Institute of Biochemistry I, Medical Faculty, Center for Molecular Medicine Cologne (CMMC), Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Germany

Dictyostelium genomics: how it developed and what we have learned from it

A. A. NOEGEL, G. GLÖCKNER

Received November 26, 2012, accepted December 5, 2012

Prof. Dr. Angelika A. Noegel, Institute of Biochemistry I, Medical Faculty, University of Cologne, Joseph-Stelzmann-Straße 52, 50931 Köln, Germany
noegel@uni-koeln.de

Dedicated to Prof. Dr. Theo Dingermann, Frankfurt, on the occasion of his 65th birthday.

Pharmazie 68: 474–477 (2013) *doi: 10.1691/ph.2013.6516*

Dictyostelium *discoideum* is the most prominent member of the social amoebae. It has been used as an experimental system since more than 50 years and a large number of scientists worldwide work on different aspects such as chemotaxis, cytoskeleton, differentiation and development. *Dictyostelium* shares more features with animals than fungi although it diverged much earlier in evolution. Many of the results obtained with *D. discoideum* can therefore be transferred to animals making *D. discoideum* a valuable model organism. Targeted gene inactivation using homologous recombination is easy and mutant phenotypes can be readily isolated due to the haploid nature of its genome. Furthermore, a variety of techniques and tools are available that facilitate the experimental work; its genome and that of several *Dictyostelidae* has been sequenced and most recently a high-resolution genome wide nucleosome map for *D. discoideum* has been generated.

1. Introduction

Dictyostelium discoideum is an excellent model organism in which signal transduction, cell migration, cell differentiation, generation of multicellularity, pattern formation and development can be easily studied. It exists as single celled amoeba and feeds on bacteria and fungi (Fig. 1). Under starvation conditions cells come together in response to a cAMP signal forming a multicellular aggregate in which cells differentiate and enter a developmental cycle that ends with the formation of a fruiting body consisting of a head carrying spores supported by a stalk. Under appropriate conditions the spores germinate and progress through the life cycle again. *D. discoideum* can be grown in large quantities in axenic media which makes biochemical studies easy. Also, all developmental stages are equally accessible for biochemical and other experiments. Various molecular techniques are available: mutants can be generated through homologous recombination or RNA silencing; random mutagenesis is achieved by REMI (Restriction Enzyme Mediated Integration). Mutants can conveniently be isolated since *D. discoideum* is haploid and phenotypes are not masked by functional alleles.

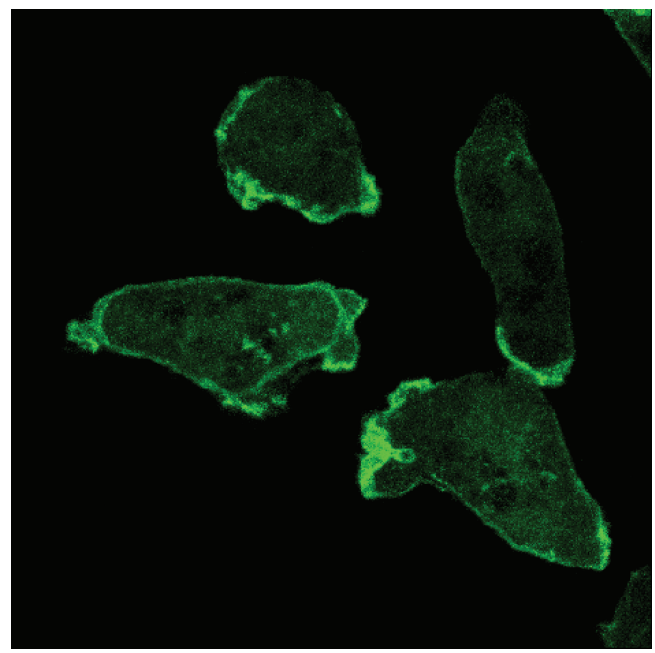

Fig. 1: *D. discoideum* amoebae immunolabelled for CAP (cyclase associated protein). The protein is distributed throughout the cytosol and is enriched near the plasma membrane, in particular in zones where there is high F-actin dynamics (kindly provided by Maria Stumpf and Rosemarie Blau-Wasser)

2. Analysis of the *D. discoideum* genome

The *Dictyostelium* genome is 35 Mb in size and is carried on six chromosomes varying in size from 4 to 7 Mb. In 1998, a DFG funded genome sequencing project was initiated that led to the sequence determination of the largest chromosome C2 (Glöckner et al. 2002). Subsequently, an international consortium was established with groups from the UK, US and Japan in which we determined and analyzed the sequence of the full genome (Eichinger et al. 2005). One of the first publications resulting from the efforts to sequence the genome came out in 1999 and was co-authored by Theo Dingermann (Table 1) (Szafranski

Table 1: Milestones in *Dictyostelium* genomics

Milestones	Year of publication
Identification of integration sites of transposons in *Dictyostelium discoideum*	Szafranski et al. (1999)
Transposon content in *Dictyostelium discoideum*	Glöckner et al. (2001)
Sequence and analysis of chromosome 2	Glöckner et al. (2002)
Sequence and analysis of the *Dictyostelium discoideum* genome	Eichinger et al. (2005)
Sequence and analysis of the *Polysphondylium pallidum* and *Dictyostelium fasciculatum* genomes	Heidel et al. (2011)
Sequence and analysis of the *Dictyostelium purpureum* genome	Sucgang et al. (2011)
The nucleosome map of the *Dictyostelium discoideum* genome	Chang et al. (2012)

et al. 1999). Theo as a member of the Sequencing Consortium (Glöckner et al. 2002) was closely associated with this scientific endeavor and supported it, in particular when it came to the analysis of the various DNA repeat elements which represent ∼10% of the *D. discoideum* genome, and in case of non-LTR retrotransposons are associated with tRNAs, a topic which Theo has studied quite extensively (Marschalek et al. 1989; Szafranski et al. 1999; Glöckner et al. 2001; Winkler et al. 2002).

The sequence has revealed a remarkably high total number of app. 13,400 genes, which is close to the one encoded by the 180-Mb *Drosophila* genome, arguing that the seemingly simple organism has a complexity nearly as high as a true multicellular organism. The high number of genes is achieved by close spacing of the genes and short 5' and 3' untranslated sequences. Introns are present, however, the number per gene is low, moreover, they are rather short (mean size ∼146 bp). The data also confirmed an evolutionary origin for *Dictyostelium* that is closer to animals and fungi than to plants. Whole genome analysis showed that *Dictyostelium* diverged early along the branch leading to the Metazoa. However, due to the rapid divergence along the line leading to the fungi, *Dictyostelium* proteins are more similar to human orthologs than are those of *Saccharomyces cerevisiae* (Eichinger et al. 2005).

The sequence analysis revealed an unusual telomere and centromere organization in *D. discoideum*. Normally, telomeres are generated by the enzyme telomerase which adds a short nucleotide sequence in a highly repetitive manner to the ends of the chromosome. In the *D. discoideum* genome we did not identify such repetitive sequences. Instead, telomeres are formed by sequences derived from the so-called rDNA palindrome. This palindrome is an extrachromosomal element which carries the rDNA genes and is highly amplified to approximately 100 copies per cell in *D. discoideum*. Centromeres are organized as transposon derived extensive regions at the tip of each chromosome (Glöckner and Heidel 2009).

A remarkable finding was the identification of a surprisingly large number of G protein coupled receptors (GPCRs). GPCRs are characterized by seven transmembrane regions that allow detection and transduction of a large variety of extracellular signals. They are subdivided into six families, and prior to our genome analysis only members of one family, the CAR/CRL (cAMP receptor/cAMP receptor-like) family, have been known. The genome sequence revealed the presence of numerous and surprisingly diverse additional GPCRs. There are 61 such receptors in the genome corresponding to a significant fraction of the *Dictyostelium* proteome (∼0.5% of the encoded genes), a proportion lying between that of fungi (e.g. *N. crassa*: 0.1%) and animals (e.g. *Drosophila*: ∼2%; vertebrates: 2–4%) (Eichinger et al. 2005; Heidel et al. 2011).

Detailed analysis revealed that the family of GPCRs showed an unexpected complexity in *D. discoideum* and the presence of members of all major GPCR families except family 1 (β-adrenergic, light and odorant receptors) and family 4 (pheromone receptors). In addition to members of the secretin (family 2) and metabotropic glutamate/$GABA_B$ (family 3) families the *Dictyostelium* genome also encodes 25 potential frizzled/smoothened (family 5) receptors. Frizzled receptors have roles in cell proliferation, cell polarity, gastrulation and tissue formation and were thought to be an innovation of multicellular metazoa. Their presence in *Dictyostelium* was therefore unexpected (Prabhu and Eichinger 2006).

For transcription factors, we observed a different situation and detected only a limited number. Moreover, the class of bHLH transcription factors was entirely missing from the *D. discoideum* genome, whereas SH2 domain-containing STATs had been already previously identified (Kawata et al. 1997). SH2 domains had not been found in non-metazoans and *Dictyostelium* is the only non-metazoan organism wherein SH2 domain-phosphotyrosine signalling has been proven to occur.

3. Genome analysis of Dictyostelids

The knowledge of the genome sequence facilitated and facilitates all aspects of subsequent work on this organism and also affects research in other species. To allow a deeper understanding of eukaryotic evolution and exploit the genome sequence in depth, comparisons to genomes of other social amoebae are needed. Therefore, in a next step we sequenced the genomes of further members of the Dictyostelidae.

The social amoebae fall into four major groups according to a small subunit ribosomal RNA (SSU rRNA) phylogeny and the signaling molecules used for entering the developmental cycle (Schaap et al. 2006). An initial rough characterization of several species of all groups using a whole genome survey sequencing approach showed that all species had approximately the same genome sizes, albeit with differing nucleotide composition biases. In a DFG-funded project we sequenced and analyzed the *D. fasciculatum* (*DF*) and *P. pallidum* (*PP*) genome. Their genome sizes are 31 and 33 Mb, respectively, the AT content is 66% for *DF* and 68% for *PP*. In contrast to *D. discoideum* (*DD*), *PP* and *DF* have conventional telomere organization built from hexameric repeats and with > 1% strongly reduced numbers of transposable elements as compared to the 10% in *DD*. The number of protein coding genes is similar between species. With all presently available data from genome analysis we could construct a phylogenetic tree of Amoebozoa which places *PP* and *DF* at the base of the Dictyostelidae and shows that *DD* is the most derived strain (Glöckner and Noegel 2012) (Fig. 2).

4. A nucleosome map for *D. discoideum*

The eukaryotic genome is packaged into chromatin, a DNA protein complex with the nucleosome as basic unit. The organizing principle of the nucleosome is a histone octamer with ∼147 base pairs of DNA wrapped around. The nucleosomes are spaced by linker DNA that ranges from 10 to 50 bp in length. This structure affects all stages of transcription. In general, promoter regions

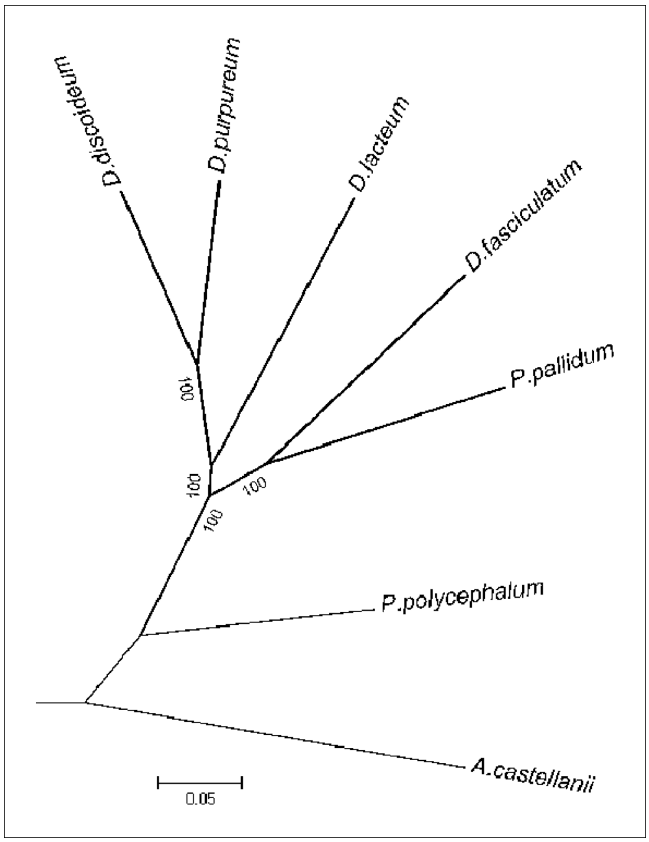

Fig. 2: Phylogeny of Amoebozoa with sequenced genomes based on a concatenated data set of 30 genes. *P. polycephalum*, *Physarum polycephalum*, *A. castellanii, Acanthamoeba castellanii*

and other transcriptional regulatory sites are nucleosome-free, whereas genic regions are occupied with arrays of well-phased nucleosomes.

The multiple interactions between histones and DNA generate a stable protein-DNA complex under physiological conditions. Construction of genomic maps of nucleosomes showed however a rapid unwinding of nucleosomes and eviction of histones indicating that nucleosomes are not static but have dynamic properties that are regulated by various protein complexes (Giresi et al. 2006; Li et al. 2007). Most important for the dynamic properties are the post translational modifications of histones which affect their tails and globular domains and loosen inter- or intra-nucleosomal DNA-histone interactions (Kouzarides 2007).

Data from yeast indicated that the primary information for positioning nucleosomes is embedded within the sequence of the genome (Segal et al. 2006; Lee et al. 2007). Acting on this are chromatin remodeling complexes that use ATP-hydrolysis to alter histone-DNA contacts and lead to a transient unwrapping of the end DNA from the octamer or allow movement of the nucleosome (Li et al. 2007). Together, these components and events regulate transcription, replication, DNA repair and epigenetic processes. Identifying nucleosome positions *in vivo* is therefore important for predicting how genes are regulated. Methods for the identification have evolved rapidly during the last couple of years.

The primary tool to identify genome-wide nucleosome positions is micrococcal nuclease (MNase). MNase cleaves between nucleosomes and preferentially digests linker DNA before nucleosome protected DNA. The nucleosomal DNA fragments can then be sequenced and mapped on the genomic DNA.

Genome wide maps of nucleosome locations are now available for a variety of species. We teamed up with the Pugh and Schuster lab at Penn State and determined the nucleosome positions in *D. discoideum* which, due to the A/T-richness of the genome, was particularly exciting. The analysis revealed that the nucleosomes were relatively enriched with G/C near their midpoints and that the nucleosome border regions were distinguished by homo- and heteropolymeric tracts of A and T, emphasizing the notion that the nucleosome positioning is shaped by the underlying DNA sequence. Furthermore, the *Dictyostelium* chromatin appears to be organized in di-nucleosome units, which is so far the only example for such an organizational principle. Apart from this, the nucleosomes are positioned as in most multicellular organisms. Around the transcription start sites the + 1 nucleosome is not placed over the transcription start sites, as in unicellular fungi, but is located downstream, as in multicellular animals and plants. This predicts the presence of a paused RNA polymerase II. In agreement, genes coding for components of the NELF pausing factor, a complex of proteins that is responsible for pausing RNA, have been found. Furthermore, the overall nucleosome organization in the vegetative and the aggregation stage did not differ which is in agreement with data from other organisms. However, a modal shift in the spacing between genic nucleosomes in the two states was observed (Chang et al. 2012). A similar increase in spacing was seen in yeast during a starvation response (Zhang et al. 2011) and in human cells at quiescent/repressed regions of the genome (Vaillant et al. 2010; Valouev et al. 2011).

5. Conclusions

Sequence and analysis of the *Dictyostelium* genome has opened new avenues for research. It has given us an unprecedented chance to quickly search the genome for proteins and sequences at dictybase (http://dictybase.org/index.html) since the year 2002. This had a tremendous impact on the research of the individual scientists in the field. It also initiated further sequencing efforts that were highlighted recently (Glöckner and Noegel 2012) and led to new research areas such as the establishment of a nucleosome map for the *D. discoideum* genome. In the future we will see many more results coming out of these efforts.

References

Chang GS, Noegel AA, Mavrich TN, Müller R, Tomsho L, Ward E, Felder M, Jiang C, Eichinger L, Glöckner G, Schuster SC, Pugh BF (2012) Unusual combinatorial involvement of poly-A/T tracts in organizing genes and chromatin in *Dictyostelium*. Genome Res 22:1098–1106.

Eichinger L, Pachebat JA, Glöckner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hemphill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Goodhead I, Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Madan Babu M, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Loulseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabbinowitsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G, Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A (2005) The genome of the social amoeba *Dictyostelium discoideum*. Nature 435:43–57.

Giresi PG, Gupta M, Lieb JD (2006) Regulation of nucleosome stability as a mediator of chromatin function. Curr Opin Genet Dev 16:171–178.

Glöckner G, Szafranski K, Winckler T, Dingermann T, Quail MA, Cox E, Eichinger L, Noegel AA, Rosenthal A (2001) The complex repeats of *Dictyostelium discoideum*. Genome Res 11:585–594.

Glöckner G, Eichinger L, Szafranski K, Pachebat JA, Bankier AT, Dear PH, Lehmann R, Baumgart C, Parra G, Abril JF, Guigó R, Kumpf K, Tunggal B, Cox E, Quail MA, Platzer M, Rosenthal A, Noegel AA; Dictyostelium Genome Sequencing Consortium (2002) Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. Nature 418:79–85.

Glöckner G, Heidel A (2009) Centromere sequence and dynamics in *Dictyostelium discoideum*. Nucl Acids Res 37:1809–1816.

Glöckner G, Noegel AA. Comparative genomics in the Amoebozoa clade. (2012) Biol Rev Camb Philos Soc 88:215–225.

Heidel AJ, Lawal HM, Felder M, Schilde C, Helps NR, Tunggal B, Rivero F, John U, Schleicher M, Eichinger L, Platzer M, Noegel AA, Schaap P, Glöckner G (2011) Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. Genome Res 21:1882–1891.

Kawata T, Shevchenko A, Fukuzawa M, Jermyn KA, Totty NF, Zhukovskaya NV, Sterling AE, Mann M, Williams JG (1997) SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in *Dictyostelium*. Cell 89:909–916.

Kouzarides T (2007) Chromatin modifications and their function. Cell 128:693–705.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C (2007) A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet 39:1235–1244.

Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128:707–719.

Marschalek R, Brechner T, Amon-Böhm E, Dingermann T (1989) Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. Science 244:1493–1496.

Prabhu Y, Eichinger L (2006) The *Dictyostelium* repertoire of seven transmembrane domain receptors. Eur J Cell Biol 85:937–946.

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 (2006) Molecular phylogeny and evolution of morphology in the social amoebas. Science 314:661–663.

Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, Wang JP, Widom J (2006) A genomic code for nucleosome positioning. Nature 442:772–778.

Sucgang R, Kuo A, Tian X, Salerno W, Parikh A, Feasley CL, Dalin E, Tu H, Huang E, Barry K, Lindquist E, Shapiro H, Bruce D, Schmutz J, Salamov A, Fey P, Gaudet P, Anjard C, Babu MM, Basu S, Bushmanova Y, van der Wel H, Katoh-Kurasawa M, Dinh C, Coutinho PM, Saito T, Elias M, Schaap P, Kay RR, Henrissat B, Eichinger L, Rivero F, Putnam NH, West CM, Loomis WF, Chisholm RL, Shaulsky G, Strassmann JE, Queller DC, Kuspa A, Grigoriev IV (2011) Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. Genome Biol 12:R20.

Szafranski K, Glöckner G, Dingermann T, Dannat K, Noegel AA, Eichinger L, Rosenthal A, Winckler T (1999) Non-LTR retrotransposons with unique integration preferences downstream of *Dictyostelium discoideum* tRNA genes. Mol Gen Genet 262:772–780.

Vaillant C, Palmeira L, Chevereau G, Audit B, d'Aubenton-Carafa Y, Thermes C, Arneodo A (2010) A novel strategy of transcription regulation by intragenic nucleosome ordering. Genome Res 20:59–67.

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A (2011) Determinants of nucleosome organization in primary human cells. Nature 474:516–520.

Winckler T, Dingermann T, Glöckner G (2002) *Dictyostelium* mobile elements: strategies to amplify in a compact genome. Cell Mol Life Sci 59:2097–2111.

Zhang L, Ma H, Pugh BF (2011) Stable and dynamic nucleosome states during a meiotic developmental process. Genome Res 21:875–884.