

## The Genome of *Dictyostelium discoideum*

Adam Kuspa and William F. Loomis

### Summary

The *Dictyostelium discoideum* genome has been sequenced, assembled and annotated to a high degree of reliability. The parts-list of proteins and RNA encoded by the six chromosomes can now be accessed and analyzed. One of the initial surprises was the remarkably large number of genes that are shared with plants, animals, and fungi that must have been present in their common progenitor over a billion years ago. The genome encodes a total of about 10,300 proteins including protein families involved in cytoskeletal control, posttranslational protein modification, detoxification, secondary metabolism, cell adhesion, and signal transduction. The genome has a higher proportion of homopolymeric tracts and simple sequence repeats, such as [CAA]<sub>n</sub>, than most other genomes. Triplet repeats in translated regions produce the highest known proportion of polyglutamine tracts in any known proteome. Phylogenetic analyses based on complete proteomes confirm that the amoebozoa are a sister group to the animals and fungi, distinct from plants and early diverging species such as *Leishmania*, *Plasmodium*, or *Giardia*. The completed *Dictyostelium* sequence opens the door to large-scale functional exploration of its genome.

**Key Words:** DNA sequence; proteome; amoebozoa; phylogeny.

### 1. Introduction

The advantages to knowing the complete genome sequence for efficient and productive functional analyses in any organism are becoming increasingly apparent. Model systems for which the genome has not yet been sequenced must present unique attributes to be the subject of continued investigations. The genome defines the information content available to the organism and allows one to predict potential physiological processes. It is the starting point for molecular manipulations to test those predictions.

In 1998, an international consortium began sequencing the *Dictyostelium discoideum* genome and completed the entire sequence early last year (1). High-throughput shotgun sequencing of DNA enriched from the individual chromosomes allowed contigs to be assembled on the basis of overlapping sequences. When combined with high-resolution physical maps, the sequence of each of the six chromosomes could be assembled from one end to the other. The finished sequence conformed well to earlier low-resolution physical maps that defined the chromosomes and provided landmarks along them (2,3). Fewer than 300 gaps remain, and many of these are known to consist entirely of complex repetitive elements such as retrotransposons. Coverage is estimated to include at least 99% of the protein coding information. As one measure of the completeness of the genome sequence, 966 of 967 previously well characterized *Dictyostelium* genes were identified in the assembled sequence. The lengthy and challenging task of sequencing the genome has been worth the effort because it can now be used to characterize the structure of the chromosomes and the genetic information they carry. All the sequence information is publicly available in a convenient and attractive form at dictyBase.org (4).

## 2. The Chromosomes

The nuclear genome of *Dictyostelium* consists of six chromosomes totaling 34 Mb and approx 100 copies of a linear 88-kb palindrome that carries genes for the ribosomal RNAs and no other functional genes (1,5). Each cell has several hundred mitochondria, each of which carries a 57-kb genome that comprises about 30% of the total cellular DNA (6). All of these genetic elements have been sequenced and annotated; however, the major interest lies in the chromosomes.

Early studies on the number of chromosomes in *Dictyostelium* based on Giemsa and Hoechst staining reported that there were seven (7,8). It now turns out that one of the stained structures is an aggregate of the 100 or so copies of the 88-kb palindrome that together hold 9 Mb of DNA, slightly more than any of the individual chromosomes (5). Such aggregates may normally function in the segregation of the palindrome copies at cytokinesis, but they cannot be considered a chromosome. Physical separation of the chromosomes on pulsed-field gels and long-range physical mapping showed that there are only six chromosomes (2,3,9). Moreover, repetitive *Dictyostelium* inverted repeat sequence (DIRS) elements were shown to form six complex clusters that mapped to one end of each of the chromosomes (2,3). Previous cytological evidence had suggested that the *Dictyostelium* chromosomes are telocentric, and *in situ* hybridization with DIRS showed six strongly stained regions, each at the end of a chromosome (1). These DIRS clusters also localize near the nuclear membrane, which is consistent with the behavior of subtelomeric repeats in other organisms (10).

The sequences of these putative centromeric regions presented a major challenge to assembly because of their complex repeated nature. Only the DIRS region of chromosome 1 was completely assembled, as the result of having above-average coverage (**I**). The 187-kb terminal region of chromosome 1 contains 14 complete or near-complete DIRS elements as well as several complete and partial copies of eight other long terminal repeat (LTR), non-LTR, and DNA transposons. It is likely that centromeric functions are encoded within, or near, the DIRS elements. However, there is no functional evidence demonstrating the centromeric function of the DIRS clusters or its neighboring sequences.

Although the DIRS clusters are established features marking one subtelomeric end of each chromosome, the telomeres themselves remain somewhat of a mystery. Previous work had suggested that the chromosomes and palindromic elements terminate in AG<sub>1-8</sub> repeats that could be extended by a telomerase (**2,5,11**). Dense clusters of repeats at the ends of the chromosome assemblies made it impossible to distinguish one from another and determine the sequence to the very end of the chromosomes (**I**). However, there were 12 “floating” contigs with complex repetitive elements on one end and specific short segments of the rDNA element on the other end that could be derived from the actual telomeres. Because there are two ends to each of the six linear chromosomes, there are just enough of these contigs to account for the telomeres. Differences in the repeat elements of some of these contigs allowed them to be physically mapped to the ends of individual chromosomes. Others could be assigned based on the prevalence of their composite reads among the reads from enriched chromosomes. In this way, each putative telomeric contig has been tentatively assigned to a chromosome end (**I**). The presence of a short portion of the palindromic sequence at the distal end of each chromosome raises the possibility that these sequences act as signals for telomere addition to both the rDNA palindrome and the chromosomes.

Although the palindromic rDNA elements are thought to be autonomously replicating mini-chromosomes, they are ultimately encoded by a master copy embedded in chromosome 4 (**5,12**). Further characterization of the master copy locus by the sequencing project revealed the likely junctions between the embedded element and the rest of the chromosome. The locus carries a complete half element and extends past the asymmetric center ending in a G/C-rich sequence that could snap back to form a hairpin primer/template for extrachromosomal replication (**I**). Such a transcription-based replication process could explain the complete absence of sequence variation between the two halves of the element and between the complete elements.

The sizes of the six chromosomes range from 8.5 Mb (chromosome 2) to 3.5 Mb (chromosome 6). Otherwise, they have few differences in gene density,

number of complex repeats, or gaps. There is a perfect inverted repeat of 1.5 Mb on chromosome 2, but it appears to have entered the genome of strain AX4, the one used for sequencing, when its progenitor strain AX3 was isolated 35 yr ago (2,3,13).

A considerable number of duplications encompassing several kilobases of DNA have occurred relatively recently (1). There are 269 pairs of genes encoding nearly identical proteins and 351 other gene families that contain 3–81 members. Most of the genes in the larger families are clustered, with the most similar family members closest to each other in physical distance along the chromosomes. These observations indicate that most duplication occurs in adjacent positions along the chromosomes. Twenty percent of the tRNA genes occur as closely linked pairs with nearly identical sequence, also suggesting a recent wave of duplications.

Each chromosome is studded with simple sequence repeats that can be generated by slippage of the lagging strand during replication and further extended and contracted by unequal crossing over (1). About 10% of the genome consists of quite long homopolymers, as well as repeats of two, three, and six nucleotides. In intergenic regions the A+T content of these repeats is 99.2%, which is much higher than the average base composition for the same regions (85% A+T). In coding regions, the simple sequence repeats consist mainly of triplets that encode polyglutamine, polyasparagine, or polythreonine. There is one or more homopolymer tract in one-third of all predicted protein coding genes. In fact, a higher proportion of the *Dictyostelium* genome encodes polyglutamine and polyasparagine than has been observed in any other sequenced eukaryote.

### 3. Protein-Coding Genes

Several Hidden Markov Model programs designed to recognize protein-coding genes have been trained with manually annotated *Dictyostelium* genes and used to predict protein sequences in the 34-Mb of the *Dictyostelium* genome. Information from each of these automated predictions has been consolidated with the GFMerge program developed by the Pathogen Sequencing Unit at the Sanger Institute and then subjected to manual curation by the team at dictyBase and the rest of the consortium (1). In an effort to include all potentially functional genes, the initial criteria were quite permissive. A total of 13,541 genes were predicted, but 2000 of these encoded proteins of less than 100 amino acids, many of which are unlikely to be functional. Using the simplifying assumption that half were mispredictions, the number of genes was estimated by the consortium to be about 12,500 (1). However, the definition of a gene is a subject of debate. Olsen started with the 13,541 predicted genes and then subtracted genes encoding proteins with less than 50 amino acids (786),

**Table 1**  
**Predicted Protein-Coding Genes of *Dictyostelium discoideum***  
**Compared With Other Organisms**

Feature	Species				
	<i>D. discoideum</i>	<i>S. cerevisiae</i>	<i>A. thaliana</i>	<i>D. melanogaster</i>	<i>H. sapiens</i>
Genome size (Mb)	34	13	125	180	2917
Number of genes	12,500	5538	25,498	13,676	31,400
Gene spacing (kbp/gene)	2.5	2.2	4.9	13.2	132.5
Mean coding size (amino acids)	518	475	437	538	447
% genes with introns	69	5	79	38	85
Mean intron size (bp)	146	ND	170	ND	3365
Mean no. of introns (in spliced genes)	1.9	1.0	5.4	4.0	8.1
Total a.a. encoded (thousands)	7021	2471	11,143	9267	9838

Modified from Eichinger et al. (1).

recently duplicated (nearly identical) genes (355), apparent pseudogenes (1659), and genes from retrotransposons (434) to arrive at 10,307 protein-coding genes (14). As a result of uncertainties in predicting transcriptional signals and protein stability, the total number of genes is likely to be in the range of 10,000–10,600. The complement of predicted protein coding genes identified 99% of the previously characterized genes and sequenced cDNAs (15). Although such measures suggest that the current predicted proteome is nearly complete, continuing manual curation and experimental verification will improve the inventory.

On average, there is a gene in every 2.5 kb of sequence, a gene density similar to that of the yeasts (see Table 1). Compared with most other eukaryotes, *Dictyostelium* genes are smaller and have fewer introns, which are themselves shorter, but encode proteins of about the same average length (see Table 1). The exception is *Saccharomyces cerevisiae*, which has a more compact genome and smaller genes than *Dictyostelium*. Considering that *Dictyostelium* was long thought to be a relatively simple organism, it was surprising to find that it

encodes twice as many proteins as *S. cerevisiae* and almost as many as *Drosophila* (see **Table 1**). The human genome only encodes about twice as many proteins as *Dictyostelium*.

#### 4. The Proteome

The major protein families found in *Dictyostelium*, such as G protein-coupled receptors (GPCRs), protein kinases, and transcription factors, were discussed in the paper presenting the *Dictyostelium* genome and many have recently been further analyzed (**1,16**). The most striking aspect of the proteome is the diversity of protein types among the broad classes of proteins and superfamilies (**17**). For example, *Dictyostelium* has at least one member of each of the major subfamilies of ABC transporters that are found in mammals (**18**). *Dictyostelium* also possesses a large number of Frizzled/smoothed and GABA<sub>B</sub> GPCRs that were previously thought to be specific to metazoa (**19**). One of them, Gr1E, has recently been shown to be sensitive to an antagonist specific to GABA<sub>B</sub> receptor and to be a functional GABA<sub>B</sub> receptor (Anjard and Loomis, submitted).

Global analyses of protein domains in the *Dictyostelium* genome also revealed some interesting insights and surprises. The presence or absence of Pfam domains within eukaryotic proteomes can be determined with increasing resolution as genome sequences accumulate. There are 53 Pfam domains found in *Dictyostelium*, animals, and fungi that are not present in any fully sequenced plant genome (see **Fig. 1**). These domains either arose soon after plants diverged and before *Dictyostelium* diverged from the line leading to animals or they were lost from all plants. The major classes of domains in this group of proteins include those involved in small and large G protein signaling (e.g., regulator of G protein signaling [RGS] proteins), cell cycle control, and domains involved in signaling. It also appears that glycogen storage and utilization arose (or was retained) as a metabolic strategy soon after the plant/animal divergence, because glycogen synthetase appears in this evolutionary interval.

Also particularly striking are the cases in which otherwise ubiquitous Pfam domains appear to be completely absent in one group or another. For instance, *Dictyostelium* appears to have lost the genes that encode collagen domains and basic helix-loop-helix (bHLH) transcription factors. Metazoa, on the other hand, appear to have lost receptor histidine kinases that are common to plants and fungi, whereas *Dictyostelium* has retained 14 of them. The current patterns of gene retention and loss in eukaryotes are likely to change as more genomes are sequenced, and it may turn out that lineage-specific gene loss may be more species-specific than it now appears. For instance, an animal may yet be discovered to have receptor histidine kinases, in the same way that *Ciona* was found to have a cellulose synthase gene similar to that of *Dictyostelium*'s and

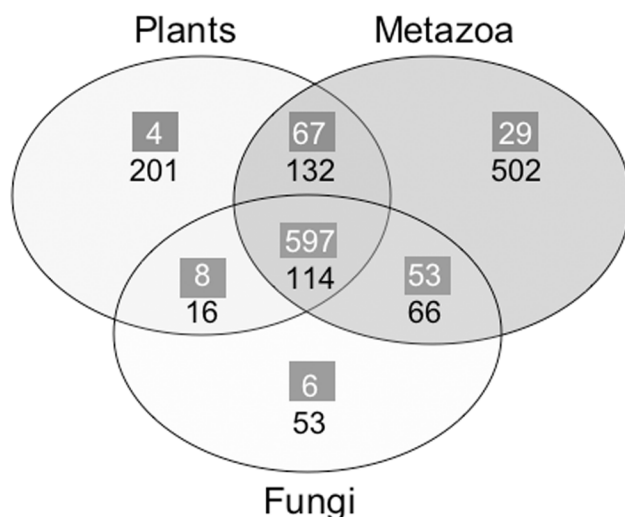


Fig. 1. Distribution of Pfam domains among the eukaryotes. The number of eukaryotic Pfam protein domains present in the major groups of organisms is shown. The numbers of domains present in *Dictyostelium* are boxed. The metazoa are represented by *Homo sapiens*, *Fugu rubripes*, *Caenorhabditis elegans*, *Drosophila melanogaster*; the fungi are represented by *Neurospora crassa*, *Aspergillus nidulans*, *Schizosaccharomyces pombe*, and *Saccharomyces cerevisiae*; and the plants are represented by *Arabidopsis thaliana*, *Oryza. Sativa*, and *Chlamydomonas reinhardtii*. Modified from **ref. 1**, in which a complete description of the analysis and a listing of the domains can be found.

plants were recently found to have “animal-specific” SH2 domain proteins (**1,20**).

Bacterial orthologs can be recognized for at least 1450 genes that must have been in the common ancestor of plants and animals because they exist in at least one proteome within each of the major groups of eukaryotes (Olsen and Loomis, unpublished observations). About one-quarter of these genes are most similar to orthologs found among the archaeobacteria, the likely progenitor of eukaryotes. However, about one-half of these genes are most similar to orthologs found in the proteobacteria that are likely to have entered the eukaryotic genome when a proteobacterium became an established endosymbiont and gave rise to mitochondria. Likewise, about one-fifth of these genes are most similar to orthologs in cyanobacteria that may have been the major food source as eukaryotes took up a predatory life style. The remaining genes appear to have gradually entered the eukaryotic genome from other bacterial types, chiefly the actinobacteria.

There have been more recent cases of gene transfer from bacteria to specific lineages among the eukaryotes. *Dictyostelium* appears to have benefited from horizontal gene transfer (HGT) of genes for such properties as a resistance to tellurite, which is so far unique among eukaryotes (1). Moreover, *Dictyostelium* clearly lost the eukaryotic form of thymidylate synthase and acquired a completely unrelated, rare form of the enzyme found in a few bacteria (21). Predictions of HGT from bacteria to a particular eukaryote suffer from incomplete sampling of eukaryotic genomes. As more genomes are completed, the tests for HGT become more stringent. For instance, 18 genes in the *Dictyostelium* genome were proposed as candidates for HGT (1), but the recent release of the draft genome of *Entamoeba histolytica* showed that one of these was present in their common ancestor and was not recently acquired from a bacterial species (22).

Looking at the protein repertoire of *Dictyostelium* and the other major phylogenetic groups, it becomes clear that the common ancestor had a broad array of proteins and that specific ones were amplified into superfamilies of more specialized proteins in particular lineages. Very few functioning proteins cannot be traced back to a gene that was present when plants and animals shared a common ancestor. Lineage-specific gene loss has turned out to be much more common than was previously supposed, and invoking multiple independent losses to explain the extant phylogenetic pattern is no longer thought to be implausible (23).

## 5. Genome-Based Phylogeny

The phylogeny of *Dictyostelium* has been clarified as more and more genome sequence has accumulated over the past 15 years. Based on sequence comparisons of small ribosomal subunit RNAs (18S), *Dictyostelium* had been thought to be among the disparate group of early diverging eukaryotes that are quite distinct from the crown group of organisms (24). However, Loomis and Smith realized that the unusually high A+T base composition of the *Dictyostelium* genome could easily skew phylogenetic interpretations made from rRNA sequences, and began to compare the available protein sequences (25). These initial protein sequence comparisons told a very different story, and suggested that *Dictyostelium* proteins are actually more similar to mammalian proteins than are the fungal proteins (25,26). This observation was confirmed and extended by an analysis of more than 100 protein sequences, predicted from the genome project, that indicated that the amoebozoa were monophyletic and a sister group to the animals and fungi (27).

Olsen and Loomis extended the phylogenetic protein sequence comparisons of eukaryotes to include thousands of clusters of orthologs from organisms with complete or near completed genome sequences (28). They examined the predicted proteomes of *Dictyostelium* and 22 other eukaryotes and assembled a



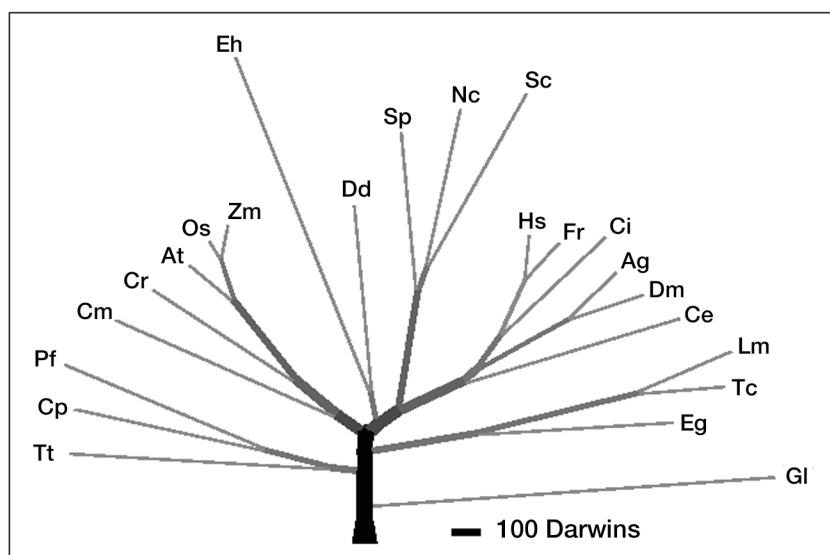


Fig. 2. Proteome based phylogeny of eukaryotes. A phylogenetic tree based on a set of 5908 eukaryotic clusters of orthologs (ECOs) shared by most eukaryotes and rooted on seven archaeobacterial proteomes (1,28). Modified from Song et al. (22). One Darwin is equivalent to 1/2000 the divergence between *Saccharomyces cerevisiae* and *Homo sapiens*. From left to right the organisms shown are: *Tetrahymena thermophila*, *Cryptosporidium parvum*, *Plasmodium falciparum*, *Cyanidioschyzon merolae*, *Chlamydomonas reinhardtii*, *Arabidopsis thaliana*, *Oryza sativa*, *Zea mays*, *Entamoeba histolytica*, *Dictyostelium discoideum*, *Schizosaccharomyces pombe*, *Neurospora crassa*, *Saccharomyces cerevisiae*, *Homo sapiens*, *Fugu rubripes*, *Ciona intestinalis*, *Anopheles gambiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Leishmania major*, *Trypanosoma cruzi*, *Euglena gracilis*, and *Giardia lamblia*.

set of 5908 eukaryotic clusters of orthologs (ECOs) based on a new model of protein sequence divergence (28). From this, they derived a phylogenetic tree of the eukaryotes rooted on a set of seven archaeobacteria, with all of the available completed genomes, that confirms *Dictyostelium's* placement among the amoebzoa and within the crown group (see Fig. 2). It appears that *Dictyostelium* diverged after the plant/animal split, along the line leading to the animals. This tree also explains the higher similarity between *Dictyostelium* and animal proteins relative to fungal proteins. Higher rates of evolutionary change along the fungal lineage have lead to more highly divergent proteins.

ECOs that include members from early diverging organisms as well as at least one crown organism must have been present in the common ancestor of the crown group eukaryotes. Likewise, ECOs that include members from a

plant as well as at least one from the branch leading to animals must have been present in this common ancestor. A total of 2258 such ancestral ECOs were found that could be used to determine patterns of gene loss in specific lineages. The greatest loss of these “ancient genes” occurred in the fungi, for which 40% cannot be recognized in the complete genomes of *Neurospora*, *Schizosaccharomyces*, or *Saccharomyces* (29). *Dictyostelium* has lost 35% of these ancient genes, whereas *Drosophila* has lost only 17% and *Arabidopsis* has lost only 12% (reviewed in ref. 28). The distribution of eukaryotic Pfam domains among eukaryotes revealed a similar pattern of gene retention and gene loss (see Fig. 1 and ref. 1). The plants, metazoa, fungi, and *Dictyostelium* all share 32% of the eukaryotic Pfam domains. Consistent with the phylogeny, Fungi and metazoa share more Pfam domains (119) than do *Dictyostelium* and metazoa (82). Intriguingly, *Dictyostelium* carries a considerable number of Pfam domains that are uniquely found among the metazoa (29). Thus, valuable clues to the functions of proteins containing these domains may come from studies in *Dictyostelium*.

## 6. Comparisons With Another Amoebozoa

Whole-genome comparisons among related species have yielded dramatic insights, as illustrated by studies of yeasts, fruit flies, and mammals (30–34). The amoebozoa lack the morphological traits needed for precise taxonomic categorization, so sequence comparisons are more critical for classification and genome characterization. Previous analysis of 100 genes has clustered *Dictyostelium* and *Entamoeba* as representative genera of the amoebozoa (27). They represent the two major arms of the conosa lineage: the free-living mycetozoa and the amitochondrial archamoeba, respectively.

The *Dictyostelium* genome was the first of the amoebozoa to be completely sequenced and remains the only free-living amoeba sequence available. The genome of the human pathogen *Entamoeba histolytica* has been subjected to deep shotgun sampling and assembly into unordered scaffolds, so most of its coding capacity is known (35). These two genomes have been compared with each other and with other eukaryotic genomes in an effort to identify amoeba specific properties (22). Of the 1500 orthologous gene families shared between the two amobae, most are also shared with plant, animal, and fungal genomes. Surprisingly, only 42 gene families could be defined as distinct to the amoeba lineage. Among the amoeba-specific proteins are a large number that contain repeats of the FNIP domain, the function of which is unknown. The transcription factor CudA was only known previously in *Dictyostelium*, but an ortholog is now known to exist in *Entamoeba* (22,36). The amoebozoa-specific genes may prove to be useful for designing of diagnostics or novel therapies for amoebal pathogens such as *Entamoeba* or *Acanthamoeba*.

The small number of lineage-specific genes indicates an ancient split in the conosa lineage. When *Entamoeba* is included in the phylogenetic analysis of ECOs, the expanded tree indicates that the divergence of these two amoebae is greater than the divergence between the budding and fission yeasts and probably happened shortly after the amoebozoa split from the opisthokont lineage (see **Fig. 2**).

## 7. Prospects for Functional Studies

Achieving a meaningful understanding of a single eukaryotic cell will be an enormous task. Moreover, an understanding of the emergent properties of robustness and evolutionary adaptability inherent in all genomes will necessitate a thorough exploration of the genomic potential of a number of organisms (37,38). History has demonstrated that this will come from the study of relatively simple systems such as *Dictyostelium*, to which powerful technologies can be brought to bear. The *Dictyostelium* genome sequence opens up enormous possibilities for functional studies. Groups from around the world have begun global investigations of gene function through directed knockout strategies and expression profiling of mutants using DNA microarrays (e.g., **ref. 39**). The “molecular anatomy” of *Dictyostelium* is being defined by *in situ* hybridization to establish the temporal and spatial patterns of gene expression throughout development (40–42). Specialists in all areas of eukaryotic biology will be able to enrich the initial interpretations and make useful extrapolations to other species.

Homology comparisons between proteins remains the most reliable and efficient way of deriving functional predictions because they allow information from other species to be integrated and used to make testable hypotheses. Although making functional inferences from data obtained with other species has its limits, the steady accumulation of sequence and functional data offers the possibility of continuous refinement of the predictions. There are a significant number of predicted *Dictyostelium* proteins that have close homologs in other species but whose function in any species remains elusive. For example, there are numerous *Dictyostelium* orthologs to human genes implicated in various diseases that could be fruitfully studied (1). Studies in *Dictyostelium* could provide information on the basic cellular function of these proteins that might be applicable to understanding human pathologies.

Additional, relatively unexplored areas of genome function in *Dictyostelium* remain. For example, the extent to which micro-RNAs (miRNA) regulate expression is an open question. The genome sequence indicates that many of the components needed for miRNA-mediated regulation are present, but bioinformatics analyses of the genome sequence and cDNA databases have so far failed to uncover potential miRNAs on the basis of cross species similarities

(43,44). Novel small noncoding RNAs have been identified, but it is not known whether any of them function as regulatory miRNAs (45).

Determination of the *Dictyostelium* genome sequence has marked a turning point in functional analyses of this organism. Over the last few years, information in the Preliminary Directory of *Dictyostelium* Genes, which was based on the sequences of contigs several years before the complete assembly of the chromosomal sequences, has proven immensely useful to those working with *Dictyostelium*. Genes encoding novel cGMP binding proteins, transcription factors, lipid phosphatases and kinases, histidine kinases, and members of the GPCR and ABC superfamilies were recognized and used in molecular genetic studies that have begun to provide exciting new insights. More such studies can be expected in the years to come. Improvements in data structures for describing biological information will facilitate comparisons between systems. The mechanistic details of a biological process need not be identical in *Dictyostelium* for them to illuminate functions in other species.

## References

1. Eichinger, L., Pachebat, J. A., Glockner, G., et al. (2005) The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**, 43–57.
2. Loomis, W. F., Welker, D., Hughes, J., Maghakian, D., and Kuspa, A. (1995) Integrated maps of the chromosomes in *Dictyostelium discoideum*. *Genetics* **141**, 147–157.
3. Kuspa, A. and Loomis, W. F. (1996) Ordered yeast artificial chromosome clones representing the *Dictyostelium discoideum* genome. *Proc. Natl. Acad. Sci. USA* **93**, 5562–5566.
4. Kreppel, L., Fey, P., Gaudet, P., Just, E., Kibbe, W. A., Chisholm, R. L., and Kimmel, A. R. (2004) dictyBase: a new *Dictyostelium discoideum* genome database. *Nucleic Acids Res.* **32(Database issue)**, D332–D333.
5. Sugang, R., Chen, G., Liu, W., et al. (2003) Sequence and structure of the extra-chromosomal palindrome encoding the ribosomal RNA genes in *Dictyostelium*. *Nucleic Acids Res.* **31**, 2361–2368.
6. Ogawa, S., Yoshino, R., Angata, K., et al. (2000) The mitochondrial DNA of *Dictyostelium discoideum*: complete sequence, gene content and genome organization. *Mol. Gen. Genet.* **263**, 514–519.
7. Robson, G. E. and Williams, K. L. (1977) The mitotic chromosomes of the cellular slime mould *Dictyostelium discoideum*: a karyotype based on Giemsa banding. *J. Gen. Microbiol.* **99**, 191–200.
8. Zada-Hames, I. M. (1977) Analysis of karyotype and ploidy of *Dictyostelium discoideum* using colchicine induced metaphase arrest. *J. Gen. Microbiol.* **99**, 201–208.
9. Cox, E. C., Vocke, C. D., Walter, S., Gregg, K. Y., and Bain, E. S. (1990) Electrophoretic karyotype for *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. USA* **87**, 8247–8251.

10. Louis, E. J. (2002) Are *Drosophila* telomeres an exception or the rule? *Genome Biol* **3**, REVIEWS0007.1–07.6.
11. Emery, H. S. and Weiner, A. M. (1981) An irregular satellite sequence is found at the termini of the linear extrachromosomal rDNA in *Dictyostelium discoideum*. *Cell* **26**, 411–419.
12. Welker, D. L., Hirth, K. P., and Williams, K. L. (1985) Inheritance of extrachromosomal ribosomal DNA during the asexual life cycle of *Dictyostelium discoideum*: examination by use of DNA polymorphisms. *Mol. Cell. Biol.* **5**, 273–280.
13. Loomis, W. F. (1971) Sensitivity of *Dictyostelium discoideum* to nucleic acid analogues. *Exp. Cell Res.* **64**, 484–486.
14. Olsen, R. M. (2005) How many protein coding genes does *Dictyostelium* have, in *Dictyostelium Genomics* (Loomis, W. F. and Kuspa, A., eds.), Horizon Scientific, Norwich, UK: pp. 265–278.
15. Urushihara, H., Morio, T., Saito, T., et al. (2004) Analyses of cDNAs from growth and slug stages of *Dictyostelium discoideum*. *Nucleic Acids Res.* **32**, 1647–1653.
16. Loomis, W. F. and Kuspa, A. (eds.) (2005) *Dictyostelium Genomics*. Horizon Scientific, Norwich, UK.
17. Anjard, C. (2005) Multigene Families of *Dictyostelium*, in *Dictyostelium Genomics* (Loomis, W. F. and Kuspa, A., eds.), Horizon Scientific, Norwich: pp. 59–82.
18. Anjard, C. and Loomis, W. F. (2002) Evolutionary analyses of ABC transporters of *Dictyostelium discoideum*. *Euk. Cell* **1**, 643–652.
19. Hereld, D. (2005) Signal transduction via G-protein-coupled receptors, trimeric G-proteins and RGS proteins, in *Dictyostelium Genomics* (Loomis, W. F. and Kuspa, A., eds.), Horizon Scientific, Norwich: pp. 103–124.
20. Williams, J. G. and Zvelebil, M. (2004) SH2 domains in plants imply new signaling scenarios. *Trends Plant Sci.* **9**, 161–163.
21. Myllykallio, H., Lipowski, G., Leduc, D., Filee, J., Forterre, P., and Liebl, U. (2002) An alternative flavin-dependent mechanism for thymidylate synthesis. *Science* **297**, 105–107.
22. Song, J., Xu, Q., Olsen, R., et al. (2005) Comparing the *Dictyostelium* and *Entamoeba* genomes reveals an ancient split in the Conosa lineage, *PLoS Comput Biol.* **1**(7):e71.
23. Bapteste, E. and Gribaldo, S. (2003) The genome reduction hypothesis and the phylogeny of eukaryotes. *Trends Genet.* **19**, 696–700.
24. McCarroll, R., Olson, G. J., Stahl, X. D., Woese, C. R., and Sogin, M. L. (1983) Nucleotide sequence of the *Dictyostelium discoideum* small-subunit ribosomal ribonucleic acid inferred from the gene sequence: evolutionary implications. *Biochemistry* **22**, 5858–5868.
25. Loomis, W. F. and Smith, D. W. (1990) Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. *Proc. Natl. Acad. Sci. USA* **87**, 9093–9097.
26. Loomis, W. F. and Smith, D. W. (1995) Consensus phylogeny of *Dictyostelium*. *Experientia* **51**, 1110–1115.

27. Bapteste, E., Brinkmann, H., Lee, J. A., et al. (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl. Acad. Sci. USA* **99**, 1414–1419.
28. Olsen, R. and Loomis, W. F. (2005) A collection of amino acid replacement matrices derived from clusters of orthologs *J. Mol. Evol.* **61**(5), 659–665.
29. Kuspa, A. (2005) Whole-genome functional analyses in *Dictyostelium*, in *Dictyostelium Genomics* (Loomis, W. F. and Kuspa, A., eds.), Horizon Scientific, Norwich, UK: pp. 279–296.
30. Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254.
31. Kellis, M., Patterson, N., Birren, B., Berger, B., and Lander, E. S. (2004) Methods in comparative genomics: genome correspondence, gene identification and regulatory motif discovery. *J. Comput. Biol.* **11**, 319–355.
32. Kellis, M., Birren, B. W., and Lander, E. S. (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624.
33. Gibbs, R. A., Weinstock, G. M., Metzker, M. L., et al. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521.
34. Richards, S., Liu, Y., Bettencourt, B. R., et al. (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and cis-element evolution. *Genome Res.* **15**, 1–18.
35. Loftus, B., Anderson, I., Davies, R., et al. (2005) The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865–868.
36. Fukuzawa, M., Hopper, N., and Williams, J. (1997) CudA: a *Dictyostelium* gene with pleiotropic effects on cellular differentiation and slug behaviour. *Development* **124**, 2719–2728.
37. Alm, E. and Arkin, A. P. (2003) Biological networks. *Curr. Opin. Struct. Biol.* **13**, 193–202.
38. McAdams, H. H., Srinivasan, B., and Arkin, A. P. (2004) The evolution of genetic regulatory systems in bacteria. *Nat. Rev. Genet.* **5**, 169–178.
39. Van Driessche, N., Demsar, J., Booth, E. O., et al. (2005) Epistasis analysis with global transcriptional phenotypes. *Nature Genet.* **37**, 471–477.
40. Maeda, M., Kuwayama, H., Yokoyama, M., et al. (2000) Developmental changes in the spatial expression of genes involved in myosin function in *Dictyostelium*. *Dev. Biol.* **223**, 114–119.
41. Tsujioka, M., Yokoyama, M., Nishio, K., et al. (2001) Spatial expression patterns of genes involved in cyclic AMP responses in *Dictyostelium* development. *Devel. Growth Differ.* **43**, 275–283.
42. Maeda, M., Sakamoto, H., Iranfar, N., et al. (2003) Changing patterns of gene expression in *Dictyostelium* prestalk cell subtypes recognized by in situ hybridization with genes from microarray analyses. *Euk. Cell* **2**, 627–637.

43. Martens, H., Novotny, J., Oberstrass, J., Steck, T. L., Postlethwait, P., and Nellen, W. (2002) RNAi in *Dictyostelium*: the role of RNA-directed RNA polymerases and double-stranded RNase. *Mol. Biol. Cell* **13**, 445–453.
44. Graf, S., Borisova, B. E., Nellen, W., Steger, G., and Hammann, C. (2004) A database search for double-strand containing RNAs in *Dictyostelium discoideum*. *Biol. Chem.* **385**, 961–965.
45. Aspegren, A., Hinas, A., Larsson, P., Larsson, A., and Soderbom, F. (2004) Novel noncoding RNAs in *Dictyostelium discoideum* and their expression during development. *Nucleic Acids Res.* **32**, 4646–4656.





<http://www.springer.com/978-1-58829-623-8>

Dictyostelium discoideum Protocols

Eichinger, L. (Ed.)

2006, XVI, 564 p. 186 illus., Hardcover

ISBN: 978-1-58829-623-8

A product of Humana Press