

Genomics: Discovery in the dirt

Soil microbes are notoriously hard to culture, so how can we make the ground yield its secrets? Virginia Gewin finds that genetic sequencing — of samples not species — may be the answer.

Virginia Gewin



Leonardo da Vinci once remarked that "We know more about the movement of celestial bodies than about the soil underfoot." You could argue that his insight still holds true the best part of 500 years later. But new genomic technologies mean that the microscopic bodies that enliven soil may be about to get the attention they deserve — if not as individuals, then as communities.

Twice as much carbon is stored in Earth's soil as exists in the plants that grow from it and the animals that depend on them. It is the soil's microbes that are responsible for recycling this carbon, and other nutrients. Living in the fractal jumble of weathered rock, mineral particles and decaying organic matter are a cast of thousands, some say millions, of species. These soil organisms occupy an endless foam of tiny niches; they purify water, detoxify harmful substances and recycle waste products. They restore carbon dioxide to the air and make the atmosphere's nitrogen available to plants. Without them, continents would be deserts — home to little more than lichen, and not much of that.

Knowing which processes soil microbes are responsible for, and how, is increasingly important to everyone from farmers to climate planners. Microbiologists want to see how the organisms communicate with each other and refine the niches in which they live. Drug developers want to know how the soil microbes poison each other with antibiotics, and other commercially minded researchers think they could discover useful industrial enzymes or additives for biofuels. But beyond a gross understanding of inputs and outputs, the specific ecological roles of microbial species or communities in the dirt remain elusive.

The main stumbling block has been that up to 99% of soil microbes cannot be grown in laboratory cultures — the traditional way to study microorganisms (see '[Cultural renaissance](#)'). But as genetic technology has improved, it has provided ways around this. First, it managed to reveal the general level of biodiversity through the sheer quantity of different sequences, then it allowed scientists to trawl for individual genes. Now technology offers the promise of extracting not just the genomes of the creatures in the soil, but — in a sense — the genome of the soil

itself.

“ Once the diversity of the microbial world is catalogued, it will make astronomy look like a pitiful science. — Julian Davies ”

The physical, chemical and biological complexity of the soil make this sort of study a daunting prospect, even for a hugely ambitious gene-sequencer such as Craig Venter. Following his work on the human genome, Venter has been using the sequencing muscle of his institutes to look at samples from seas around the world. But soil's greater complexity makes it the ultimate challenge for such studies, and arguably the biggest open frontier in biology today.

Julian Davies, a pioneer of soil genomics at the University of British Columbia in Canada, thinks that finding out what is really going on in the soil will turn our world upside down — making that below us more wonderful than that above. "Once the diversity of the microbial world is catalogued," says Davies, "it will make astronomy look like a pitiful science."

The full extent of this diversity is still up for debate. But its magnitude was suggested 20 years ago by Norman Pace, a microbiologist currently at the University of Colorado, Boulder. Pace adapted techniques that Carl Woese, a microbiologist at the University of Illinois at Urbana-Champaign, had used to probe the genetic relationships between all living things. Woese's work reclassified the microbial world using variations in the sequence of a highly conserved RNA molecule; Pace applied his technique to DNA taken from the environment.

Although Pace's studies excelled at revealing the diversity in soil, they were less good at suggesting what that diversity did. "We knew enough to know that we were ignorant about most of the organisms, but we still couldn't get our hands on them," says Jo Handelsman, a plant pathologist at the University of Wisconsin, Madison.

The problem was partly solved as researchers refined methods for taking genes from the vast number of organisms they couldn't culture and inserting them into lab workhorses such as *Escherichia coli* — allowing analysis of particular genes and functions. Handelsman dubbed these studies of DNA from uncultured organisms 'metagenomics', and was one of the first to take advantage of it.

Because many antibiotics have been derived from soil microbes that do grow in cultures, such drugs were an obvious thing to look for in those that do not. So far, Handelsman and her collaborators have turned up antibiotics called turbomycin A and B (ref. [1](#)), and Davies has discovered one he named terragine². Companies were swiftly set up to capitalize on the promise of 'functional screens' for possible antibiotics or industrial compounds; they included Diversa, based in San Diego, California, and TerraGen Discovery — now part of Cubist Pharmaceuticals in Lexington, Massachusetts.

But none of the dozen or so compounds found through metagenomic studies of soil seems set to be useful. Davies, who founded TerraGen in 1996, resignedly describes the situation as "rather disappointing". One of the problems is that the screens ignore a lot of the soil's natural diversity; genes from many of the microbes may simply not be expressed when cloned into *E. coli*. In fact, terragine is one of them: it was isolated from genes inserted into recombinant streptomycetes.

The utility of resistance

Although the search continues for new antibiotics, using metagenomics to

understand soil's role in antibiotic resistance holds perhaps greater promise. Indeed, Davies has a theory that antibiotics evolved as signalling molecules that drove the development of sensing and evasion strategies in microbes. The idea is gaining popularity and leading to exciting links between environmental and medical microbiology. Recent work by Gerard Wright, a microbiologist at McMaster University in Ontario, demonstrates that soil microbes have amassed a bevy of tactics to elude the onslaught of antibiotics. On average, every one of 480 strains of *Streptomyces* that Wright harvested from soils across Canada was able to survive 7 or 8 of the 21 antibiotics presented to it — to many of which it had probably never been exposed³. Tracing genes for antibiotic resistance is an area that is ripe for study by metagenomics, enthuses Handelsman.

The search for enzymes to improve the specificity, efficiency and sustainability of industrial reactions has already had some success. In one desert soil sample, researchers at Diversa unearthed more than 100 enzymes that cut up esters and lipids for use in such reactions. With only 200 esterase enzymes known before this work, discovering half as many again is no small accomplishment. For the past five years, Diversa has turned its attention to developing products, but many sit stuck in a bottleneck of government regulations, as approval processes can take several years to navigate.



AP PHOTO/M. HOUSTON

Of course, the functions of genes found in soil do not have to be commercially useful in order to be ecologically enlightening. In simpler environments, such as the sea, metagenomic studies have yielded remarkable discoveries — for example, microbes that make use of sunlight without the benefit of chlorophyll. These organisms are no marginal oddity, having recently been shown to account for 13% of the microbes in the shallow parts of the ocean⁴.

Soil ecologists may be envious of such a breakthrough in marine systems, but recent findings suggest that soil's secrets may not remain hidden much longer. Christa Schleper is a microbial ecologist at the University of Bergen, Norway, with a laboratory that produces libraries of large DNA fragments. She recently discovered that some of the oxidation of ammonia that goes on in the soil is being carried out by genes that come not from bacteria, but from archaea. The archaea were established as a separate kingdom of life by Woese, and until recently they were seen as being of ecological importance only in marginal and extreme environments. Now they are being discovered playing mainstream roles. Schleper's archaea findings^{5,6} are already causing scientists to reexamine the nitrogen cycle — the conversion of nitrogen to and from forms needed by plants and animals.

The success of studies such as Schleper's will not only spur work on as-yet-uncultured organisms, but also generate hypotheses about community dynamics and functions. Schleper notes that soil contains a lot of archaea, but that in moderate environments they are much less diverse than the bacteria they live alongside. So, did they always inhabit moderate environments, being overrun at a later point by more diverse bacteria? Or did they invade, and if so, how did they acquire their niche?

Nitty gritty

"We need more studies of the genes discovered by functional screens" in order to take such work further, says Rolf Daniel, a microbiologist at the Georg-August

University in Göttingen, Germany. "Massive sequencing efforts are the starting point." Instead of cloning genes into bacteria and then studying the proteins that are expressed, large-scale sequencing allows scientists to analyse the genes directly, comparing their sequences with those of genes of known function. Ecology groups in the United States and Europe are trying to piece together the funding necessary for such sequencing capacity.

This is where Venter comes in. He plans to use the 'shotgun' sequencing technique — in which all the DNA in a sample is cut into small fragments, sequenced and then pieced together — to create a genetic inventory of the planet. In 2004, he produced a paper that sequenced the DNA in a sample of the Sargasso Sea: it yielded 1.2 million new genes and evidence for 148 new species⁷. In addition to the marine samples taken as part of his *Beagle*-emulating voyages aboard the *Sorcerer II*, Venter is also taking soil samples. Soil searchers have greeted this sally on to — indeed, into — their turf with a degree of scepticism and even jealousy. Only a handful of soil labs are funded well enough to construct metagenomic libraries like Schleper's, whereas the various institutes he has set up make Venter a one-man superpower in sequencing terms. Eugene Madsen, a microbial ecologist at Cornell University in Ithaca, New York, grudgingly agrees that Venter's enormous sequencing ability will probably be good for the research community. But he wonders whether Venter will end up with anything more than a hodgepodge of random sequences.

“ The soil work is like trying to assemble 40,000 human genomes simultaneously. — Craig Venter ”

Venter freely admits that, compared with the clear waters of the Sargasso Sea, soils are proving a challenge. Sea water is well mixed and unstructured. You can take as much of it as you want — the Sargasso work took 200 litres — remove the DNA, sequence it, and come up with a pretty good picture of what was there and in what proportions. The diversity in soils means that sequencing a sample of more than a gram is overly ambitious, and this in turn means that the DNA

from that gram needs to be amplified before it can be shotgunned. Venter concedes that this can create artefacts unless care is taken, and his Sargasso Sea work has been faulted for sequencing sample contaminants.

The whole truth

The focus of research need not be on single genes. Putting together individual microbial genomes from the jumble of library sequences is an obvious goal — albeit one that is as yet out of reach for samples that contain thousands of species. The most complicated system attempted so far was a community consisting of a handful of microbe species that lived in the extremely acidic waters draining from an old iron mine⁸. Even in this case, the genomes of just two species have been mostly completed, but the work is seen as a major step towards the ultimate goal of determining the ecological role of community members.



"Assembling genomes from the environment is a far more complicated task than the human genome," Venter says. Producing the sequence of a human's 3 billion base pairs required enough shotgunned snippets of DNA to cover the whole genome five to eight times over. "The soil work is like trying to assemble 40,000 human genomes simultaneously with only one-half to one-quarter of the sequencing coverage." Handelsman estimates that getting enough coverage to assemble the genomes of the species in 1 gram of soil would mean sequencing 250 billion base pairs; Venter shotgunned just 1 billion base pairs for the Sargasso Sea paper.

Fertile ground: comparing genetic sequences from the sea floor and soil can uncover key processes.

C. SMITH/UNIVERSITY OF HAWAII

Schleper has constructed three libraries in total, covering 3 billion base pairs of DNA from just two different soil samples, a sandy ecosystem and a mixed forest soil, but hasn't sequenced them. She welcomes help doing the high-throughput work. "The interesting biology is interpreting and looking through it," she notes.

Even with all the sequencing power that money can buy, stitching together genomes of thousands of species is difficult owing to the plentiful horizontal transfer of genes. According to Jonathan Eisen, a genomicist at The Institute for Genome Research in Rockville, Maryland, the largest number of complete genomes so far assembled from a mixed DNA sample is just two: a study he was involved in managed to disentangle the genomes of a *Wolbachia* symbiont and its *Drosophila melanogaster* host. Eisen stresses that even that was not easy.

To Schleper and others, however, the assembly of whole genomes is not necessarily an appropriate goal. Given the enormous amount of genetic exchange among microbes, treating the soil itself as an organism may be more valid. This is the approach used in a recent collaboration between Diversa and the US Department of Energy's Joint Genome Institute at Walnut Creek, California. The researchers put fragmentary sequences from their data into categories or 'bins' based on the metabolism they represent, rather than the gene or organism. They compared the microbial communities in an agricultural soil with samples from sea water and with microbes from ocean-floor sediments recovered from sites where dead whales have decomposed. The comparisons showed that even incomplete DNA could characterize an environment and suggest ecological roles within it.

Whether or not individual genomes are pieced together, there is no doubt that the sequencing of soil genomes will be fertile ground for hypotheses about how microbial communities drive the ecological processes of a region. At the moment, says George Kowalchuk, a microbial ecologist at the Netherlands Institute of Ecology in Heteren, "even if you have the entire community genome, you're still far from predicting how it works." But at least we are heading in the right direction.

"I think this is an area that will have far greater impact on science than sequencing the human genome," says Venter. Not everyone might want to go quite that far, but Schleper is happy to insist: "This is really a new era in microbial ecology."

Virginia Gewin is a freelance science writer based in Portland, Oregon.

Article brought to you by: [Nature](#)

[▲ Top](#)

References

1. Gillespie, D. E. *et al. Appl. Environ. Microbiol.* **68**, 4301–4306 (2002). | [Article](#) | [PubMed](#) | [ISI](#) | [ChemPort](#) |
2. Wang, G. Y. *et al. Org. Lett.* **2**, 2401–2404 (2000). | [Article](#) | [PubMed](#) | [ISI](#) | [ChemPort](#) |
3. D'Costa, V. M. , McGrann, K. M. , Hughes, D. W. & Wright, G. D. *Science* **311**, 374–377 (2006). | [Article](#) | [PubMed](#) | [ChemPort](#) |
4. Sabeji, G. *et al. PLoS Biol.* **3**, e273 (2005). | [PubMed](#) |
5. Quaiser, A. *et al. Environ. Microbiol.* **4**, 603–611 (2002). | [Article](#) | [PubMed](#) | [ISI](#) | [ChemPort](#) |
6. Treusch, A. H. *et al. Environ. Microbiol.* **7**, 1985 -1995 (2005). | [Article](#) | [PubMed](#) | [ISI](#) | [ChemPort](#) |
7. Venter, J. C. *et al. Science* **304**, 66 - 74 (2004). | [Article](#) | [PubMed](#) | [ISI](#) |
8. T

[▲ Top](#)