

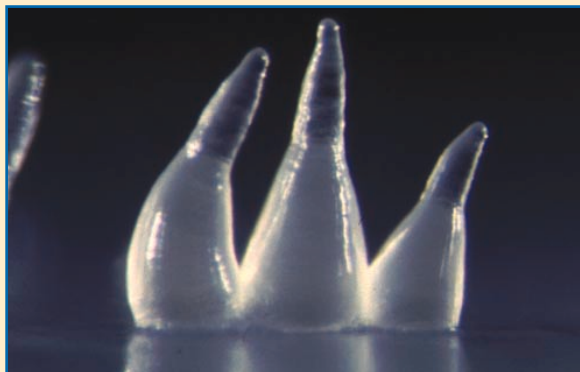
## Genome sequencing

## Stick it in the family album

You may feel that you have some relations that resemble the cellular slime mould *Dictyostelium discoideum*, shown here. If so, you are right. For years there has been debate about where this organism sits on the tree of life, and whether it belongs with plants or animals. But a partial genomic sequence — that of chromosome 2, described by Gernot Glöckner *et al.* in this issue (*Nature* **418**, 79–85; 2002) — confirms its closer kinship with animals.

*Dictyostelium* — *Dicty* to its friends — is a soil amoeba but nonetheless a eukaryote: an organism with a membrane-bound nucleus. It has long been recognized as an excellent model organism. It shows much of the genetic flexibility of yeast, and has complex signalling pathways as well as chemotactic behaviours like those seen, for instance, in white blood cells.

A multinational team of sequencing centres ([www.uni-koeln.de/dictyostelium/consortium.shtml](http://www.uni-koeln.de/dictyostelium/consortium.shtml)) is sequencing the roughly 34 million bases of the *Dicty* genome and mining them for insights into eukaryotic life. But sequencing is dogged by the problem of high A–T base-pair content, and subsequent analysis is also challenging



PIC CREDIT

because of the high repetitive content (around 10%).

To surmount some of the hurdles in sequencing, the various groups constructed libraries from each chromosome separately and carried out shotgun sequencing on each library. But the chromosomal libraries are only about 60% pure, so each group is also generating a whole-genome shotgun. By pooling the data from several libraries, Glöckner *et al.* could fish for the maximum number of pieces of chromosome 2 using known chromosome 2 genes as bait. Tying these pieces together and fitting them onto a backbone of clone-based sequences yielded a mostly complete chromosome sequence, which at 8.1 million bases is about 25% of the genome.

Analysis of chromosome 2 indicates that there are 2,799

protein-coding genes and 73 for transfer RNA. On this basis, the entire genome should contain 10,500 to 11,500 genes, with a density similar to that in budding and fission yeasts. Comparison with sequenced eukaryotic genomes revealed about 45% matches of protein-coding genes, and 55% unique to *Dicty*. Although this number seems high, it is in line with estimates for the available sequences of other eukaryotes.

At the same time, analysis of chromosome 2 has already uncovered more family resemblance, as it encodes proteins similar to several cytoskeleton-related and signalling proteins in animals.

With sequences of the remaining five chromosomes coming shortly, *Dicty* will soon need a page in the family photo album.

Chris Gunter