhizobacteria as biofertilizers. Plant and Soil. 2003;255(2):571–586.
44. Bashan Y, de-Bashan LE, Prabhu SR, Hernandez JP. Advances in plant growth-promoting bacterial inoculant technology: formulations and practical perspectives (1998–2013). Plant and Soil. 2014;378(1–2):1–33.
45. Herrmann L, Lesueur D. Challenges of formulation and quality of biofertilizers for successful inoculation. Applied Microbiology and Biotechnology. 2013;97(20):8859–8873.
46. Adesemoye AO, Torbert HA, Kloepper JW. Plant growth-promoting rhizobacteria allow reduced application rates of chemical fertilizers. Microbial Ecology. 2009;58(4):921–929.
47. Bhardwaj D, Ansari MW, Sahoo RK, Tuteja N. Biofertilizers function as key player in sustainable agriculture by improving soil fertility, plant tolerance and crop productivity. Microbial Cell Factories. 2014;13(1):66.
48. Tilman D, Balzer C, Hill J, Befort BL. Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(50):20260–20264.
49. Cordell D, Drangert JO, White S. The story of phosphorus: global food security and food for thought. Global Environmental Change. 2009;19(2):292–305.
50. Geisseler D, Scow KM. Long-term effects of mineral fertilizers on soil microorganisms—a review. Soil Biology and Biochemistry. 2014;75:54–63.
51. Ruiz-Lozano JM, Aroca R, Zamarreño AM, Molina S, Andreo-Jiménez B, Porcel R, *et al*. Arbuscular mycorrhizal symbiosis induces strigolactone biosynthesis under drought and improves drought tolerance in lettuce and tomato. Plant, Cell and Environment. 2016;39(2):441–452.
52. Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, *et al*. Core microbiomes for sustainable agroecosystems. Nature Plants. 2018;4(5):247–257.
53. Mendler K, Chen H, Parks DH, Lobb B, Hug LA, Doxey AC. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids Research. 2019;47(9):4442–4448.
54. European Parliament. Regulation (EU) 2019/1009 on EU fertilising products. Official Journal of the European Union. 2019;L170:1–114.
55. Compant S, Sessitsch A, Mathieu F. The 125th anniversary of the first postulation of the soil origin of endophytic bacteria: a tribute to MLV Galippe. Plant and Soil. 2012;356(1–2):299–301.
56. Kumar P, Dubey RC, Maheshwari DK. Bacillus strains isolated from rhizosphere showed plant growth-promoting and antagonistic activity against phytopathogens. Microbiological Research. 2012;167(8):493–499.
57. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al*. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–589.
58. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, *et al*. Visualization and analysis of

gene expression in tissue sections by spatial transcriptomics. Science. 2016;353(6294):78–82.
59. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology. 2019;37(5):540–546.
60. Sessitsch A, Pfaffenbichler N, Mitter B. Microbiome applications from lab to field: untapping the potential of second-generation inoculants? Trends in Plant Science. 2019;24(10):883–892.
61. Nødvig CS, Nielsen JB, Kogle ME, Mortensen UH. A CRISPR-Cas9 system for genetic engineering of filamentous fungi. PLOS ONE. 2015;10(7):e0133085.

How to Cite This Article

Craven P. Metagenomic and bioinformatics analysis of Sebacina vermifera for next-generation biofertilizer development. Journal of Agricultural Digitalization Research. 2025;6(2):77-97.
doi:10.54660/JADR.2025.6.2.77-97.

Creative Commons (CC) License