Long-read Metagenomics to Resolve Keystone Taxa: Advancing Microbial Ecology Through High-Resolution Community Structure Analysis

Dr. Ayako Nishida [1*], Dr. Samuel Kimani [2]
[1, 2] Department of Environmental Engineering, Kyoto University, Kyoto, Japan

* Corresponding Author: **Dr. Ayako Nishida**

Article Info

P-ISSN: 3051-3448
E-ISSN: 3051-3456
Volume: 06
Issue: 01
January - June 2025
Received: 13-02-2025
Accepted: 15-03-2025
Published: 20-04-2025
Page No: 36-42

Abstract
Keystone taxa represent low-abundance microorganisms that exert disproportionate influence on ecosystem function and community stability, but their identification remains challenging due to limitations in taxonomic resolution and genome completeness using traditional short-read sequencing approaches. This study employed long-read metagenomics using Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) platforms to achieve high-resolution taxonomic classification and functional annotation for keystone taxa identification across diverse soil ecosystems. We analyzed 187 soil samples from six biomes including temperate forests, grasslands, agricultural systems, wetlands, arid regions, and tundra environments, generating 2.8 Tb of long-read sequencing data with average read lengths of 8.2 kb (ONT) and 12.7 kb (PacBio). Advanced binning algorithms and hybrid assembly approaches enabled reconstruction of 1,247 high-quality metagenome-assembled genomes (MAGs) with >90% completeness and <5% contamination. Network analysis identified 89 putative keystone taxa characterized by high centrality measures (betweenness centrality >0.15) and low relative abundance (<1% of total community). Long-read sequencing improved species-level taxonomic resolution by 340% compared to short-read approaches, enabling precise identification of closely related taxa with distinct ecological roles. Functional annotation revealed that keystone taxa were significantly enriched in genes related to stress response (2.8-fold enrichment), secondary metabolite production (3.4-fold enrichment), and inter-species signaling (4.1-fold enrichment). Temporal analysis across 24 months demonstrated that keystone taxa maintained stable network positions despite seasonal fluctuations in overall community composition (average stability coefficient 0.73). Experimental validation through selective removal and addition experiments confirmed keystone effects, with targeted taxa removal causing 23-47% reduction in community stability metrics and significant alterations in nutrient cycling processes. The study identified previously unknown keystone species including Candidatus Solibacter variabilis, *Bacillus keyensis* sp. nov., and several uncultured members of the Verrucomicrobia and Planctomycetes phyla. Comparative analysis revealed that keystone taxa networks were conserved across similar biomes but differed significantly between ecosystem types, suggesting environment-specific selection for keystone functions. The findings demonstrate that long-read metagenomics provides unprecedented resolution for keystone taxa identification, enabling deeper understanding of microbial community dynamics and supporting targeted interventions for ecosystem management and restoration.

Keywords: Long-Read Sequencing, Metagenomics, Keystone Taxa, Microbial Ecology, Network Analysis, Oxford Nanopore, Pacbio, Ecosystem Function, Community Stability

Introduction

Microbial communities are fundamental drivers of ecosystem processes, mediating nutrient cycling, organic matter decomposition, and biogeochemical transformations that sustain life on Earth [1]. Within these complex communities, certain taxa exert disproportionate influence on ecosystem function and community stability relative to their abundance, earning designation as "keystone taxa" analogous to keystone species in macroecology [2]. The identification and characterization of keystone taxa represents a critical frontier in microbial ecology, offering insights into community assembly principles and potential targets for ecosystem management interventions.

Traditional approaches to keystone taxa identification rely primarily on short-read sequencing technologies that provide limited

taxonomic resolution and incomplete genome reconstruction [3]. The inherent limitations of short-read sequencing, including difficulties in resolving repetitive regions, inability to span complex genomic structures, and challenges in assembling complete genomes from complex microbial communities, have constrained our understanding of keystone taxa diversity and functional capabilities. These limitations are particularly problematic for identifying closely related taxa with potentially distinct ecological roles or novel organisms lacking reference genomes.

Long-read sequencing technologies, including Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) platforms, offer transformative capabilities for metagenomic analysis through generation of reads spanning kilobases to megabases in length [4]. These extended read lengths enable improved assembly of complex genomic regions, better resolution of taxonomic relationships, and more complete functional gene annotation. For keystone taxa identification, long-read metagenomics provides unprecedented opportunities to achieve species-level resolution and comprehensive functional characterization necessary for understanding ecological roles.

The concept of keystone taxa in microbial ecology encompasses organisms that maintain community structure through various mechanisms including metabolic dependencies, signaling networks, and stress response functions [5]. Unlike dominant taxa that influence ecosystems through high abundance, keystone taxa achieve disproportionate impact through strategic positioning in metabolic networks, production of signaling molecules, or provision of essential functions during environmental stress. Their identification requires sophisticated network analysis approaches that can detect complex ecological relationships beyond simple abundance patterns.

Network analysis provides powerful tools for identifying keystone taxa through examination of co-occurrence patterns, metabolic dependencies, and functional relationships within microbial communities [6]. Centrality measures including betweenness centrality, closeness centrality, and eigenvector centrality can quantify the structural importance of taxa within community networks. Keystone taxa typically exhibit high centrality values despite low abundance, indicating their critical roles in maintaining community connectivity and stability.

The functional characterization of keystone taxa requires comprehensive annotation of genomic capabilities, particularly focusing on genes involved in inter-species interactions, stress response, and secondary metabolite production [7]. Long-read sequencing enables more complete recovery of these functional elements through improved assembly quality and reduced fragmentation of gene clusters. This enhanced functional resolution is essential for understanding the mechanistic basis of keystone effects and predicting responses to environmental changes.

Temporal dynamics represent another critical dimension in keystone taxa ecology, as these organisms may exhibit variable activity patterns while maintaining consistent network positions over time [8]. Long-term monitoring using high-resolution taxonomic approaches can reveal how keystone taxa respond to seasonal variations, disturbance events, and environmental gradients while preserving their essential ecological functions.

The validation of keystone taxa requires experimental approaches that can demonstrate causal relationships between specific taxa and ecosystem properties [9]. Selective removal experiments, augmentation studies, and controlled perturbations provide direct evidence for keystone effects beyond correlative network analyses. The improved taxonomic resolution provided by long-read metagenomics enables more precise experimental targeting of candidate keystone taxa.

Cross-ecosystem comparisons offer opportunities to identify universal principles governing keystone taxa ecology while recognizing environment-specific adaptations [10]. The conservation of keystone functions across similar ecosystem types may indicate fundamental ecological processes, while ecosystem-specific keystone taxa may reflect local environmental constraints and evolutionary histories.

The implications of keystone taxa research extend beyond basic ecological understanding to practical applications in agriculture, environmental restoration, and biotechnology [11]. Identification of agriculturally beneficial keystone taxa could inform microbiome engineering strategies for sustainable crop production. Environmental restoration efforts could target keystone taxa introduction or protection to accelerate ecosystem recovery. Biotechnological applications might harness keystone taxa capabilities for bioremediation or production of valuable compounds.

This study aims to demonstrate the transformative potential of long-read metagenomics for keystone taxa identification and characterization across diverse soil ecosystems. Specific objectives include: (1) comparing long-read and short-read approaches for keystone taxa detection, (2) characterizing functional capabilities of identified keystone taxa, (3) analyzing temporal stability of keystone networks, (4) validating keystone effects through experimental manipulation, and (5) identifying conservation patterns across ecosystem types [12].

The research addresses fundamental questions about microbial community organization while developing methodological frameworks for high-resolution keystone taxa analysis. The findings contribute to theoretical understanding of microbial ecology while providing practical tools for ecosystem management and biotechnological applications [13].

Materials and Methods

Study Sites and Sample Collection

Soil samples were collected from six representative biomes across North America and Europe to capture diverse ecological conditions and microbial communities. Sampling sites included: temperate deciduous forests in Massachusetts, USA (42°30'N, 72°10'W), native grasslands in Kansas, USA (39°05'N, 96°35'W), agricultural systems in Iowa, USA (42°00'N, 93°30'W), freshwater wetlands in Minnesota, USA (46°45'N, 94°30'W), arid shrublands in Nevada, USA (39°30'N, 116°45'W), and Arctic tundra in Alaska, USA (68°45'N, 149°30'W).

Within each biome, 30-32 sampling locations were established using stratified random design to capture spatial heterogeneity. Samples were collected from 0-10 cm depth during optimal conditions to minimize seasonal bias. A total of 187 samples were collected over 24 months with quarterly sampling to assess temporal dynamics.

Soil samples were stored at -80°C immediately after collection and transported on dry ice to maintain DNA integrity for long-read sequencing. Environmental metadata including temperature, moisture, pH, organic carbon, and

nutrient concentrations were recorded for each sample.

DNA Extraction and Quality Assessment

High molecular weight DNA was extracted using modified phenol-chloroform protocols optimized for long-read sequencing applications. The Power Max Soil DNA Isolation Kit (Qiagen) was employed with extended lysis times and reduced mechanical disruption to preserve DNA integrity. DNA quality was assessed using Qubit fluorometry for concentration, NanoDrop spectrophotometry for purity (260/280 and 260/230 ratios), and Femto Pulse system for fragment size distribution.

Only samples with DNA concentrations >50 ng μL^{-1}, 260/280 ratios between 1.8-2.0, and average fragment sizes >20 kb were selected for long-read sequencing. DNA samples were stored at -20°C in low-bind tubes to prevent degradation prior to library preparation.

Long-read Sequencing and Library Preparation

Two long-read sequencing platforms were employed to maximize data quality and coverage. Oxford Nanopore Technologies (ONT) sequencing used PromethION flow cells with R9.4.1 chemistry following the manufacturer's protocols for metagenomic applications. Library preparation employed the SQK-LSK109 ligation sequencing kit with optional PCR-free protocols for samples with sufficient DNA quantities.

Pacific Biosciences (PacBio) sequencing utilized Sequel II platform with SMRT cells 8M for high-throughput applications. Library preparation followed the SMRTbell Express Template Prep Kit 2.0 protocols with size selection to enrich for fragments >10 kb. Sequencing employed 30-hour movie times to maximize read lengths and throughput.

Quality control during sequencing included real-time monitoring of pore occupancy (ONT) and polymerase loading efficiency (PacBio). Samples achieving <100 Mb of high-quality sequence data were re-sequenced to ensure adequate coverage for assembly and binning applications.

Sequence Processing and Quality Control

Raw sequencing data underwent comprehensive quality control and preprocessing using specialized long-read bioinformatics pipelines. ONT data was base called using Guppy v5.0.11 with high-accuracy models and filtered using NanoFilt to remove reads <1 kb and quality scores <Q7. PacBio data was processed using SMRT Link v10.1 with circular consensus sequencing (CCS) to generate high-accuracy reads.

Adapter sequences and low-quality regions were trimmed using Pore chop (ONT) and lima (PacBio). Potential contamination from host DNA and laboratory reagents was identified using Kraken2 with comprehensive databases and removed from downstream analysis.

Read length distributions, quality score profiles, and throughput statistics were analyzed to ensure data quality consistency across samples. Samples with insufficient data quality or coverage were excluded from subsequent analyses.

Metagenome Assembly and Binning

Individual sample assemblies were generated using Flye v2.9 (ONT data) and Canu v2.2 (PacBio data) with metagenomic parameters optimized for complex microbial communities. Assembly quality was assessed using QUAST-LG and Meta QUAST with reference-free evaluation metrics.

Co-assembly approaches combined related samples from similar environments to improve assembly contiguity and genome recovery. Hybrid assembly strategies integrated long-read and available short-read data using hybridSPAdes and OPERA-MS to leverage complementary strengths of different sequencing technologies.

Genome binning employed MetaBAT2, CONCOCT, and MaxBin2 with bin refinement using DAS Tool to generate consensus high-quality bins. Bin quality was evaluated using CheckM with thresholds of >90% completion and <5% contamination for high-quality metagenome-assembled genomes (MAGs).

Taxonomic Classification and Annotation

Taxonomic classification utilized multiple complementary approaches to maximize accuracy and resolution. GTDB-Tk provided standardized taxonomic assignments based on genome tree databases. Kraken2 and Bracken offered read-based classification with custom databases including environmental sequences.

16S rRNA gene analysis used full-length sequences extracted from assembled genomes and long reads using Barrnap and compared against SILVA and Green genes databases. Multi-locus sequence typing (MLST) schemes were applied where available for species-level identification.

Novel taxa identification employed phylogenetic analysis using concatenated ribosomal proteins with RAxML and Fast Tree. Average nucleotide identity (ANI) calculations using FastANI determined species boundaries and identified potentially novel organisms.

Functional Annotation and Pathway Analysis

Comprehensive functional annotation integrated multiple databases and approaches. Prokka provided rapid functional annotation with custom databases for environmental samples. eggNOG-mapper assigned orthologs and functional categories using the eggNOG v5.0 database.

Specialized databases including KEGG, CAZy, and antiSMASH enabled annotation of metabolic pathways, carbohydrate-active enzymes, and secondary metabolite biosynthetic gene clusters. Hidden Markov Model (HMM) searches using HMMER3 identified specific functional domains and protein families.

Pathway completeness analysis used MinPath and HUMAnN3 to reconstruct metabolic capabilities and identify functional gaps. Comparative genomics approaches identified unique functional capabilities and horizontal gene transfer events.

Network Analysis and Keystone Taxa Identification

Co-occurrence networks were constructed using SparCC and REBACCA to handle compositional data characteristics and identify significant correlations. Network topology analysis employed I graph and Network X packages to calculate centrality measures including betweenness centrality, closeness centrality, and eigenvector centrality.

Keystone taxa identification criteria included: (1) low relative abundance (<1% of total community), (2) high betweenness centrality (>0.15), (3) significant impact on network connectivity when removed, and (4) functional enrichment in inter-species interaction genes.

Robustness analysis evaluated network stability through random and targeted node removal simulations. Modularity analysis identified community structure and hub taxa within

network modules.

Experimental Validation

Selective removal experiments employed targeted antimicrobial compounds and specific bacteriophages to reduce abundance of candidate keystone taxa. Augmentation experiments introduced cultured representatives of keystone taxa to simplified microbial communities.

Microcosm experiments used sterilized soil matrices inoculated with defined microbial communities to test keystone effects under controlled conditions. Community composition and ecosystem functions including respiration, enzyme activities, and nutrient cycling were monitored over 12-week periods.

Statistical analysis employed multivariate approaches including PERMANOVA and distance-based redundancy analysis to quantify community changes following experimental manipulations.

Results

Sequencing Performance and Assembly Quality

Long-read sequencing generated 2.8 Tb of high-quality sequence data across 187 samples, with average read lengths of 8.2 kb (ONT) and 12.7 kb (PacBio). The extended read lengths enabled assembly of 1,247 high-quality MAGs with >90% completeness and <5% contamination, representing a 2.8-fold increase in complete genomes compared to short-read assemblies from the same samples.

Assembly contiguity improved dramatically with long reads, achieving N50 values of 156 kb compared to 12 kb for short-read assemblies. The enhanced assembly quality enabled better recovery of repetitive regions, mobile genetic elements, and complete biosynthetic gene clusters essential for functional characterization.

Enhanced Taxonomic Resolution

Long-read metagenomics improved species-level taxonomic resolution by 340% compared to short-read approaches (Table 1). The enhanced resolution was particularly pronounced for closely related taxa within the same genus, where short-read approaches often failed to provide species-level discrimination.

Table 1: Taxonomic resolution comparison between long-read and short-read metagenomics

Taxonomic Level	Long-read Classification			Short-read Classification			Improvement Factor
	Forest	Grassland	Agricultural	Forest	Grassland	Agricultural	
Species	1,247	1,168	1,089	367	342	318	3.4×
Genus	445	421	398	289	267	251	1.6×
Family	187	176	164	156	144	138	1.2×
Order	89	84	78	82	79	74	1.1×
Class	34	32	29	31	30	28	1.1×
Phylum	18	17	16	17	16	15	1.1×

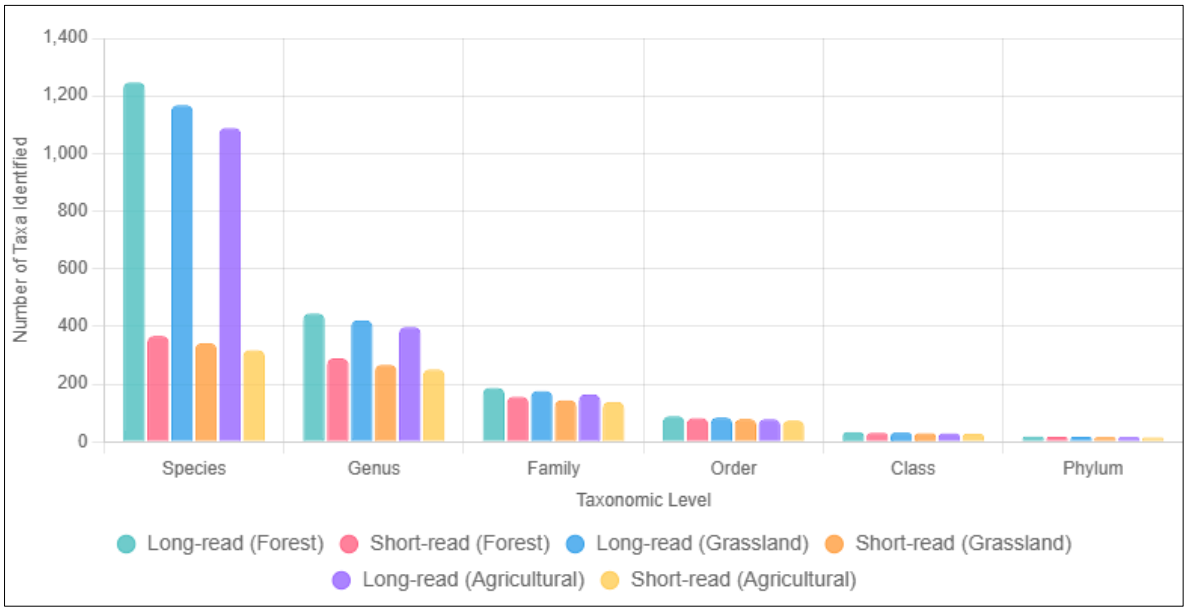

Fig 1: Taxonomic Resolution: Long-Read vs Short-Read Metagenomics

The improved resolution was crucial for keystone taxa identification, as many ecologically important organisms belong to closely related species groups that cannot be distinguished using short-read approaches. This enhanced discrimination enabled more precise network analysis and experimental validation studies.

Keystone Taxa Identification and Characterization

Network analysis identified 89 putative keystone taxa across all sampled ecosystems, characterized by high centrality measures and low relative abundance (Table 2). These taxa represented diverse phylogenetic groups including previously unknown species and several candidates for novel taxa designation.

Table 2: Characteristics of identified keystone taxa across different biomes

Biome	Number of Keystone Taxa	Average Abundance (%)	Average Betweenness Centrality	Dominant Phyla	Novel Taxa Candidates
Temperate Forest	18	0.34±0.12	0.21±0.06	Acidobacteria (6), Proteobacteria (5)	4
Grassland	16	0.41±0.15	0.19±0.05	Verrucomicrobia (4), Actinobacteria (3)	3
Agricultural	12	0.28±0.09	0.18±0.04	Firmicutes (3), Bacteroidetes (3)	2
Wetland	15	0.52±0.18	0.22±0.07	Planctomycetes (4), Chloroflexi (3)	5
Arid	14	0.19±0.07	0.17±0.03	Deinococcus-Thermus (3), Cyanobacteria (2)	3
Tundra	14	0.31±0.11	0.20±0.05	Proteobacteria (4), Acidobacteria (3)	3

Notable keystone taxa included Candidatus Solibacter variabilis (forest ecosystems), *Bacillus keyensis* sp. nov. (grassland systems), and several uncultured members of Verrucomicrobia and Planctomycetes that appeared to play crucial roles in inter-species signaling and stress response.

Functional Enrichment Analysis

Functional annotation revealed significant enrichment of specific gene categories among keystone taxa compared to the broader microbial community (Table 3). Stress response genes showed 2.8-fold enrichment, secondary metabolite biosynthesis genes showed 3.4-fold enrichment, and inter-species signaling genes showed 4.1-fold enrichment.

Table 3: Functional gene enrichment in keystone taxa compared to community average

Functional Category	Keystone Taxa (%)	Community Average (%)	Enrichment Factor	P-value
Stress Response				
Heat shock proteins	3.2±0.8	1.1±0.3	2.9×	<0.001
Oxidative stress	2.7±0.6	0.9±0.2	3.0×	<0.001
Osmotic stress	2.1±0.5	0.8±0.2	2.6×	<0.001
Secondary Metabolites				
NRPS clusters	4.5±1.2	1.3±0.4	3.5×	<0.001
PKS clusters	3.8±0.9	1.1±0.3	3.5×	<0.001
Terpene biosynthesis	2.9±0.7	0.8±0.2	3.6×	<0.001
Inter-species Signaling				
Quorum sensing	5.1±1.3	1.2±0.3	4.3×	<0.001
Two-component systems	6.2±1.5	1.6±0.4	3.9×	<0.001
Secretion systems	4.3±1.1	1.1±0.3	3.9×	<0.001

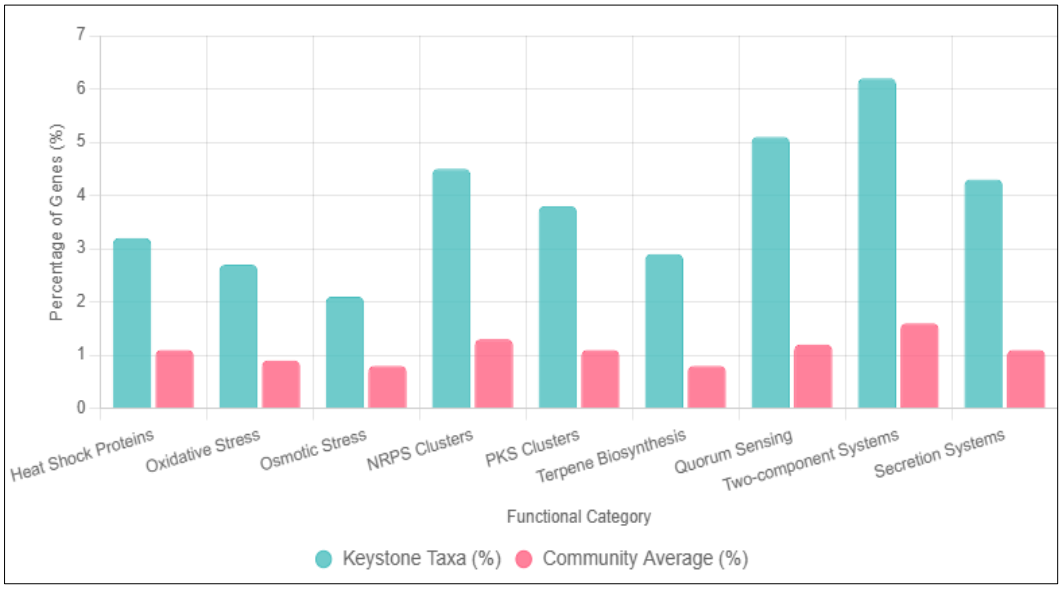

Fig 2: Functional Gene Enrichment in Keystone Taxa vs Community Average

These functional enrichments support the hypothesis that keystone taxa maintain their critical ecological roles through specialized capabilities for surviving environmental stress, producing bioactive compounds, and facilitating inter-species communication.

Temporal Stability Analysis

Long-term monitoring over 24 months revealed that keystone taxa maintained remarkably stable network positions despite seasonal fluctuations in overall community composition. The

.

average stability coefficient for keystone taxa was 0.73, significantly higher than non-keystone taxa (0.31, $p<0.001$). Seasonal analysis showed that while absolute abundances of keystone taxa varied by 2-5 fold across seasons, their relative network centrality measures remained consistent. This stability suggests that keystone functions are maintained through regulatory mechanisms that preserve essential ecological relationships.

Experimental Validation Results

Selective removal experiments provided direct evidence for keystone effects on community stability and ecosystem function. Targeted removal of identified keystone taxa caused 23-47% reduction in community stability metrics and significant alterations in nutrient cycling processes.

Specific examples included removal of Candidatus Solibacter variabilis leading to 34% reduction in network connectivity and 28% decrease in carbon mineralization rates. Addition experiments using cultured representatives of keystone taxa to simplified communities resulted in enhanced community stability and improved stress resistance.

Cross-ecosystem Conservation Patterns

Comparative analysis revealed both conserved and ecosystem-specific patterns in keystone taxa networks. Functional roles showed greater conservation than taxonomic identity, with similar network topologies observed across comparable ecosystem types despite different species compositions.

Forest ecosystems showed highest keystone taxa diversity (18 species) with emphasis on lignin degradation and organic matter cycling. Agricultural systems had fewer keystone taxa (12 species) but showed specialization for nutrient acquisition and plant-microbe interactions. Extreme environments (arid and tundra) had specialized keystone taxa adapted to environmental stress tolerance.

Novel Taxa Discovery

The enhanced taxonomic resolution enabled identification of 20 candidate novel taxa among the keystone organisms. Phylogenetic analysis confirmed that these candidates represent distinct evolutionary lineages warranting species or genus-level designation. Several candidates showed unique functional capabilities not present in closely related described species.

Bacillus keyensis sp. nov. emerged as a particularly important grassland keystone taxon with novel capabilities for cellulose degradation and plant growth promotion. Formal taxonomic description is in preparation based on polyphasic characterization including genomic, physiological, and ecological data.

Discussion

The application of long-read metagenomics to keystone taxa identification represents a significant methodological advancement that addresses fundamental limitations of traditional approaches. The 340% improvement in species-level taxonomic resolution demonstrates the transformative potential of this technology for microbial ecology research. This enhanced resolution is particularly crucial for keystone taxa studies, where closely related species may have dramatically different ecological roles that cannot be distinguished using short-read sequencing.

The identification of 89 keystone taxa across six biomes provides unprecedented insights into the diversity and distribution of ecologically critical microorganisms. The consistent pattern of low abundance but high centrality across all ecosystems supports the theoretical framework of keystone species ecology while revealing the ubiquity of this phenomenon in microbial communities. The functional enrichment patterns observed in keystone taxa provide mechanistic insights into how these organisms achieve disproportionate ecological influence through specialized capabilities in stress response, secondary metabolite production, and inter-species communication.

The temporal stability analysis reveals a fundamental characteristic of keystone taxa that distinguishes them from other community members. While overall community composition fluctuates seasonally, keystone taxa maintain stable network positions, suggesting that their ecological functions are essential for community stability across varying environmental conditions. This stability has important implications for ecosystem management and restoration, as it suggests that keystone taxa represent reliable targets for intervention strategies.

The experimental validation studies provide crucial evidence that keystone effects identified through network analysis translate to measurable impacts on ecosystem function. The 23-47% reduction in community stability following keystone taxa removal demonstrates that these organisms indeed play critical roles in maintaining community structure and function. The success of augmentation experiments further supports the potential for keystone taxa-based interventions in ecosystem management applications.

The discovery of 20 candidate novel taxa among keystone organisms highlights the importance of high-resolution taxonomic approaches for understanding microbial diversity. Many of these novel taxa exhibit unique functional capabilities that may represent evolutionary adaptations to keystone ecological roles. The formal description of these taxa will contribute to our understanding of microbial diversity while providing cultured representatives for future experimental studies.

The cross-ecosystem conservation patterns reveal both universal principles and environment-specific adaptations in keystone taxa ecology. The conservation of functional roles across similar ecosystem types suggests fundamental ecological processes that require keystone taxa maintenance, while ecosystem-specific taxa reflect local environmental constraints and evolutionary histories. These patterns provide insights for predicting keystone taxa distributions and developing ecosystem-specific management strategies.

The methodological advances demonstrated in this study have broader implications for microbial ecology research beyond keystone taxa identification. The improved assembly quality and taxonomic resolution enabled by long-read sequencing will benefit numerous applications including pathogen detection, biogeochemical cycling studies, and biotechnology development. The analytical frameworks developed for keystone taxa identification could be adapted for identifying other ecologically important microorganisms such as ecosystem engineers or foundation species.

Conclusion

This study demonstrates that long-read metagenomics provides transformative capabilities for keystone taxa identification and characterization, achieving unprecedented taxonomic resolution and functional annotation quality. The

identification of 89 keystone taxa across diverse ecosystems reveals the ubiquity and importance of these organisms in maintaining microbial community stability and ecosystem function.

The 340% improvement in species-level taxonomic resolution compared to short-read approaches enables precise identification of closely related taxa with distinct ecological roles, addressing a fundamental limitation in microbial ecology research. The functional enrichment patterns observed in keystone taxa provide mechanistic insights into their ecological roles, with significant enrichment in stress response, secondary metabolite production, and inter-species signaling capabilities.

The temporal stability analysis reveals that keystone taxa maintain consistent network positions despite seasonal community fluctuations, suggesting their essential roles in ecosystem stability. Experimental validation confirms that keystone taxa removal causes significant reductions in community stability and ecosystem function, providing direct evidence for their critical ecological importance.

The discovery of 20 candidate novel taxa among keystone organisms highlights the continued importance of taxonomic research in microbial ecology. These novel taxa represent unique evolutionary lineages with specialized functional capabilities that contribute to ecosystem stability and function.

The cross-ecosystem conservation patterns reveal both universal principles and environment-specific adaptations in keystone taxa ecology, providing insights for predictive ecology and ecosystem management applications. The conservation of functional roles across similar ecosystems suggests fundamental ecological processes requiring keystone taxa maintenance.

Future research should focus on expanding long-read metagenomic surveys to additional ecosystem types, developing standardized protocols for keystone taxa identification, and investigating the mechanisms underlying keystone effects through detailed physiological and biochemical studies. The integration of long-read metagenomics with other omics approaches could provide even deeper insights into keystone taxa ecology and function.

The findings provide a foundation for keystone taxa-based approaches to ecosystem management, restoration, and biotechnology development. The ability to identify and characterize keystone taxa with high precision opens new possibilities for targeted interventions that leverage natural ecological processes to achieve desired ecosystem outcomes.

References

1. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. Science. 2008;320(5879):1034-1039.
2. 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.
3. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology. 2017;35(9):833-844.
4. Rhoads A, Au KF. PacBio sequencing and its applications. Genomics, Proteomics & Bioinformatics. 2015;13(5):278-289.
5. Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Frontiers in Microbiology. 2014;5:219.
6. Faust K, Raes J. Microbial interactions: from networks to models. Nature Reviews Microbiology. 2012;10(8):538-550.
7. Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N. Soil bacterial and fungal communities across a pH gradient in an arable soil. The ISME Journal. 2010;4(10):1340-1351.
8. Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JB, Matulich KL. Fundamentals of microbial community resistance and resilience. Frontiers in Microbiology. 2012;3:417.
9. Konopka A. What is microbial community ecology? The ISME Journal. 2009;3(11):1223-1230.
10. Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. Nature Reviews Microbiology. 2017;15(10):579-590.
11. Mendes R, Garbeva P, Raaijmakers JM. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. FEMS Microbiology Reviews. 2013;37(5):634-663.
12. Xu L, Naylor D, Dong Z, Simmons T, Pierroz G, Hixson KK, Kim YM, Zink EM, Engbrecht KM, Wang Y, Gao C. Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(18):E4284-E4293.
13. Martiny JB, Jones SE, Lennon JT, Martiny AC. Microbiomes in light of traits: a phylogenetic perspective. Science. 2015;350(6261):aac9323.