

## The importance of sample archiving in microbial ecology

*S. Craig Cary and Noah Fierer*

For many microbial ecology studies, the samples collected are irreplaceable — microbial ecologists must therefore develop robust strategies for long-term storage and archiving of samples in order to fully develop, and protect, the scientific record.

For microbial ecologists who are studying marine sediments, the human gut, soils or indeed any environment, the samples that we collect and analyse clearly have immense value. It is surprising therefore how little attention we pay to the long-term archiving of these samples or the nucleic acids extracted from them. We think this is a barrier to progress in microbial ecology. As a research community, we must develop strategies to archive samples and, where possible, make them available to other researchers.

The importance of data archiving is well recognized in microbial ecology, as it allows the scientific community to reanalyse published data and make direct comparisons across studies. Thus, considerable time and effort is devoted to proper data archiving, with nearly every scientific journal and funding agency requiring it in some form. By contrast, there are few, if any, comparable efforts to archive samples from microbial ecology studies, even though they are typically more valuable than the initial data that have been gleaned from them; the data can often be regenerated, but the samples are probably irreplaceable in both time and space. The world is changing rapidly, and the samples that we collect today cannot necessarily be replaced. Failure to effectively archive these samples means that they are truly lost forever.

The utility of preserving biological specimens has been recognized for centuries and the value of specimens housed in herbaria, museums and culture collections is regularly demonstrated in research areas that range from ecology to molecular biology<sup>1</sup>. Thus, what we propose for microbial ecology is not new. However, what is unique is that microbial ecologists can archive whole communities (or at least DNA from those communities) and not just individual organisms or strains. Of course, effective sample archiving is neither cheap nor easy, but there are several crucial arguments for its importance in microbial ecology in general and also for the funding agencies that support this work.

Data generation is cheap and is becoming cheaper. The data that we can now generate are often of higher

quality than the data we could generate only a few months ago. What this means is that reanalysing hundreds of samples using the 'latest and greatest' technologies can often be faster and can yield more robust and informative results than can be gained from trying to piece together pre-existing data sets that were often generated using vastly different methodological approaches. For example, results obtained by characterizing bacterial communities with 16S rRNA gene sequences can be strongly influenced by the choice of PCR primers, the PCR conditions and the DNA extraction methodologies used<sup>2,3</sup>. Therefore, comparing data sets generated with different methods can be compromised by a range of unpredictable biases. Reanalysing samples can be advantageous as it permits direct comparison with other sample sets or allows one to investigate an aspect of a community that was not the focus of the initial analyses (for example, investigating fungal communities in a sample that was initially only analysed for bacteria). The value of reanalysing previously collected samples can be particularly relevant when new technologies become available that can provide more information about the microbial communities found in the samples. For example, longer-read sequencing technologies will prove valuable for assembling whole genomes from shotgun metagenomic data. Such methods cannot be used on samples that are improperly archived but collecting new samples may be infeasible or impossible.

Samples are often more expensive and time-consuming to acquire than they are to sequence and analyse. For example, sample collection from ephemeral deep-sea hydrothermal vents requires the use of both a research vessel and a submersible, just as collecting soil samples from remote areas in Antarctica requires costly logistical support. Even more mundane samples can be valuable as one never knows how they might prove useful in the future. For example, soils collected from before a hurricane or an oil spill could be crucial for documenting how such disturbances affect below-ground microbial communities, just as the analysis of

*S. Craig Cary is at the Department of Biological Sciences, University of Waikato, Private Bag 3105, Hamilton, 3240, New Zealand; Noah Fierer is at the University of Colorado at Boulder, Department of Ecology & Evolutionary Biology, 216 UCB, Boulder, CO 80309-0216, USA. emails: [caryc@waikato.ac.nz](mailto:caryc@waikato.ac.nz); [noah.fierer@colorado.edu](mailto:noah.fierer@colorado.edu) doi:10.1038/nrmicro3382*  
Published online  
3 November 2014

archived soils collected from polar regions will be crucial for documenting the effects of ongoing climate change. Likewise, faecal samples that have been collected from endangered animals or from human populations that have never received antibiotics will be immensely valuable in the future, as will soil samples collected at the start of long-term experimental manipulations. Without archived samples, the effects of ocean acidification, disease, climate change, changes in land use, pollution and other important temporal changes will be far more difficult to assess with any certainty. The dilemma is that when one collects samples, one does not know how or when they might prove useful in the future — if at all. However, it is our contention that quality field samples (or at least the DNA from those samples) should never be discarded. If the results obtained from analysing the samples are of sufficient importance to publish in the scientific literature then the samples should be of sufficient importance to archive.

The practice of archiving samples can mean different things to different people, from keeping a handwritten list of the samples that are stored in one specific freezer to storing well-curated samples in a vault that is capable of surviving a global catastrophe, such as the Svalbard Global Seed Vault. There is already an extensive literature prescribing 'best practices' for archiving biological specimens<sup>4</sup>. Our goal here is not to prescribe a single archiving strategy; every study is different, and an archiving approach that may work with one sample type (or one budget) might not be broadly applicable. Likewise, we are not proposing that all archived samples be made publicly available; this would be impossible for samples comprising limited amounts of material (for example, low-biomass air filter samples) or samples that are subject to legal controls (for example, human faecal samples, which must be handled following strict protocols). Rather, what we are proposing is that researchers carefully consider at the onset of a project how their samples will be archived for extended periods of time, made publicly available if possible, and not lost to science should the laboratory suffer a freezer failure, a man-made or natural disaster or the loss of a crucial laboratory document that contains the sample details. Despite what many of us may wish, no laboratory or scientist exists in perpetuity — we publish and we perish. Scientists graduate, get fired, switch jobs, retire and pass away, and when they do, a lifetime of valuable samples may go with them.

Ideally, it is best to archive not only the nucleic acids but also the whole samples from which they came, in case they are needed for additional analyses. This might not always be feasible; sufficient amounts of sample may not remain after nucleic acid extractions or there could simply be inadequate space for storage. However, storing DNA is the next best thing to storing whole samples, as it is reasonably cheap and the archived DNA could be amplified to produce more if needed. Forensic laboratories archive thousands of samples on a daily basis and there are technologies available that are fairly user-friendly, cheap and that offer the possibility of room-temperature storage of DNA for

decades<sup>5</sup>. Such an approach is currently being used by the Antarctic Genetic Archive in New Zealand, where DNA samples are stored at room temperature at a low per-sample cost and made available to other researchers upon request.

Any discussion of archiving samples for microbial analyses usually quickly shifts to a debate about the best methods to use. Such debates are often based on hearsay and anecdotal information rather than on actual data, and there is clearly a need to quantitatively assess the best approaches for long-term storage of samples. Storage at ultra-low temperatures (typically  $-80^{\circ}\text{C}$  or  $-70^{\circ}\text{C}$ ) is widely considered to be the gold standard, but it is not known whether samples change over time even when stored at such temperatures. Likewise, it is often too expensive and risky to rely on ultra-low temperature storage. Just as the amount of clothing packed for a trip will always match the size of the suitcase used, it is an unwritten rule that all freezers will rapidly become filled to capacity, no matter how many freezers are purchased. Freezers also consume inordinate amounts of energy and are notoriously prone to catastrophic failure during weekends, holidays and other inopportune times. Other methods (freeze-drying, storage at  $-20^{\circ}\text{C}$  and perhaps storage in ethanol) might be effective, but we often do not know how effective they are and there is not likely to be one perfect storage method. The key is to know what storage approach is best for a given sample type, the magnitude of the storage-induced biases and whether those biases are sufficiently small to render the samples useful for future research questions.

Collections of archived samples could be a boon to research in microbiology and microbial ecology, enabling researchers to conduct globally relevant research with previously collected samples, saving the time, expense and environmental impacts associated with redundant collection efforts. We advocate that funding agencies provide funds to cover the costs associated with long-term sample storage and archiving. This is obviously asking a lot given the cash-strapped state of many funding agencies, and although the costs of archiving might be negligible compared with the cost of acquiring the samples, they are not insignificant. However, it is a financially prudent decision. After all, a funding agency that has just spent millions of dollars (or rubles) for a research team to collect samples from a lake in Antarctica that is covered by kilometres of ice probably does not want to see those samples lost in a freezer somewhere, accidentally discarded or unintentionally thawed.

1. Johnson, R. C. Gene banks pay big dividends to agriculture, the environment, and human welfare. *PLoS Biol.* **6**, e148 (2008).
2. Martin-Laurent, F. *et al.* DNA extraction from soils: old bias for new microbial diversity analysis methods. *Appl. Environ. Microbiol.* **67**, 2354–2359 (2001).
3. Wintzingerode, F. V., Göbel, U. B. & Stackebrandt, E. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* **21**, 213–229 (1997).
4. OECD. OECD Best Practice Guidelines for Biological Resource Centres [OECD Publishing, 2012].
5. Ivanova, N. V. & Kuzmina, M. L. Protocols for dry DNA storage and shipment at room temperature. *Mol. Ecol. Resour.* **13**, 890–898 (2013).

#### Competing interests statement

The authors declare no competing interests.

“ it is an unwritten rule that all freezers will rapidly become filled to capacity, no matter how many freezers are purchased ”