Agri Articles
(e-Magazine for Agricultural Articles)
Volume: 06, Issue: 05 (SEP-OCT, 2026)
Available online at http://www.agriarticles.com
©Agri Articles, ISSN: 2582-9882

The Secret Lives in the Soil: How DNA Is Regrouping Soil Community

*Mekala Manjusha and Anu Rajan S
Dept. of Agricultural Microbiology, College of Agriculture, Vellayani, Kerala, India
Corresponding Author's email: mekalamanjusha2000@gmail.com

A single gram of soil hides more life (10^8) than there are people on Earth - and until recently, we could only identify 1% of it. A single gram of soil contains up to ten billion microorganisms: bacteria, fungi, archaea, algae and protozoa, competing for space and food in tangled pores and mineral surfaces. This invisible community helps in nutrient cycling, builds the porous structure that lets roots breathe, keeps plant diseases in check, and even decides how much greenhouse gas escapes into the atmosphere. Most of the soil microbiome were, quite literally, unculturable - invisible to humans who try to put them in lab. The discrepancy between the number of microbes observed in an environmental sample (using direct microscopy) and the number of those microbes that can actually be cultivated on standard agar culture plates is explained through the great plate count anomaly.

The invisible investors of nature

In 1985, microbiologists James Staley and Allen Konopka noticed something very strange. they took a handful of soil and looked directly under microscope but when they tried to culture it there is none left which paved a way for the great plate count anomaly theory. It turns out many soil microbes are simply not designed for the artificial conditions provided in lab. Some exist in a strange, dormant state called the viable-but-non-culturable (VBNC) state, alive but refusing to grow (Harmani, 2012). A lot of microbes preferred environmental conditions and nutritional requirements are yet to be discovered. All these reasons left microbiologists in an odd position where, they could describe soil's chemistry and physics in very intensive way but they can't show microbes in reality.

How DNA is helping to look straight into soil communities

The breakthrough came from giving up on culturing artificially and reading microbial DNA straight out of the soil. This shift from organism to DNA is molecular ecology which meant researchers no longer needed an entire microbe to be present in lab it's genome itself is enough to recognise it's presence. Three major techniques do most of the work. Metabarcoding amplifies a short, standardised stretch of DNA, the 16S gene for bacteria and archaea, the ITS region for fungi, to produce a quick taxonomic library on microbes present in soil (Taberlet *et al.,* 2012). Metagenomics goes further, by sequencing all the DNA in a sample at once, revealing not just identity but genetic potential: what these organisms are biochemically capable of doing. Finally metatranscriptomics captures the messenger RNA being actively made at the moment of sampling, showing which genes are switched on right now rather than merely present in the genome, explaining the actual work done by the genes. Different genes serves different kind of tagger. The 16S rRNA gene is the classic all purpose marker, but because bacteria often carry several slightly different copies of it, it can blur close relatives together. Single copy genes like *rpoB* or *gyrB* sharpen that resolution. And some markers are chosen not for identity but for job description: *nifH* identifies organisms capable of fixing nitrogen, while *amoA* and *pmoA* mark those that oxidise ammonia or methane, letting ecologists connect a microbe's family tree directly to its function in the ecosystem.

Bioinformatic approach

Once you have millions of DNA reads, you need a sensible way to group them into meaningful units, otherwise sequencing noise and true biological diversity become impossible to tell apart. For years the standard was the Operational Taxonomic Unit (OTU): sequences sharing roughly 97% similarity were lumped together, on the reasoning that this loosely tracks species level differences (Fasolo *et al.,* 2024). Modern pipelines increasingly favours Amplicon Sequence Variants (ASVs) instead, which use statistical error correction to resolve differences down to a single nucleotide rather than a fixed similarity cutoff. The choice genuinely changes the ecological story a dataset tells, sometimes inflating or deflating diversity estimates depending on which method is used, which is why comparing results across older and newer studies requires some caution (Fasolo *et al.,* 2024).

Whole genome directly from soil

The most striking advance of the last decade is the ability to reconstruct entire genomes of organisms that have never been cultured at all, through what are called metagenome assembled genomes (MAGs). Shotgun sequencing reads every fragment of DNA in a soil sample; assembly software stitches overlapping fragments into longer contigs; binning algorithms then sort those contigs into genome sized bundles based on shared composition and abundance patterns; and quality control tools check that each reconstructed genome is complete and uncontaminated enough to trust. The end result is a genome for an organism (Kang *et al.,* 2019). This genome first approach has also triggered a quiet taxonomic revolution. Classifying microbes on a single gene like 16S is not always a prime method since there is horizontal gene transfer among unrealted microbes and loses resolution at deeper branches of the tree of life. Phylogenomics instead combines dozens to hundreds of conserved marker genes into a single, far sturdier tree. Built on this foundation, the Genome Taxonomy Database (GTDB) has been steadily replacing over a century of traditional, sometimes inconsistent naming conventions with one grounded directly in genomic evidence and it now spans hundreds of thousands of genomes, the majority representing organisms known only from their DNA (Chaumeil *et al.,* 2020). Species boundaries themselves have been re-drawn along the way. Rather than the old, laborious DNA-DNA hybridisation experiments, researchers now lean on Average Nucleotide Identity (ANI), with roughly 95% similarity serving as a practical, reproducible line between one prokaryotic species and the next (Jain *et al.,* 2018).

Emerging techniques to study soil microbes

Newer sequencing platforms are pushing resolution down to the level of individual strains. Long-read technologies from PacBio and Oxford Nanopore can read entire gene stretches in one continuous pass, untangling repetitive genomic regions that outpasses older, shorter-read methods. The payoff has been dramatic: a 2025 long read survey of Danish soils and sediments, fittingly named the Microflora Danica project uncovered genomes for more than 15,000 previously undescribed prokaryotic species in one sweep, expanding science's known catalogue of bacterial and archaeal life by roughly 8% (Sereika *et al.,* 2025). Meanwhile, single-cell genomics can now sequence one isolated microbial cell at a time, and a technique called proximity ligation (Hi-C) helps untangle which stray DNA fragments truly belong to the same organism within a messy mixed community (Bickhart *et al.,* 2022). Machine learning has quietly become part of the microbiologist's toolkit too. Algorithms trained to recognise short DNA "fingerprints" can sort billions of sequencing reads into taxonomic groups in a fraction of the time older alignment based methods would need, and newer deep learning models are beginning to make educated guesses about organisms so unfamiliar that no close relative exists in any database at all.

Conclusion

Knowing precisely who lives in soil and what they are biochemically capable of has direct, practical stakes. It enhances our ability to track how farming practices, deforestation or urban expansion declines biodiversity long before visible symptoms appear. It guides the search for

new biofertilisers and biocontrol agents and a large share of the antibiotics used in modern medicine. The genomic databases we rely on remain highly biased toward microbes that are easy to culture or medically important, leaving vast stretches of soil specific life under represented.

References

1. Bickhart, D. M., Kolmogorov, M., Tseng, E., Portik, D. M., Korobeynikov, A., Tolstoganov, I., Uritskiy, G., Liachko, I., Sullivan, S. T., Shin, S. B., and Zorea, A. 2022. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat. Biotechnol*. 40(5): 711-719.
2. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinforma*. 36(6): 1925-1927.
3. Fasolo, A., Deb, S., Stevanato, P., Concheri, G., and Squartini, A. 2024. ASV vs OTUs clustering: Effects on alpha, beta, and gamma diversities in microbiome metabarcoding studies. *PLoS One,* 19(10): e0309065.
4. Harwani, D. 2012. The great plate count anomaly and the unculturable bacteria. *Int. J. Sci. Res.* 2: 350-351.
5. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun*. 9(1): 5114.
6. Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* 7: e7359.
7. Sereika, M., Mussig, A. J., Jiang, C., Knudsen, K. S., Jensen, T. B. N., Petriglieri, F., Yang, Y., Jørgensen, V. R., Delogu, F., Sørensen, E. A., and Nielsen, P. H. 2025. Genome-resolved long-read sequencing expands known microbial diversity across terrestrial habitats. *Nat. Microbiol.* 10(8): 2018-2030.
8. Staley, J. T. and Konopka, A. 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Ann. Rev. Microbiol*. 39(1): 321-346.
9. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., and Willerslev, E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21(8): 2045-2050.