

Minireview

The *Dictyostelium* genome: the private life of a social model revealed?

Robert Insall

Address: School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK.

E-mail: R.H.Insall@bham.ac.uk

Published: 9 May 2005

Genome Biology 2005, **6**:222 (doi:10.1186/gb-2005-6-6-222)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2005/6/6/222>

© 2005 BioMed Central Ltd

Abstract

The complete genome sequence of *Dictyostelium*, a widely studied social amoeba, reveals unexpected complexities in genome structure, and cell motility and signaling, most notably the presence of a large number of G-protein-coupled receptors not previously found outside animals and the absence of receptor tyrosine kinases.

The social amoeba *Dictyostelium discoideum* is widely studied, in particular because aspects of its lifestyle are especially suitable for experiments that are difficult in other organisms. It has an intriguing way of becoming multicellular, following growth as unicellular amoebae. Starving cells stream together by chemotaxis towards autocrine signals and form aggregates that can contain millions of cells. These differentiate into complex fruiting bodies which somewhat resemble those of fungi. This behavior makes *Dictyostelium* an excellent organism for studying chemotaxis and movement, as well as the cell-cell interactions and differentiation required to make an ordered structure out of a pile of cells. It has also resulted in an unfortunate tendency, seen in a thousand reviews and grant applications, to call *Dictyostelium* a 'simple' model organism. In truth, *Dictyostelium* species are highly adapted and extremely successful, and can be found in almost any soil anywhere on the globe. They eat some organisms (mostly bacteria) and try not to be eaten by others (such as nematodes). There is no room for simplicity in this lifestyle, and the newly published genome sequence [1] reveals an organism that is complex and highly evolved, even if a number of gene families of great importance in multicellular animals and plants are absent.

The *Dictyostelium* genome

This complexity is clear from the finished genome of *D. discoideum*, which contains coding sequence for approximately 12,500 proteins [1]. Yeasts, by comparison, encode only

about 5,500 proteins, and the multicellular (and unarguably complex) *Drosophila melanogaster* only about 13,700. The *Dictyostelium* genes are packed in a compact genome of about 34 megabases (Mb), which is far smaller than the 180-Mb genome of *Drosophila* and a tiny fraction of the sprawling human genome of 2,851 Mb (which still encodes less than twice the number of proteins found in *Dictyostelium*, despite the near 100-fold larger genome).

The relatively large number of genes in *Dictyostelium* was a surprise, albeit one that had been anticipated as genomic studies progressed. Several of the large gene families of multicellular animals are missing, and the number of cell types needed to complete differentiation is a fraction of those required in *Drosophila*. This leads to the question of why *Dictyostelium* contains nearly as many genes as *Drosophila*. Eichinger *et al.* [1] find that as many as 20% of all predicted proteins in the *D. discoideum* genome have appeared relatively recently in its evolutionary history, and in particular that a number of large gene families appear to have been recently duplicated. These families are frequently involved in processes such as motility and signaling, *Dictyostelium*'s particular specialities.

Who is *Dictyostelium*?

Dictyostelium's phylogenetic relationship to multicellular animals has been a contentious issue. Early studies based on

rRNA sequence homology suggested that *Dictyostelium* was an extreme outlier, more closely related to unusual organisms such as the primitive unicellular protist *Giardia* than to animals [2]. To experienced *Dictyostelium* researchers this always seemed improbable as the behavior of *Dictyostelium* closely resembles that of motile mammalian cells such as macrophages, and key proteins (for example the small GTPase ARF1) are almost 100% identical to animal forms. Phylogenetic trees based on protein structure [3,4] suggest that *Dictyostelium* diverged from the animal line at about the same time as plants. Eichinger *et al.* [1] go further, using complete proteome comparisons to establish a clear identity that agrees with earlier protein-based results. In this tree (summarized in Figure 1), *Dictyostelium* diverges from the animal lineage before fungi and yeasts, but after plants. From the point of view of its use as a model organism, the evolutionary distance between *Dictyostelium* and human is actually less than that between human and yeast, because the yeast lineage has experienced a higher rate of evolutionary change. This, again, will not surprise researchers; in a range of processes from motility to lipid signaling, *Dictyostelium* and not *Saccharomyces* appears to be the closer relative of animal cells.

One relationship that will have surprised many in the field is with *Entamoeba*, another motile amoeba whose genome has recently been sequenced [5]. *Entamoeba* is an intestinal parasite of mammals, causing diseases such as amoebic dysentery - an antisocial amoeba to *Dictyostelium*'s social amoeba, perhaps. In keeping with its parasitic lifestyle, *Entamoeba* has some unusual traits. In order to grow, it absolutely requires reducing conditions, such as are found in the large intestine, and it derives its energy from fermentation rather

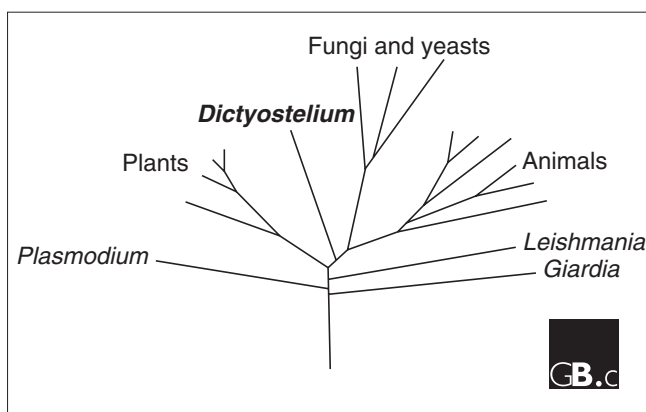


Figure 1
The position of *Dictyostelium* in eukaryotic phylogeny. Whole-proteome comparisons of *Dictyostelium* and representatives of a variety of other groups, rooted on a number of archaeal species, were used to generate this phylogenetic tree (modified from Eichinger *et al.* [1]). *Dictyostelium* diverges from the animal line shortly after the plants and shortly before fungi and yeasts. In many respects *Dictyostelium* is closer to animals than are the fungi, because of the greater rate of divergence of the fungal lineage.

than oxidative metabolism. Consequently, it has no mitochondria (small structures called mitosomes are apparently evolutionary relics) and shares various lifestyle adaptations with pathogens such as *Trichomonas* and *Giardia*, which are phylogenetically extremely distant. Nevertheless, protein-sequence analysis shows that *Entamoeba* and *Dictyostelium* are in fact close cousins [6], suggesting that the loss of mitochondria and oxidative metabolism is evolutionarily recent. This offers great opportunities for using *Dictyostelium* as a tool for understanding amoebiasis and generating new therapies.

Codon and amino-acid bias

Analysis of the genome allows Eichinger *et al.* [1] to make quantitative what 'Dictyologists' have long suspected. First, the AT-richness of *Dictyostelium* DNA is well known. Predicting introns and extragenic sequences is difficult using conventional methods, but this is compensated for by a sharply defined, extreme change from around 70% AT in coding sequences to more than 90% AT elsewhere. The resulting long stretches of poly(AT) also make the cloning of large inserts and PCR difficult, hence the use of whole-chromosome shotgun sequencing to accomplish the *Dictyostelium* genome sequence. Eichinger *et al.* [1] now show that the bias towards AT is so extreme that it biases the choice of amino acids in proteins. Amino acids that are encoded by AT-rich codons (asparagine, lysine, isoleucine, tyrosine and phenylalanine) are commoner in *Dictyostelium* proteins than in other organisms, whereas amino acids encoded by GC-rich codons (proline, alanine, arginine and glycine) are rarer. Similarly, those familiar with *Dictyostelium* know that coding sequences frequently contain bizarre-looking repeats of a single amino acid, most frequently asparagine, similar to the dynamic triplet repeats found in human genes such as the Fragile X locus [7]. The *Dictyostelium* repeats are apparently translated to form poly-asparagine, which makes up a substantial fraction of some proteins. The description of the whole genome allows the large scale of these repeats in *Dictyostelium* to be appreciated: a staggering 34% of predicted proteins contain tracts of 15 residues or more that are composed of only one or two types of amino acids, and 3.3% of all the amino acids specified by the genome are encoded by simple repeats.

Signaling and multicellularity

Dictyostelium's sociability is founded on large-scale and complex signaling between individual cells. Multiple signaling pathways convey the density of bacterial food and the density of cells eating the food, as well as the better-known signals that mediate chemotaxis once cells decide to aggregate, and that set the proportions of differentiated cells in the fruiting body. The genome contains two surprises related to signaling - an unexpectedly large number of G-protein-coupled receptors (GPCRs) is present, but receptor tyrosine kinases (RTKs) are absent.

Earlier work on cyclic AMP signaling identified a family of GPCRs, designated cAR1-cAR4, in *Dictyostelium* [8]. It was also clear that at least two folic-acid receptors are G-protein-coupled [9], and recent work fed by the Japanese *Dictyostelium* cDNA project revealed a small number of additional receptors that resemble cAR1-cAR4 [10]. The complete genome, however, reveals a further 48 putative GPCRs in three families that had not previously been seen outside the animal kingdom. This discovery raises numerous questions. First and foremost, what are all these receptors detecting: interactions with other *Dictyostelium* cells, food location, or identification of other as yet unknown environmental cues? One group of receptors, related to the Frizzled/Smoothed receptors of animals, is usually associated with intercellular signaling, but there are few clues to the roles of the others. The second question is why the additional receptor families are present in *Dictyostelium* but not in yeasts and other fungi. The answer may be that their common ancestor contained at least four families of GPCRs but that the fungal lineage, unlike *Dictyostelium's* ancestors, lost three.

The absence of RTKs is a surprise in the opposite direction. Tyrosine phosphorylation is known to occur in *Dictyostelium*, but the inability of several groups to find RTKs led to a suspicion, now confirmed by the complete genome, that kinases other than RTKs were responsible. This has led to the conclusion that RTK signaling appeared late in evolution, after *Dictyostelium* diverged from the animal line. Other aspects of tyrosine kinase signaling are present, in particular several phosphotyrosine-binding SH2 domains. The real surprise comes from the *Entamoeba* genome. Having identified *Entamoeba* as a close relative of *Dictyostelium*, it was a great surprise to see several RTKs in its genome [5]. The ancestral cells that evolved into *Dictyostelium*, *Entamoeba*, animals and fungi plainly had a diverse range of signaling receptors, which was subject to considerable amplification and loss as species adapted to different niches. One of the key downstream elements of RTK signaling is a pathway based on the small GTPase Ras. *Dictyostelium* contains numerous Ras proteins [11], and the genome predicts a remarkable 25 RasGEFs, the proteins that connect RTK stimulation to activation of Ras in mammalian cells. Clearly, *Dictyostelium* uses some other, as yet entirely unknown, mechanism to connect the outside world to Ras.

Actin-based motility

Dictyostelium has become one of the best models for studying actin-based motility for a number of reasons, including ease and cost of handling, straightforward mutagenesis, and now, of course, the completed genome project. The *Dictyostelium* lifestyle is, in fact, highly focused on motility. Phagocytosis, essential for survival of the amoebae in the wild, is mainly driven by the same set of proteins that drive cell movement [12], while chemotaxis drives both the location

of bacterial food and the process of multicellular aggregation. The genome reflects this specialization: Eichinger *et al.* [1] identify an amazing 71 previously unknown, putative actin-binding proteins, as well as a novel class of actin-related proteins. The systems that regulate actin polymerization are also disproportionately well represented, though surprises remain. Although there are 18 members of the Rho family of small GTPases, Rho itself is missing, as are Rho effector proteins such as ROCK. Most aspects of *Dictyostelium* and mammalian cell movement appear very similar, and myosin II-based contractility (which is important for movement in both cell types) is largely regulated by Rho and ROCK in mammals. It remains to be seen whether a different pathway performs the same job in *Dictyostelium*. Similarly, the Rho family-member Cdc42 is essential for cell polarity in animal and fungal cells, but is not present in the *Dictyostelium* genome. Aggregating *Dictyostelium* are as polar as any mammalian cell, however, and various Cdc42-binding proteins such as the Wiskott-Aldrich syndrome protein (WASP) are present. Presumably one of the other Rho family members - perhaps a Rac such as RacE - substitutes for Cdc42, and *Dictyostelium* may not have subdivided the functions of Rac and Cdc42 in the way that animal cells have done.

Questions like these await coherent, genome-wide studies of the functions of entire gene families, which would have been impossible without a complete genomic sequence. This could be the biggest long-term consequence of the huge collaboration that has enabled the elucidation of the complete genome - knowledge of the entire protein complement of the organism switches the focus away from experiments on single genes, and enables researchers to think in terms of whole processes or complete pathways. Whether or not *Dictyostelium* researchers alter their experimental philosophy, the field will never be the same again.

References

- Eichinger L, Pachebat JA, Glöckner G, Rajandream M-A, Sugang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, *et al.*: **The genome of the social amoeba *Dictyostelium discoideum*.** *Nature* 2005, **435**:43-57.
- McCarroll R, Olsen GJ, Stahl XD, Woese CR, Sogin ML: **Nucleotide sequence of the *Dictyostelium discoideum* small-subunit ribosomal ribonucleic acid inferred from the gene sequence: evolutionary implications.** *Biochem* 1983, **22**:5858-5868.
- Loomis WF, Smith DW: **Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison.** *Proc Natl Acad Sci USA* 1990, **87**:9093-9097.
- Baldauf SL, Doolittle WF: **Origin and evolution of the slime molds (Mycetozoa).** *Proc Natl Acad Sci USA* 1997, **94**:12007-12012.
- Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, *et al.*: **The genome of the protist parasite *Entamoeba histolytica*.** *Nature* 2005, **433**:865-868.
- Bapteste E, Brinkmann H, Lee JA, Moore DV, Sensen CW, Gordon P, Durufle L, Gaasterland T, Lopez P, Muller M, Philippe H: **The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*.** *Proc Natl Acad Sci USA* 2002, **99**:1414-1419.
- Caskey CT, Pizzuti A, Fu YH, Fenwick RG, Jr., Nelson DL: **Triplet repeat mutations in human disease.** *Science* 1992, **256**:784-789.

8. Chen MY, Insall RH, Devreotes PN: **Signaling through chemoattractant receptors in *Dictyostelium***. *Trends Genet* 1996, **12**:52-57.
9. Hadwiger JA, Lee S, Firtel RA: **The G α subunit G α 4 couples to pterin receptors and identifies a signaling pathway that is essential for multicellular development in *Dictyostelium***. *Proc Natl Acad Sci USA* 1994, **91**:10566-10570.
10. Raisley B, Zhang MH, Hereld D, Hadwiger JA: **A cAMP receptor-like G protein-coupled receptor with roles in growth regulation and development**. *Dev Biol* 2004, **265**:433-445.
11. Chubb JR, Insall RH: ***Dictyostelium*: an ideal organism for genetic dissection of Ras signaling networks**. *Biochim Biophys Acta* 2001, **1525**:262-271.
12. Gerisch G, Muller-Taubenberger A: **GFP-fusion proteins as fluorescent reporters to study cell organelle and cytoskeleton dynamics in chemotaxis and phagocytosis**. *Meth Enzymol* 2003, **361**:320-337.