

RESEARCH HIGHLIGHT

Two different genomes that produce the same result

Richard H Kessin*

Abstract

Despite considerable differences in genomic sequence, the developmental program of gene expression between two similar *Dictyostelium* species is remarkably similar.

Have you ever wondered who determined the first DNA sequence? Or how hard it was? Well, I can't say it was the very first, but nearly 45 years ago, George Streisinger and his colleagues mutated the lysozyme gene of phage T4 with acridine, which they knew caused frameshifts, and then they caused a second site suppression (a mutation at a second site that suppressed the effect of the first) with another round of mutagenesis, restoring lysozyme activity. The amino acids encoded by the DNA between the two mutations should, in theory, have been changed - and they were. Knowing the changed amino acids and the genetic code, the group determined the actual DNA sequence. It was 23 nucleotides long and the complete study must have taken five people a year [1]. There was an extra prize, however. The work confirmed that there were no 'commas' between codons. Reading the paper as a graduate student, I thought it was wonderful. And it was.

Now we sequence genomes with such speed that our problem is to make use of the information and not be overwhelmed by it. For the past few decades we have been obsessed with sequences from various organisms and have mastered the art of building phylogenetic trees to reveal distant evolutionary relationships, but comparing the genomes and the transcriptomes of similar organisms can also be revealing. Parikh *et al.* [2] have assembled a team of molecular biologists and informaticians to ask a number of interesting questions about the development of two outwardly very similar species, the slime molds *Dictyostelium discoideum* and *Dictyostelium*

purpureum, now that the sequence of *D. purpureum* as well as that of *D. discoideum* is available (R Sucgang *et al.*, unpublished, (see [3])). The two species are social amoebae, single-celled creatures that live in the soil and eat bacteria until they run out of food. Then they do an extraordinary thing - the amoebae aggregate in groups of 50,000 or so and undergo a synchronous development such that, after 24 hours, they have created a fruiting body composed of a ball of resistant spores on top of a stalk of dead cells: the spores can then be dispersed to a more favorable environment. For movies of these organisms undergoing synchronous aggregation and development go to [4] - it's worth the trouble.

The two species are very similar in appearance and behavior, and the chemoattractant aggregation signal for both species is cyclic AMP (cAMP). *D. purpureum* makes the stalk of the fruiting body a little differently and the spore mass is purple (*D. discoideum* is light yellow) but that is about the extent of the obvious morphological differences. And yet the genome sequences are different - as different, according to Parikh *et al.* [2], as those of humans and bony fishes, despite the fact that *D. discoideum* and *D. purpureum* group within the same clade within the many species of social amoebae, according to phylogenies constructed from ribosomal RNA gene (*rrnA*) sequences [5]. The overall sequence homology of the orthologues is 61.8%. Parikh *et al.* [2] find that the two genomes retain certain gross similarities - both are remarkably AT-rich - but the coding and intergenic sequences have diverged. The questions they then ask are: Do the two species retain the same programs of development despite the differences in genomes? Do the genes necessary to make spores or stalk cells turn on at the same time in each species? How many genes are orthologs; that is, similar by virtue of direct descent from the same ancestral gene? And how many genes are transcribed, and which genes are transcribed the most or the least?

To analyze and compare the transcriptomes of the two species, Parikh *et al.* [2] have abandoned the difficulties of microarray analysis in favor of RNA-sequencing (RNA-seq) [6]. The latter method has a greater dynamic range and cross hybridization is not the problem in

*Correspondence: rhk2@columbia.edu
Department of Pathology and Cell Biology, Columbia University, 630 W. 168th St.
New York, New York, 10032, USA

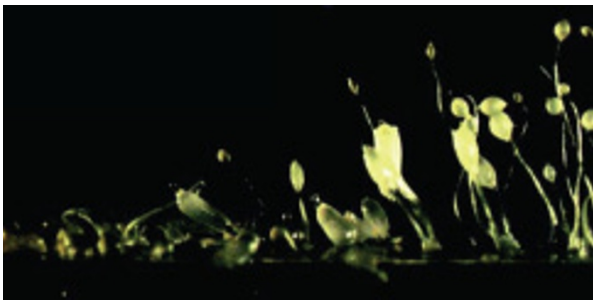


Figure 1. *Dictyostelium discoideum* has a multicellular development, the latter stages of which are shown in this figure. After aggregating by chemotaxis, the cells form a mound, differentiate into two cell types and then, over the next 12 hours, construct a fruiting body consisting of 80,000 viable spores on a stalk created by 20,000 dead stalk cells. *D. purpureum* has a similar development, except for an earlier formation of the stalk and the synthesis of a purple dye in the spore mass. Both species aggregate by chemotaxis toward sources of cAMP. The high synchrony of development makes these experiments possible. Image reproduced from [12].

RNAseq that it is in microarray analysis. Transcripts were collected at 4-hour intervals during the synchronous development of the fruiting body of each species and converted into cDNAs. Fragments of the cDNAs were sequenced in reads of 35 base pairs, and the reads mapped onto the genomes of *D. discoideum* or *D. purpureum*. Any transcript that did not map to a unique sequence was not counted, which will eliminate repetitive elements would be eliminated. This means that actin genes, of which there are a number, would not be counted, nor would the transcripts coding for the mysterious poly-asparagine tracts found in thousands of *Dictyostelium* proteins.

There is interesting data in the transcriptome analysis and the authors provide a nice tool, DictyExpress [7], to explore them, even for those not well versed in computational biology. The important finding is that among the transcripts that are mapped back to the two genomes, there are many orthologs - 7,619 to be exact (out of a predicted total of 12410 genes for *D. purpureum* and 13992 for *D. discoideum*) - and to a great extent they are transcribed in the same groups and in the same temporal order during development in the two species. Almost all genes are regulated during development, either up or down. The synchrony of development and the improved quantitation of RNA-seq (compared with microarrays) make these comparisons possible. Despite the differences in genome sequence, the regulation of developmental gene expression is maintained. Transcripts that are induced during development are coordinated with the slight differences in timing - *D. purpureum* takes 4 hours longer than *D. discoideum* to reach a particular developmental stage, and the appearance of the relevant

transcripts is delayed as well. Many previously characterized genes are regulated almost identically in the two species.

What is the value of this molecular comparative anatomy? Some essential detail is perhaps lost in the statement of Parikh *et al.* [2] likening the difference between *D. discoideum* and *D. purpureum* genomes to the differences between the genomes of bony fish and humans. The differences in sequence between the two slime molds will surely not be spread evenly over the genomes. In structural genes, important functional elements of the protein sequence tend to be conserved, leaving other sequences to diverge. Occasionally, a lack of conservation can be telling - the cell-cell recognition proteins of different species, for example, might be expected to be species-specific and vary in discrete regions [8]. Amazingly, the amoebae of these two species will co-aggregate because of their mutual chemotaxis towards higher levels of cAMP, but they subsequently sort out before forming a fruiting body, as Raper and Thom showed long ago [9].

But there is a long standing problem with *Dictyostelium* development and that concerns the responsible transcription factors - or rather their paucity [10]. It has been known for years that development in *Dictyostelium* is accompanied by shifts in the expression patterns of many genes. In fact, it seems as if the cells switch from expressing one set of genes to expressing another, exactly at the time they switch from being unicellular to being multicellular. Parikh *et al.* [2] now show that the cells alter the abundance of almost every mRNA in the transcriptome during development, so one might expect that transcription factors would be central to the regulation of *Dictyostelium* development, as they are in *Drosophila*, for example. But this may not be the case - *Dictyostelium* researchers have looked for developmental mutants by mutagenesis screens with restriction-enzyme-mediated mutagenesis (REMI), a form of insertional mutagenesis, for the past 18 years, but only a handful of the hundreds of mutants found are in canonical transcription factors. Of such transcription factors, two Mybs, one GATA, two bZIPs, CRTF and a STAT have been found, but a close correlation of any of these with any developmental program or coordinated gene expression in *Dictyostelium* has been elusive (see [4] for the roles of these factors and the phenotypes of their mutants). One exception is *srfa*, a transcriptional regulator similar in sequence to mammalian serum-response factor, whose loss by mutation results in the depression of transcripts involved in spore formation. *D. discoideum* and *D. purpureum* have the lowest known number of transcription factors relative to their genome size [8].

There are a number of possible explanations for these findings. One is that transcription factor genes have been

mutated and associated developmental defects have been observed, but the gene products were not recognized as gene regulatory proteins because they had no homology with known transcription factors. A mutation in the *D. discoideum* G-box binding factor (GBF), for example, blocks post-aggregation development, but it is a non-canonical transcription factor. Another possibility is that the extraordinary conserved temporal expression of many orthologous transcripts in prestalk and prespore cells in the two species could be controlled by some means in addition to traditional transcription factors and recognition sites.

The exceptional AT-richness of promoter regions - 95% in most cases - invites comparison with another organism with a similarly sized AT-rich genome - *Plasmodium falciparum*. In this case too, transcriptional regulation has been difficult to study in detail, although recently a family of AP2 (Apicomplexan apetala2) transcription factors have been shown to be linked to sporozoite specific genes[11]. These have weak homology with plant AP2 factors and, like GBF, bind sequences that have some GC content. Perhaps, with the exception of GBF and a few others, we are just not seeing the *Dictyostelium* transcription factors.

The extraordinary synchrony of development of *Dictyostelium* species and the quantitative advantages of RNA-seq are powerful partners, but such comparisons could be imagined in developing lineages within a particular species, such as different breeds of domesticated animals. How do the neural crest cells that make the snout of a greyhound differ from those of a bulldog? Is it just a few sequences that differ? Or a matter of transcript number? Is the transcript repertory the same but in one case there are more progenitors? These methods might be applied to find out. I am not suggesting

sacrificing puppies (perhaps fish would be better subjects), but it is the kind of thing that Darwin would have liked to know.

Published: 27 April 2010

References

1. Terzaghi E, Okada Y, Streisinger G, Emrich E, Inouye M, Tsugita A: **Change of a sequence of amino acids in phage T4 lysozyme by acridine-induced mutations.** *Proc Natl Acad Sci USA* 1966, **56**:500-507.
2. Parikh A, Miranda ER, Katoh-Kurasawa M, Fuller D, Rot G, Zagar L, Curk T, Sucgang R, Chen R, Zupan B, Loomis WF, Kuspa A, Shaulsky G: **Conserved developmental transcriptomes in evolutionarily divergent species.** *Genome Biol* 2010, **11**:R35.
3. **Dictyostelium purpureum welcome page** [<http://genomes.dictybase.org/purpureum>]
4. **dictyBase** [<http://www.dictybase.org>]
5. Schaap P, Winckler T, Nelson M, Alvarez-Curto E, Elgie B, Hagiwara H, Cavender J, Milano-Curto A, Rozen DE, Dingermann T, Mutzel R, Baldauf SL: **Molecular phylogeny and evolution of morphology in the social amoebae.** *Science* 2006, **314**:661-663.
6. Shendure J: **The beginning of the end for microarrays?** *Nat Methods* 2008, **5**:585-587.
7. **DictyExpress** [<http://www.aillab.si/dictyexpress>]
8. Benabentos R, Hirose S, Sucgang R, Curk T, Katoh M, Ostrowski EA, Strassmann JE, Queller DC, Zupan B, Shaulsky G, Kuspa A: **Polymorphic members of the lag gene family mediate kin discrimination in *Dictyostelium*.** *Curr Biol* 2009, **19**:567-572.
9. Raper KB, Thom C: **Interspecies mixtures of the Dictyosteliaceae.** *Am J Bot* 1941, **28**:69-78.
10. Shaulsky G, Huang E: **Components of the *Dictyostelium* gene expression regulatory machinery.** In *Dictyostelium Genomics*, edited by Loomis WF, Kuspa A. Norwich, UK: Horizon Scientific Press; 2005: 83-102.
11. Lindner SE, De Silva EK, Keck JL, Llinás M: **Structural determinants of DNA binding by a *P. falciparum* ApiAP2 transcriptional regulator.** *J Mol Biol* 2010, **395**:558-567.
12. Ostrowski EA, Shaulsky G: **Learning to get along despite struggling to get by.** *Genome Biol* 2010, **10**:218

doi:10.1186/gb-2010-11-4-114

Cite this article as: Kessin RH: Two different genomes that produce the same result. *Genome Biology* 2010, **11**:114.