



◀ Fig. 1. Images of dinoflagellates that have been identified as *Amphidinium operculatum*. But there are ~150 other described species in the genus; are all these the same species? Barcoding should help to resolve such uncertainties. (a) Illustration for the type description of *A. operculatum* Claparède & Lachmann 1859; (b) graphic image (<http://microscope.mbl.edu>); (c) light micrograph of CCMP strain 123 (<http://microscope.mbl.edu>); (d) light micrograph of CCAP strain 1102/06 (Edmund Nash); (e) non-motile cell (Mona Hoppenrath & Shauna Murray).

Sorting out what we mean by a species, and bringing order to higher level groupings, are important activities for microbial taxonomists.

But **Phil Williamson** and his colleagues argue that the real priorities are more prosaic, yet pragmatic: 'what exactly is out there?' and 'what features should we use to routinely distinguish organisms of different kinds?'

problem for such groups has, however, been somewhat side-stepped; first, by use of the 'candidatus' label for taxa that have yet to be cultured (and hence don't fully comply with the current ICSP Code); and second, by a range of de-replication procedures, including ribotyping (using 16S and 18S rRNA tag sequences) and whole-cell fingerprinting (e.g. based on mass spectrometry). Although ribotyping isn't generally known as barcoding – the shorthand term for the molecular sequencing identification framework for higher plants and animals (www.barcodinglife.org) – it is basically the same thing: characterization using a short and species-specific DNA sequence from a standard position in the genome.

Protist issues

For protists (used broadly, not implying monophyly), at least 200,000 species have been formally described, based on phenotypic features, and several thousand representative strains are maintained in culture. But the evolutionary history of protists is extremely complex, and while phylum-level identities are reasonably clear, relationships between taxa are far from straightforward. Other major problems include the near-impossibility of long-term culture of many parasitic, and highly host-specific, groups (yet these are arguably of greatest economic importance); a lack of congruence of ICBN and

Barcoding is a benchmarking process. On their own, barcodes cannot confer new species status, nor evaluate phylogenetic relationships. Nevertheless, when used with an appropriate bio-informatics database, relatively short DNA sequences (a few hundred base pairs) offer rapid and effective identification with good species separation (unambiguous outcomes for >90% of case studies to date). Costs are currently \$3 per DNA extract, with the analysis process taking a couple of hours. The aim is to achieve at least an order of magnitude reduction in costs and time within a decade, so that bio-barcode analysers can be routinely used in every science laboratory and school, and taken on every field-collection trip.

For animals, the COI (cytochrome c oxidase I) locality is the preferred gene region, while the suitability of a chloroplast gene is currently being investigated for plants. For protists, the

type locality (Norway). If all the *A. operculatum* strains match by barcoding, that will be reassuring. If they don't, some further work – with more detailed characterization – will be necessary, since it is not obvious which strain should retain the *A. operculatum* name and be declared the epitype, as the new standard for future reference. The original (1859) type illustration of *A. operculatum* is shown in Fig. 1, together with some other images of organisms currently considered to be that species.

International co-operation

A co-ordinated, international approach to protist barcoding led by culture collections, and assisted by other molecular and non-molecular approaches, will have other desirable outcomes. In particular, barcodes will provide a robust quality control of what is kept in culture, whilst avoiding unnecessary replication. This is important since collections pro-

culture-based cross-referencing. Such an advance will re-invigorate and potentially revolutionize the study of protists, whilst greatly enhancing the value of relevant culture collections.

Phil Williamson

School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ (e p.williamson@uea.ac.uk)

Robert A. Andersen

Provasoli-Guillard National Center for Culture of Marine Phytoplankton, Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME 04575, USA (e randersen@bigelow.org)

Frithjof C. Küpper

NERC Culture Collection for Algae and Protozoa, Scottish Association for Marine Science, Oban PA37 1QA (e frithjof.kuepper@sams.ac.uk)

Why every protist

Modern systematics isn't stamp-collecting. Even for microbial taxa of relatively limited known diversity, no-one expects to encounter the complete set. Efficient identification systems are, however, essential to link biodiversity and ecosystem services, to help achieve effective disease control and to exploit biotechnological resources. All these applications require unambiguous and standardized naming frameworks, anchoring the name to a real specimen or culture. That standard framework is currently provided by the International Codes for Botanical and Zoological Nomenclature, and the International Committee on Systematics of Prokaryotes (ICBN, ICZN and ICSP). Unfortunately, there are serious mismatches between the operational requirements of these Codes and microbial reality. At the current rate of progress, it will take more than 600 years to properly describe the ~10 million 'known unknowns', and potentially an order of magnitude higher for 'unknown unknowns'.

For prokaryotes (bacteria and archaea), diversity issues are particularly acute. In a single litre of seawater, there can be three times more prokaryotic operational taxonomic units than the global total of ~6,000 recognized species. The naming

ICZN protocols in the protist realm (14% of generic names for plants have also been used for non-plants); and a lack of consensus on suprafamilial systematics (despite claims to the contrary in the November 2006 issue of *Microbiology Today*). Furthermore, most original descriptions for protist species are based on light microscopy and ink drawings, not only making species identification for some groups an inherently subjective and specialist occupation, but also potentially hiding major genetic diversity.

The barcoding solution

To help resolve many of the contradictions and uncertainties in protist taxonomy, genetic barcoding is the way forward, starting with material, particularly type strains, in internationally recognized culture collections. Such an approach was unanimously agreed by 40 protist experts from 12 countries (Australia, Canada, Denmark, France, Germany, Japan, Malaysia, Netherlands, Norway, Russia, UK and USA) at a workshop last November in Portland, Maine, funded by the Sloan Foundation and co-organized by the Culture Center for Marine Phytoplankton (hosted by Bigelow Laboratory for Ocean Science) and the NERC Culture Collection of Algae and Protozoa (Scottish Association for Marine Science).

needs a barcode

Canadian Barcode of Life Network will initially compare COI data for around 1,000 DNA extracts (for ~100 species, ~10 strains of each). If that target does not give good separation, other markers will be tested; alternatives anyway will be necessary for protists lacking mitochondria suitable for rRNA analyses.

Testing the system

Having several, independently collected isolates of what is considered to be the same species is clearly of great value to launch (and test) a protist barcoding initiative. For example, there are currently at least eight strains of the marine dinoflagellate considered to be *Amphidinium operculatum* (and many other strains of '*Amphidinium* sp.') in seven culture collections in five countries. This material has been collected from many locations, including the south-west Pacific, but none from the

vide the reference for species identity; they also serve as patent depositories (under the Budapest Treaty of 1977) and provide model organisms for physiological and biochemical studies. Mis-labelling of cultures due to human error can occur, even in extremely well-managed collections. Without a routine genetic identity check, such mistakes are likely to be perpetuated.

In conclusion

Gene sequencing approaches to taxonomy are not without critics, who have expressed concern that principles are being sacrificed for the sake of expediency. But barcoding is intended to complement, not replace, traditional methods for understanding and classifying whole organisms. For protists, the emphasis is on facilitating identification, confirming in hours what otherwise might take many months of paper-based or

Further reading

- Hebert, P.D.N. & Gregory, T.R. (2005). *Syst Biol* 54, 852–859.
- Jørgensen, M.F., Murray, S. & Daughbjerg, N. (2004). *J Phycol* 40, 351–365.
- Patterson, D.J. (1999). *Am Nat* 154 suppl., S96–S124.
- Pedrós-Alió, C. (2006). *Trends Microbiol* 14, 257–263.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch D.M., Huse, S.M., Neal, P.R., Arrietta, J.M. & Herndl, G.J. (2006). *Proc Natl Acad Sci U S A* 103, 12115–12120.