Eukaryotic Cell

Conserved Gene Regulatory Function of the Carboxy-Terminal Domain of Dictyostelid C-Module-Binding Factor

Anika Schmith, Marco Groth, Josephine Ratka, Sara Gatz, Thomas Spaller, Oliver Siol, Gernot Glöckner and Thomas Winckler

Eukaryotic Cell **2013, 12(3):460. DOI: 10.1128/EC.00329-12. Published Ahead of Print 25 January 2013.**

EC
Journals.ASM.org

Conserved Gene Regulatory Function of the Carboxy-Terminal Domain of Dictyostelid C-Module-Binding Factor

Anika Schmith,[a] Marco Groth,[b] Josephine Ratka,[a] Sara Gatz,[a] Thomas Spaller,[a] Oliver Siol,[a,c] Gernot Glöckner,[d,e] Thomas Winckler[a]

Department of Pharmaceutical Biology, Institute of Pharmacy, University of Jena, Jena, Germany[a]; Genome Analysis, Leibniz Institute for Age Research—Fritz Lipmann Institute, Jena, Germany[b]; Institut de Génétique Humaine, CNRS, UPR 1142, Montpellier, France[c]; Institute for Biochemistry I, Medical Faculty, University of Cologne, Cologne, Germany[d]; Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany[e]

C-module-binding factor A (CbfA) is a jumonji-type transcription regulator that is important for maintaining the expression and mobility of the retrotransposable element TRE5-A in the social amoeba *Dictyostelium discoideum*. CbfA-deficient cells have lost TRE5-A retrotransposition, are impaired in the ability to feed on bacteria, and do not enter multicellular development because of a block in cell aggregation. In this study, we performed Illumina RNA-seq of growing CbfA mutant cells to obtain a list of CbfA-regulated genes. We demonstrate that the carboxy-terminal domain of CbfA alone is sufficient to mediate most CbfA-dependent gene expression. The carboxy-terminal domain of CbfA from the distantly related social amoeba *Polysphondylium pallidum* restored the expression of CbfA-dependent genes in the *D. discoideum* CbfA mutant, indicating a deep conservation in the gene regulatory function of this domain in the dictyostelid clade. The CbfA-like protein CbfB displays ~25% sequence identity with CbfA in the amino-terminal region, which contains a JmjC domain and two zinc finger regions and is thought to mediate chromatin-remodeling activity. In contrast to CbfA proteins, where the carboxy-terminal domains are strictly conserved in all dictyostelids, CbfB proteins have completely unrelated carboxy-terminal domains. Outside the dictyostelid clade, CbfA-like proteins with the CbfA-archetypical JmjC/zinc finger arrangement and individual carboxy-terminal domains are prominent in filamentous fungi but are not found in yeasts, plants, and metazoans. Our data suggest that two functional regions of the CbfA-like proteins evolved at different rates to allow the occurrence of species-specific adaptation processes during genome evolution.

Cells of the social amoebae (dictyostelids) live in the soil and feed on bacteria. Under unfavorable conditions, dictyostelid cells can aggregate and form multicellular fruiting bodies that hold dormant spores to ensure the survival of the species (1, 2). Insight into the genome structures of related dictyostelids has revealed an unexpected degree of genome flexibility and genetic diversity, suggesting that the study of these organisms may be valuable in understanding the evolutionary forces that drive genomic adaptation processes (3–5).

The model species *Dictyostelium discoideum* has an unusual genome, where only ~22% of the nucleotides are GC and 65% of the DNA codes for proteins (3). Compared with other dictyostelid genomes that present similar gene-dense environments (4), the *D. discoideum* genome harbors a surprisingly high percentage of mobile genetic elements (6). Transposon activity can be disastrous for the host if it leads to insertional mutagenesis, nonallelic homologous recombination, or induction of chromosome breaks (7, 8). Therefore, the activity of these parasitic elements must be carefully controlled by the host to maintain genome stability. *D. discoideum* has emerged as an excellent model to study the interactions between transposable elements and compact host genomes (9).

The retrotransposon TRE5-A is characterized by three outstanding features: an active population of elements in the *D. discoideum* genome is maintained (10), gene disruption rarely occurs because integration is targeted to the vicinity of tRNA genes (11), and a considerable amount of minus strand ("antisense") RNA is produced from a promoter located at the 3′ end of the element, the C module (12, 13). The large amount of minus strand RNA in growing *D. discoideum* cells suggests that TRE5-A retrotransposition may be regulated by posttranscriptional silencing. The C-module-binding factor A (CbfA) was discovered in a search for cellular factors that bind to the C module *in vitro* and may modulate TRE5-A amplification by regulating the TRE5-A minus strand of RNA (14).

Attempts to delete CbfA from *D. discoideum* cells by gene replacement have been unsuccessful. As an alternative, a *cbfA* gene knock-in mutant was generated in which the codon at position 455 was replaced with a TAG (*amber*) stop codon (15). Because of the low frequency of *amber* codons in the *D. discoideum* genome, the expression of an *amber* suppressor tRNA gene does not produce a phenotype (16) but rather allows readthrough translation of the full-length protein from the *cbfA* (*amber*) mRNA. Because of the low efficacy of *amber* suppression, the JH.D strain produces less than 5% of the full-length CbfA protein (15).

In cells of the CbfA mutant JH.D, both plus- and minus-strand RNA from the retrotransposon TRE5-A is diminished and the mobility of the endogenous retrotransposon population is drastically reduced (13). Thus, CbfA is a positive regulator of TRE5-A amplification. Expression and retrotransposition of TRE5-A can be restored in JH.D cells by the ectopic expression of the full-length CbfA protein. Interestingly, the expression of the isolated carboxy-terminal domain (CTD) of CbfA in the mutant cells was sufficient for complete restoration of the TRE5-A transcript levels, even though TRE5-A retrotransposition was unaffected (13).

Received 23 November 2012 **Accepted** 18 January 2013
Published ahead of print 25 January 2013
Address correspondence to Thomas Winckler, t.winckler@uni-jena.de.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/EC.00329-12.

doi:10.1128/EC.00329-12

JH.D cells are defective in phagocytosis, cytokinesis (15), and multicellular development (17). In mutant cells, the expression of the aggregation-specific adenylyl cyclase ACA is strongly reduced; therefore, cyclic-AMP (cAMP)-induced gene expression is absent. However, multicellular development of JH.D cells can be restored either by the application of exogenous cAMP pulses or by mixing mutant and wild-type cells, suggesting that the developmental phenotype of JH.D cells is not cell autonomous (17). The aggregation block in JH.D cells can be overcome by the ectopic expression of the full-length CbfA protein (17), but not by the expression of the isolated CbfA CTD (18).

The observation that the CTD of CbfA restored TRE5-A expression raised questions concerning the origin of this domain and its role as an independent gene regulatory entity in CbfA function. In this study, we show that CbfA is a general transcriptional regulator in growing *D. discoideum* cells and that most CbfA-dependent gene expression is mediated by the CbfA CTD. The CbfA protein, including the CTD, is highly conserved throughout dictyostelid evolution. Dictyostelids contain a CbfB protein. This protein is paralogous to CbfA in the amino-terminal portion of the protein but has an individual CTD with no similarity to those of other CbfA or CbfB proteins. We speculate that the two gene regulatory elements within the CbfB proteins in individual species have evolved differently to facilitate the emergence of genome-specific gene regulatory functions.

MATERIALS AND METHODS

Dictyostelids. The dictyostelid species used in this study were obtained from the *Dictyostelium* Stock Center, which can be accessed at dictyBase (http://dictybase.org). The following strains were used (dictyBase strain identification numbers are in parentheses): *D. citrinum* OH594 (DBS0235738), *D. rosarium* M45 (DBS0235885), *D. dimigraformum* AR5B (DBS0235745), *D. giganteum* WS589 (DBS0235820), *D. sphaerocephalum* GR11 (DBS0235889), *D. robustum* Ch53 (DBS0235883), *D. firmibasis* TNS-C-14 (DBS0235812), *D. intermedium* PJ11 (DBS0235829), *D. longosporum* TNS-C-109 (DBS0235836), *D. clavatum* TNS-C-189 (DBS0235739), *D. brunneum* WS700 holotype (DBS0235734), *D. brefeldianum* G-12-1 (DBS0235732), *D. minutum* 71 (DBS0235843), *D. lacteum* (DBS0235831), *Polysphondylium pallidum* WS320 (ATCC 44843), *Acytostelium subglobosum* Lb1 (DBS0235452), and *D. fasciculatum* SH3 (DBS0235810).

Isolation of the dictyostelid *cbfA* and *cbfB* gene sequences. The dictyostelid species were grown on SM/5 agar plates (19) in association with *Raoultella planticola*. Before the first multicellular stages appeared, the cells were harvested, washed, and stored as pellets consisting of 2×10^7 cells. The pellets were frozen at −80°C. Genomic DNA was prepared from the frozen cells with the Qiagen DNeasy Tissue kit. Total RNA was prepared from the frozen cells with the Qiagen RNeasy Minikit. cDNA synthesis was performed with the Qiagen Omniscript RT kit with an oligonucleotide-deoxythymidine primer.

Degenerate primers were derived from the *D. discoideum* CbfA protein sequence and multiple sequence alignments of CbfA sequences from group 4 species. DNA sequences of the orthologous *cbfA* genes were generated from overlapping PCR fragments. The PCR products were either cloned into the pGEM-T plasmid (Promega) or sequenced directly from the PCR fragments. The DNA sequences representing approximately 90% of the encoded CbfA proteins were assembled and aligned with COBALT (20).

The *D. discoideum cbfA* and *cbfB* gene sequences can be accessed via dictyBase (http://dictybase.org/) under accession numbers DDB_G0279409 and DDB_G0293470, respectively.

Information about the *cbfA* and *cbfB* genes of *D. fasciculatum* and *P. pallidum* was assembled in the Social Amoebas Comparative Genome Browser (http://sacgb.fli-leibniz.de/cgi/index.pl) from the preliminary data of the comparative genome sequencing project (4). CbfA and CbfB sequences of *A. subglobosum* were generously provided by the *Acytostelium* gene analysis team at the University of Tsukuba (H. Urushihara, personal communication). The *cbfA* and *cbfB* sequences from *D. purpureum* were generously provided by A. Kuspa (Baylor College of Medicine, Houston, TX). These sequences are now publicly available at dictyBase (5). The structures of the *cbfA* and *cbfB* gene sequences as predicted by the dictyostelid comparative genome projects were confirmed by sequencing overlapping PCR products that were generated from both genomic DNA and cDNA of the respective species.

Expression of the CbfA CTD mutant proteins. A gene fragment spanning amino acids 724 to 1000 of the *D. discoideum* CbfA protein was cloned as a BamHI fragment into the pDXA-GFP2 plasmid (21). The resulting plasmid, pDXA-DdCbfA-CTD, was later transformed into JH.D cells (15). The corresponding sequence of *P. pallidum* CbfA was isolated by PCR and cloned as mentioned above. All of the strains were cultured in standard HL-5 medium (ForMedium, Norfolk, United Kingdom) supplemented with 10 μg/ml G418. Live cells expressing the green fluorescent protein (GFP)-tagged CbfA CTD were stained with 20 μg/ml 4′,6-diamidino-2-phenylindole (DAPI) in phosphate-buffered saline (PBS) for 30 min at room temperature. The cells were washed with PBS and observed with an Axio Observer Z1 microscope (Carl Zeiss AG). The expression of GFP-tagged CbfA CTD proteins was analyzed by Western blotting as described previously (22).

Illumina RNA sequencing (RNA-seq). Total RNA was prepared from cells grown in three independent cultures containing AX2 cells (wild type), JH.D cells (the CbfA-underexpressing mutant), or JH.D cells expressing the GFP-tagged CbfA CTD from *D. discoideum* or *P. pallidum*. For library preparation, 5 μg of total RNA per sample was processed with an Illumina mRNA-Seq sample prep kit (RS-122-2001) by following the manufacturer's instructions. Library preparation was performed with the total RNA from the three independent cultures. The libraries were sequenced with a HiSeq 2000 instrument in a single-read approach with 50 cycles resulting in reads with a length of 50 nucleotides. Reads were mapped to the *D. discoideum* reference genome downloaded from dictyBase (http://dictybase.org/) with the Illumina-supplied tool ELAND (23) by using standard parameters. The expression values were calculated by determining the number of uniquely mappable reads per transcript and normalizing that value to the total number of mappable reads and the length of the respective transcript (adapted from reference 24). This approach resulted in RPKM values (reads per kilobase transcript and million mappable reads) representing the expression level of the respective transcript. The determination of differentially expressed genes was performed with the unnormalized count data with the R statistical computing (25) tools edgeR (26) and DEseq (27). Genes were considered to be differentially expressed if the *P* value was ≤0.01. Furthermore, an at least 3-fold difference in gene expression was considered CbfA-dependent regulation. Gene ontology (GO) term enrichment analyses were performed with AmiGO version 1.8 (28).

Phylogenetic analysis. The multiple-sequence alignment was performed with ClustalX 2.0 (29). The phylogenetic tree was generated with Tree-Puzzle 5.2 (30) under the Dayhoff model. The heterogeneity rate was uniform. Supporting values are derived from 1,000 puzzling steps. The tree was rooted with the *Naegleria* sequence as the outgroup.

Quantitative reverse transcription (qRT)-PCR. Total RNA was prepared from frozen pellets consisting of 2×10^7 cells with the Qiagen RNeasy Minikit according to the protocol provided. cDNA was synthesized by the RT of 500 ng of total RNA with an oligo(dT) primer and the Qiagen Omniscript RT kit. Real-time amplification of the genes was performed with the *Taq* polymerase mix from Jena Bioscience (Jena, Germany) supplemented with EvaGreen Fluorescent Gel Stain and ROX Reference Dye (both from Jena Bioscience, Jena, Germany) on a Stratagene Mx3000P instrument. The sequences of the primers used are listed in Table S1 in the supplemental material. All of the measurements were performed in triplicate. Real-time PCR signals were normalized to the

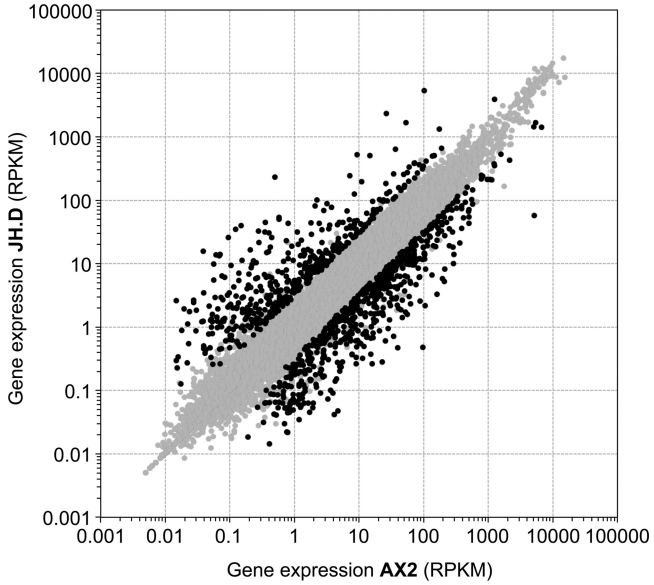

FIG 1 Transcriptome analysis of CbfA mutant JH.D cells. Shown are the results from an RNA-seq experiment with AX2 and JH.D cells. The reads were mapped to 12,317 gene models, and read counts were standardized to transcript lengths and 1 million reads (RPKM values). The mean RPKM values of three independent cultures are plotted as gray symbols. The black symbols highlight genes that were reported to show statistically significant differential levels of regulation ($P \leq 0.01$) of at least 3-fold.

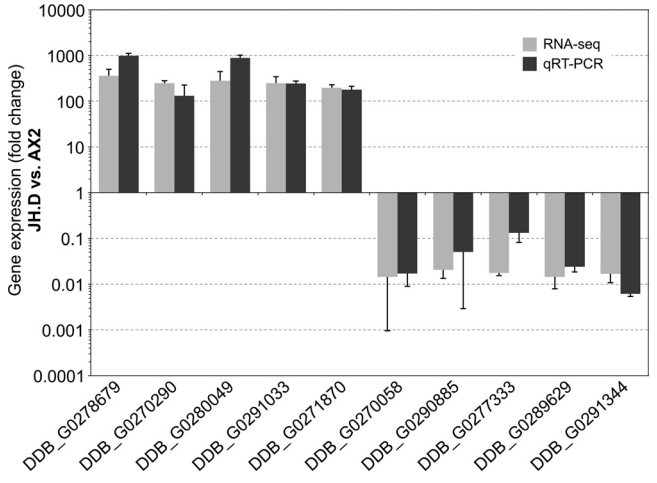

FIG 2 qRT-PCR analysis of a representative set of CbfA-regulated genes. Comparison of RNA-seq data (gray bars) with qRT-PCR data (black bars) for five CbfA-repressed and five CbfA-supported genes. The qRT-PCR was performed with cDNA prepared from three independent cultures of AX2 and JH.D cells that were different from the three cultures used for RNA-seq. Values are presented as means of three biological replicates ± the standard deviations. Note that the expression of genes is calculated for JH.D versus AX2 cells, meaning that values of >1 represent genes usually repressed by CbfA (and therefore overexpressed in JH.D), whereas values of <1 represent genes whose expression is supported by CbfA.

expression of the gene encoding *catA* (dictyBase entry DDB_G0274595). After an initial denaturing step at 95°C for 10 min, PCR was performed for 40 cycles at 95°C for 30 s, 60°C for 45 s and 72°C for 30 s. Differential gene expression was calculated according to the Pfaffl method (31) with *catA* as a reference gene.

Nucleotide sequence and RNA-seq data accession numbers. The partial *cbfA* gene sequences that were determined in this study have been deposited in GenBank under accession numbers JQ918362 (*D. giganteum cbfA*), JQ918358 (*D. citrinum cbfA*), JQ918363 (*D. intermedium cbfA*), JQ918366 (*D. sphaerocephalum cbfA*), JQ918359 (*D. clavatum cbfA*), JQ918360 (*D. dimigraformum cbfA*), JQ918361 (*D. firmibasis cbfA*), JQ918365 (*D. rosarium cbfA*), JQ918357 (*D. brunneum cbfA*), and JQ918356 (*D. brefeldianum cbfA*). The RNA-seq data have been deposited in the Gene Expression Omnibus database under accession number GSE38418.

RESULTS

CbfA is a general transcriptional regulator. The complex phenotype of CbfA-deficient mutants suggested that CbfA functions as a general transcriptional regulator in growing and developing *D. discoideum* cells. To test this hypothesis, we defined a list of CbfA-dependent genes in growing cells. To this end, we performed Illumina RNA-seq analyses on the *D. discoideum* wild-type strain AX2 and the CbfA-deficient mutant JH.D. RNA-seq on three independent cultures from each strain was highly reproducible (see Fig. S1 in the supplemental material) and yielded approximately 20 million sequence reads per RNA preparation, which were mapped to 12,317 *D. discoideum* gene models. We plotted the normalized sequence reads (RPKM values) for AX2 versus JH.D. Considerable differences in gene expression between the strains existed (Fig. 1). Statistical analysis revealed that 2,298 genes were differentially expressed between the AX2 and JH.D samples. As expected, the numbers of differentially expressed genes varied considerably, depending on the threshold setting (see Fig. S2 in the supplemental material). A 3-fold change in gene expression was considered to be indicative of CbfA-regulated gene expression. With this threshold, 1,030 genes were differentially regulated (see Table S2 in the supplemental material). Of these genes, 584 had reduced expression in JH.D cells, suggesting that these genes are positively regulated by CbfA. Fifty-seven genes were downregulated more than 20-fold in the mutant. The most pronounced CbfA-regulated gene was *abpF* (205-fold differential expression), which codes for an actin-binding protein that is most likely localized to the centrosomes. We found that 446 genes were upregulated in JH.D cells, which suggests that CbfA normally represses these genes. Eighty-six genes were upregulated by at least 20-fold in the CbfA mutant cells versus the AX2 cells. The most significant differences were a 458-fold change in the expression of a gene with unknown function (DDB_G0283511) and a 388-fold upregulation of *cyp513F1*, which encodes a cytochrome P450 enzyme. The gene *agnC*, which codes for an Argonaute protein that may be involved in processes involving posttranscriptional gene silencing, was upregulated by 225-fold in the CbfA mutant cells. We performed a qRT-PCR analysis of a representative set of 10 CbfA-regulated genes to verify the RNA-seq data. We measured the gene expression in the RNA preparations from 3 cultures that were grown independently from the cells used for RNA-seq. As indicated in Fig. 2, the qRT-PCR data confirmed the RNA-seq results.

GO term enrichment analysis with the program AmiGO set at a $P < 0.05$ cutoff indicated that a significant enrichment of the positively CbfA-regulated genes was involved in the assembly and maintenance of the actomyosin cytoskeleton and phagocytic vesicles, as well as general cellular processes, such as cytokinesis and cell morphogenesis (see Table S3 in the supplemental material). This observation correlated well with the growth characteristics of CbfA-deficient cells, including a decreased phagocytosis rate, slow growth on bacteria, and the generation of giant cells containing

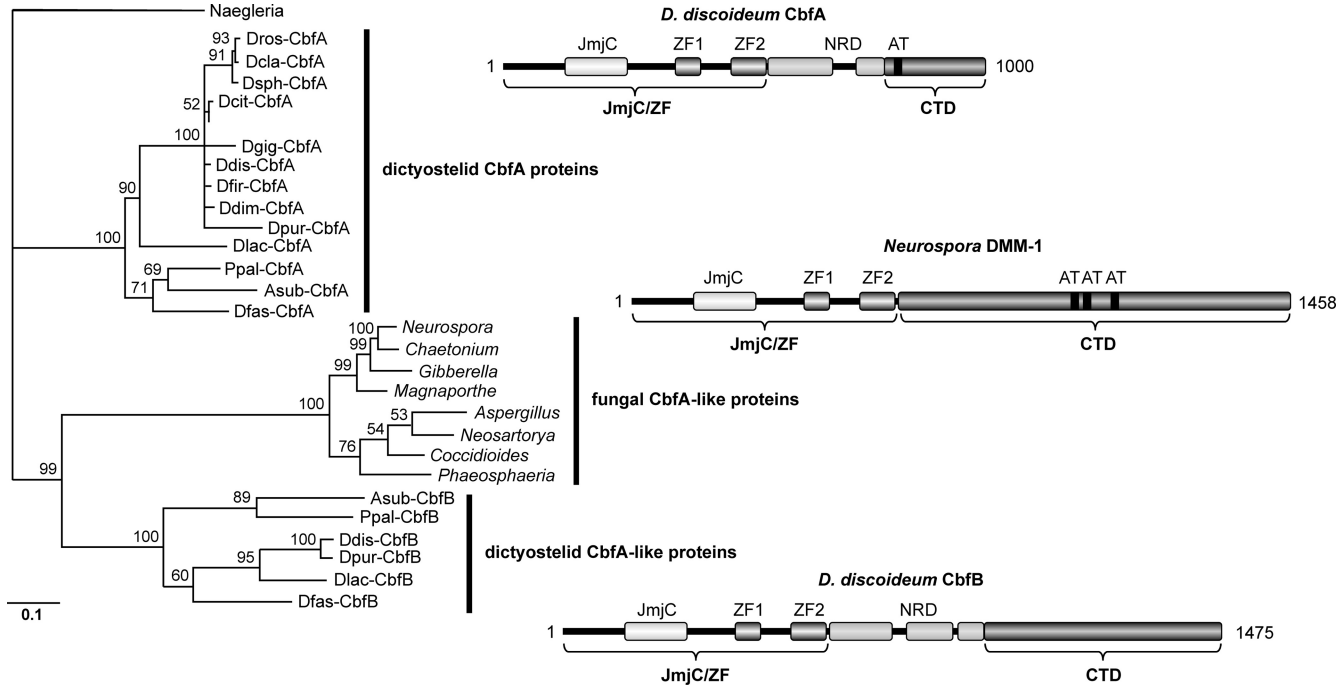

FIG 3 Phylogenetic relationships of CbfA-like proteins. Protein sequences corresponding to amino acids 1 to 553 of *D. discoideum* CbfA (JmjC/ZF region) were aligned with ClustalX 2.0 and analyzed with Tree-Puzzle 5.2. The values above the lines indicate the support values derived from 1,000 puzzling steps. In addition to the dictyostelid CbfA and CbfB proteins, the following fungal CbfA-like proteins were used in this alignment: *Aspergillus nidulans* (GI: 67528200), *Chaetonium globosum* (GI: 116199707), *Coccidioides immitis* (GI: 119188365), *Gibberella zeae* (GI: 46138433), *Magnaporthe grisea* (GI: 39973571), *Neosartorya fischeri* (GI: 119501326), *Neurospora crassa* (GI: 85078924), and *Phaeosphaeria nodorum* (GI: 111066574). Note that the *N. crassa* protein is identical to DMM-1, as reported by Honda et al. (39). The ZF2 region was also aligned with the zinc finger domain of *Arabidopsis* monoamine oxidase A repressor R1 (GI: 145360248) and a related zinc finger region in human cell division cycle-associated 7-like protein isoform 2 (GI: 188497635). Domain structures are schematically presented for *D. discoideum* CbfA, *D. discoideum* CbfB, and *N. crassa* DMM-1. JmjC, jumonji C domain; ZF, zinc finger region, AT, AT hook.

multiple nuclei (15). We performed a qRT-PCR analysis of 10 of these genes on RNA from three independently grown cultures to confirm the RNA-seq data. The genes selected exhibited between 3.3- and 31.2-fold higher expression levels in wild-type cells than in the CbfA mutant cells in the RNA-seq experiment. Surprisingly, the qRT-PCR data did not confirm the RNA-seq data (data not shown). This discrepancy, however, was not due to a general incongruence between RNA-seq and qRT-PCR data because the values for the genes measured by qRT-PCR in the RNA samples that we used for RNA-seq were comparable to the values measured by RNA-seq (not shown). Thus, it is probable that the expression of the genes putatively involved in the CbfA-dependent regulation of phagocytosis is generally variable and subject to subtle changes in culture conditions and the involvement of CbfA in this gene regulation remains uncertain. We should note, however, that the slow-growth phenotype of JH.D cells on bacteria lawns is reversed by the expression of CbfA in the mutant (data not shown).

CbfA is a highly conserved protein. *D. discoideum* CbfA is a 1,000-amino-acid multidomain protein (Fig. 3) that belongs to the family of jumonji-type transcriptional regulators. These proteins are often involved in deciphering the histone code by removing methyl groups from methyllysine or methylarginine residues in the histone tails (32, 33). A "carboxy-terminal jumonji domain" (JmjC) catalyzes this type of oxidative demethylation (reviewed in references 34 and 35). The JmjC domain is located in the amino-terminal portion of the CbfA protein, between amino acids 113 and 280. This domain is followed by two cysteine-rich, zinc finger-like regions located at positions 373 to 414 (ZF1) and 492 to 550 (ZF2). Downstream of the ZF motifs is a distinct region of 217 amino acids that consists of 50% asparagine residues; therefore, this region is referred to as the CbfA asparagine-rich domain (NRD). The CbfA NRD separates the JmjC/ZF region of the CbfA protein from the CTD (CbfA CTD), which spans 230 amino acids and contains a single AT hook.

Dictyostelids are classified into four major groups based on their phylogenetic relationships (36). *D. discoideum* belongs to the latest diverged group, 4. On the basis of the primary structure of *D. discoideum* CbfA (14), we obtained partial sequences of the *cbfA* genes from eight other group 4 species by PCR with degenerate primers. These data were complemented with full-length *cbfA* genes obtained from the recently completed genome sequencing projects for the group 4 species *D. purpureum* (5), the group 3 species *D. lacteum* (G. Glöckner, unpublished data), the group 2 species *P. pallidum* (4) and *A. subglobosum* (available in the *Acytostelium* Genome Database), and the group 1 member *D. fasciculatum* (4). Sequences of the *cbfA* genes derived from genome sequencing projects were confirmed by sequencing genomic DNA and cDNA (data not shown). An alignment of 14 dictyostelid CbfA proteins is presented in Fig. S3 in the supplemental material.

Within group 4, the level of identity between the JmjC/ZF domains ranges from 74 to 97% and covers ~550 amino acids. The CbfA proteins from the organisms in groups 1 to 3 are more than 45% identical to the JmjC/ZF region of CbfA of *D. discoideum*. CbfA-like zinc finger regions (see Fig. S4 in the supplemental ma-

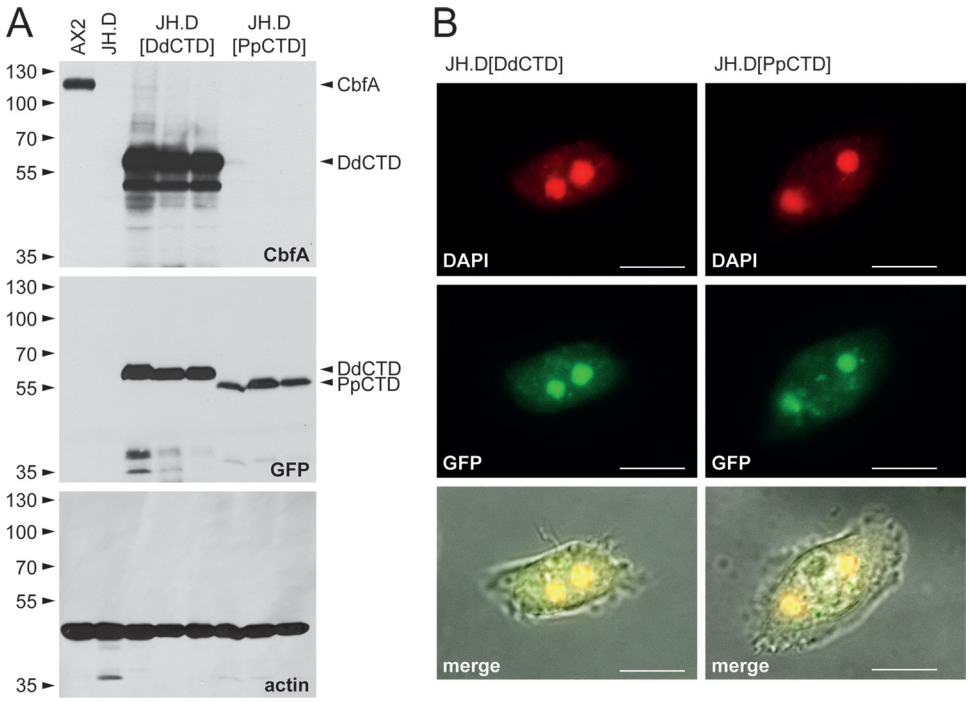

FIG 4 Expression of CbfA CTD in JH.D cells. (A) Western blot assays illustrating expression levels of the GFP-tagged *D. discoideum* CbfA CTD and the *P. pallidum* CbfA CTD in JH.D cells. Total cell extracts were prepared from AX2 and JH.D cells and the three JH.D transformants that were used for RNA-seq experiments. Shown is a single blot that was successively stained with CbfA monoclonal antibody 7F3 (which recognized *D. discoideum* CbfA CTD but not *P. pallidum* CbfA CTD), a GFP-specific antibody, and a polyclonal antiserum specific for actin (loading control). Values to the left indicate molecular mass markers (kilodaltons). (B) Cellular localization of the GFP-tagged *D. discoideum* CbfA CTD and the *P. pallidum* CbfA CTD in JH.D live cells. Top panel, DAPI staining; middle panel, GFP fluorescence; bottom panel, merged DAPI and GFP fluorescence overlaid with a phase-contrast picture. Note that a *D. discoideum* culture usually contains a fraction of multinucleated cells. Cells with two nuclei are compared here. Bars, 10 μm.

terial) are found in proteins from plants and animals; however, these proteins contain either ZF1 or ZF2 and are not linked with a JmjC domain. The CbfA-like combination of JmjC and ZF1/ZF2 domains was detected in proteins from several filamentous fungi (but not yeasts). Furthermore, this structure was detected in a protein of unknown function from the amoeboflagellate *Naegleria gruberi*. A protein with similarity to CbfA was also detected in the genome of the amoebozoan *Acanthamoeba castellanii* (Glöckner, unpublished). As discussed below, the CbfA-archetypical JmjC/ZF1/ZF2 design combined with the divergence of the CTDs defines a new family of "C-module-factor-like proteins."

CbfA CTDs are well conserved, with sequence identities ranging from 81 to 98% in group 4 and >40% in the non-group 4 species compared with *D. discoideum* CbfA. This *D. discoideum* CbfA CTD is later referred to as the CTD "core" because CbfA CTDs from non-group 4 species are considerably extended (see Fig. S3 in the supplemental material). CbfA CTD extensions range from 140, 223, 107, and 255 amino acids in CbfA from *D. lacteum*, *P. pallidum*, *A. subglobosum*, and *D. fasciculatum*, respectively. These extensions do not display any detectable sequence homology with each other, despite the strong conservation of the CTD core.

The CTD of CbfA displays deeply conserved gene regulatory functions. Having established a comprehensive list of CbfA-regulated genes in growing *D. discoideum* cells from the RNA-seq data, we next investigated whether the CbfA CTD could repair the aberrant gene expression in JH.D cells. The cells were transformed with a plasmid containing a GFP-tagged *D. discoideum* CbfA CTD variant under the control of a constitutive promoter. As shown in Fig. 4A, the expression of the CbfA CTD was higher than the endogenous CbfA levels in AX2 cells, and the CbfA CTD protein was enriched in the nuclei of JH.D transformants (Fig. 4B). These transformants were used for RNA-seq experiments, and gene expression levels were compared with those of AX2 and JH.D cells. Gene expression levels in JH.D cells that were 50% of the levels in AX2 cells were considered to be successfully restored by CbfA CTD transformation. Thus, 85% of the CbfA-activated and 76.7% of the CbfA genes were responsive to CbfA CTD. With a more stringent threshold of 80% of the wild-type expression level, we still determined that 76.5% of the CbfA-activated genes and 47.3% of the CbfA-repressed genes were complemented by the expression of the CbfA CTD in the CbfA mutant. We observed a pronounced overcomplementation in the mutant cell transformants; 69.5% of the CbfA-activated genes in the CbfA CTD-transformed mutant cells were expressed at levels higher than those observed in wild-type cells (see Table S2 in the supplemental material). The overcomplementation phenomenon in the mutant cells could result from the overexpression of the CbfA CTD. Alternatively, the JmjC/ZF region of CbfA may modulate the activity of the CTD in the full-length protein. In summary, the CbfA CTD may function to regulate the majority of the CbfA-regulated genes without requiring the JmjC/ZF domains.

We next determined whether the CbfA CTD of a dictyostelid species that diverged early would be functionally equivalent to the *D. discoideum* CbfA CTD. We cloned the CbfA CTD from *P. pallidum* and expressed the gene as a GFP-tagged protein in *D. dis-*

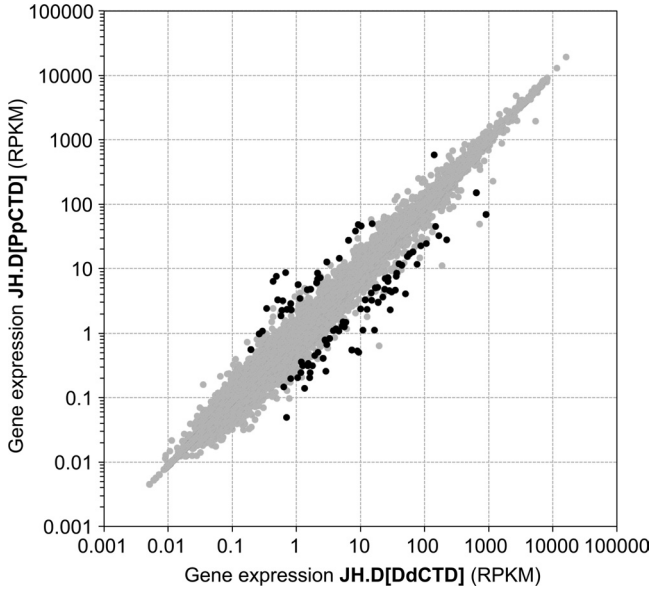

FIG 5 Gene regulatory function of *P. pallidum* CbfA CTD. Shown are the results from an RNA-seq experiment with JH.D cells expressing either the GFP-tagged *D. discoideum* CbfA CTD or the *P. pallidum* CbfA CTD. The reads were mapped to 12,317 gene models, and read counts were standardized to transcript lengths and 1 million reads (RPKM values). The mean RPKM values of three independent cultures are plotted (gray symbols). The black symbols highlight genes that were reported to show statistically significant differential levels of regulation ($P \leq 0.01$) of at least 3-fold.

coideum JH.D cells (Fig. 4A). Similar to the *D. discoideum* CbfA CTD, the *P. pallidum* CbfA CTD was localized to the nuclei of the transformed cells (Fig. 4B). RNA-seq data revealed that the expression of 73.3% of the CbfA-activated genes (73.3%) and 51.4% of the CbfA-repressed genes was restored by the *P. pallidum* CbfA CTD to at least 80% of the expression level in AX2 cells. Although the expression of the *P. pallidum* CbfA CTD in JH.D cells was considerably less than the expression of the *D. discoideum* CbfA CTD (Fig. 4A), we observed a similar overcomplementation effect on CbfA-dependent gene expression (64.0% of the CbfA-activated genes were expressed at >100% of the AX2 levels; see Table S2 in the supplemental material). Thus, overcomplementation by the CbfA CTD may, in fact, be a consequence of the missing JmjC/ZF region of the CbfA protein rather than an artifact of the overexpression process.

In Fig. 5, we plotted the RNA-seq data obtained from JH.D cells expressing the *D. discoideum* CbfA CTD versus that from the mutant cells expressing the *P. pallidum* CbfA CTD. Gene expression differences that were less than 3-fold were not considered for analysis. We determined that only 98 (0.8%) of the 12,317 genes were differentially regulated in the two data sets. Of these genes, only 31 were identified on the list of 1,030 CbfA-regulated genes, indicating a great functional overlap between the *D. discoideum* and *P. pallidum* CbfA CTDs. In fact, we found that approximately 87% of the CbfA CTD-responsive genes were restored to at least 80% of the wild-type level by both the *D. discoideum* and *P. pallidum* CbfA CTDs. This result is notable because *D. discoideum* and *P. pallidum* have different percentages of GC content in their genomes and promoter alignments of the two species are nearly impossible (4); however, the gene regulatory function of the *D. discoideum* CbfA CTD requires direct binding to the AT-rich DNA sequences within the target promoters, as suggested by *in vitro* and *in vivo* data (14, 18).

CbfA-like proteins contain conserved JmjC/ZF domains and divergent carboxy termini. "CbfA-like proteins," which we define in this study as proteins displaying the characteristic JmjC/ZF1/ZF2 architecture of dictyostelid CbfA, were detected both within and outside the dictyostelid clade. The average percentage of amino acid sequence identity between CbfA and CbfA-like proteins from dictyostelids (i.e., paralogous CbfB proteins) and fungi is approximately 25% in the protein regions with conserved architecture. The distinctive design of the zinc finger regions of CbfA (see Fig. S4 in the supplemental material) provides an indication of the evolution of CbfA and CbfA-like proteins. The ZF1 region of CbfB proteins shares the signature $CX_2CX_7\mathbf{X}X_2CX_2C$ (where **X** is any amino acid replacing cysteine in this position) with both fungal CbfA-like proteins and one of the earliest diverged dictyostelid, *D. fasciculatum*. This motif developed into $CX_2CX_7CX_2CX_2C$ in dictyostelid CbfA proteins and was further expanded by a CX_2H motif found exclusively in CbfA proteins (see Fig. S4 in the supplemental material). Importantly, a CbfA-like protein from *Naegleria gruberi* and CbfB from the dictyostelid *A. subglobosum* (group 2 dictyostelid) share a $CX_2CX_7CX_2\mathbf{X}X_2\mathbf{X}$ variant of ZF1 (**X** replacing cysteine in this position). In contrast to the design of the ZF1 region, that of the ZF2 region is highly conserved among CbfA, CbfB, and fungal CbfA-like proteins, although the distances between the two cysteine residues in the second CX_2C motif are extended by up to 36 amino acids in the CbfB proteins.

The arrangement of the cysteines in the ZF1 region suggests that the fungal CbfA-like proteins are more closely related to the dictyostelid CbfB proteins than to the CbfA proteins. A phylogenetic analysis of the JmjC/ZF regions of the CbfA-like proteins revealed a clear separation between the dictyostelid CbfA and CbfB proteins and their fungal orthologs (Fig. 3). To date, the CbfA CTD has not been found in proteins from any other organisms outside the dictyostelids. In striking contrast to the conservation of the CTDs of the CbfA proteins, the CbfB proteins have extended CTDs that do not exhibit any signs of similarity. The CbfB CTDs from the group 4 organisms *D. discoideum* and *D. purpureum* do not show any similarity, even though they are 85% identical in the JmjC/ZF regions. Similarly, the CTDs of the CbfA-like proteins from fungi do not share any traceable similarities. These observations indicate that the CTDs of the CbfA-like proteins have rapidly evolved because of either species-specific adaptation pressures or the total removal of clade-specific constraints.

DISCUSSION

Two different gene regulatory functions combined in one protein. In this study, we performed Illumina RNA-seq to extend our knowledge concerning the function of the CbfA protein. A previous DNA microarray experiment performed with CbfA-deficient mutant JH.D cells (18) had limited sensitivity and included only approximately half of the *D. discoideum* genes. Comparing the two transcriptomic approaches, we observed that only 35% of the 1,030 CbfA-dependent genes identified by RNA-seq were present on the DNA microarray. Only 42 of the 365 genes were reported as CbfA regulated in the DNA microarray experiment, where the threshold was set at a 1.5-fold change in expression, thereby suggesting that the sensitivity of the DNA microarray was lower than that of the RNA-seq approach. This assumption is supported by the observation that the detection of CbfA-dependent genes by DNA microarray analysis was biased toward strongly expressed

genes, as illustrated in Fig. S5 in the supplemental material. Both the DNA microarray experiments and the RNA-seq analyses reported CbfA positively regulated genes that were involved in phagocytosis. However, the results obtained with the two transcriptomic methods were not similar. GO term enrichment analysis of the DNA microarray data with AmiGO version 1.8, which is the same program used to analyze the RNA-seq data, reported that 16 genes that were assigned to the cellular function of "phagocytic vesicles" were regulated by CbfA (data not shown). However, nine of these genes produced conflicting results in the RNA-seq experiment and were not detected as CbfA-dependent genes. Even more confusing was the observation that qRT-PCR analysis measuring the expression of the CbfA-regulated genes that were associated with phagocytosis did not confirm the prediction from RNA-seq. Because we used different cultures for the RNA-seq and qRT-PCR experiments, we hypothesize that the expression of genes involved in phagocytosis is considerably variable and culture dependent.

Consistent with the previous DNA microarray data (18), we found with RNA-seq that the CbfA CTD core exerts the majority of its gene regulatory activity without requiring the other domains of CbfA. Thus, CbfA can be divided into two independently functioning regions: the JmjC/ZF region and the CTD. At present, we are not able to determine how the JmjC/ZF region of the CbfA protein contributes to CbfA function. In the experiments presented here, approximately two-thirds of the CbfA-mediated gene regulation was mediated by the CbfA CTD alone. Thus, the JmjC/ZF region of CbfA may only function as a regulatory domain to modulate the CbfA CTD. However, there is evidence that the JmjC/ZF portion of CbfA is required for gene regulatory functions that cannot be carried out by the CbfA CTD alone. For example, the CbfA CTD was not able to rescue the developmental defect in CbfA-deficient JH.D cells (18). The fact that CbfA-deficient cells have a strong developmental phenotype indicates that CbfB, which is expressed in the CbfA mutant at normal levels, is not able to complement the loss of CbfA. Nevertheless, we cannot exclude the possibility that the number of genes regulated by the JmjC/ZF region of CbfA has been underestimated in our study because of a partial overlap of CbfA with similar functions of the CbfB-derived JmjC/ZF domain. This question will be addressed once a CbfB-deficient mutant becomes available.

Preliminary data have suggested that the JmjC/ZF portion of CbfA is sufficient to rescue JH.D cell development in the absence of the CbfA CTD, which would imply that the two domains of the CbfA protein function independently (S. Gatz and T. Winckler, unpublished data). This finding would suggest that gene regulation mediated by the CbfA CTD depends on the direct binding to specific promoter regions in target genes; gene regulation mediated by the ectopically expressed CbfA CTD is compromised by mutations in the AT hook motif (18). It is tempting to speculate that the JmjC/ZF domain of CbfA acts as a chromatin modifier without directly binding to DNA. This possibility would assume that the CbfA zinc fingers have functional parallels with plant homeodomain (PHD) fingers, which are known to interpret the histone code by binding to nucleosome core histones in a modification-dependent manner (37, 38). CbfA displays significant similarity to the *Neurospora* DMM-1 protein in both the JmjC and ZF domains but not in the CTD. DMM-1 is required to inhibit the spreading of histone 3 lysine-9 methylation from transposable element loci into nearby genes (39). Notably, both the JmjC and ZF domains are required for the function of DMM-1, while the distinct CTD of DMM-1 is dispensable, even though the presence of multiple AT hooks in the DMM-1 CTD would have suggested a role for this region in directing the protein to its target sequences. In fact, DMM-1 requires interactions with the DMM-2 protein to interact with target DNA (39).

Rapid evolution of CTDs. In this study, we defined a new family of "C-module-binding factor-like proteins" that are characterized by the unique combination of a JmjC domain and CbfA-archetypical zinc fingers. CbfA-like proteins emerged before the dictyostelid split, and the dictyostelid clade is characterized by the emergence of a second CbfA-like protein, i.e., CbfA. Whereas CbfA-like proteins are characterized by the rapid evolution of the CTDs, the CbfA CTD is highly conserved and is detectable only in the dictyostelid clade. The question that remains is whether CbfA has diverged from a CbfB-like ancestor or vice versa. The differences in the ZF1 region in the CbfA and CbfB proteins suggest that the CbfB proteins represent the ancestral variant of the ZF1 region. This region was subsequently expanded to the dictyostelid-specific CbfA proteins by the addition of the CX_2C and CX_2H motifs (see Fig. S4 in the supplemental material). Assuming that the zinc finger(s) in the CbfA proteins may be involved in the binding to specific histone modifications as a part of an intrinsic chromatin regulatory function similar to the plant PHD fingers (37, 38), it is possible that adaptations in the ZF1 region have yielded functional specifications in the JmjC/ZF regions of the CbfA and CbfB proteins. This hypothesis can be evaluated by comparing the gene expression in CbfA- and CbfB-deficient mutants. We speculate that CbfA arose in the dictyostelid clade by the duplication of an ancestral CbfA-like protein that consisted of a CbfB-like JmjC/ZF region and an unknown CTD. The recent discovery of a CbfA-like protein within an *Acanthamoeba* protein (Glöckner, unpublished) supports the suggestion that CbfA derived from the dictyostelid lineage by gene duplication from a CbfB-like ancestor. The CbfA CTD may have been a part of the ancestral CbfA-like protein that has been maintained by evolutionary pressures to serve common gene regulatory functions. Whether the rapid divergence of the CTDs in CbfA-like proteins (including *D. discoideum* CbfB) occurred through a mutation in a preexisting CTD or by frequent domain swapping remains unknown. The extended CbfA CTDs in the earlier dictyostelids may mediate functions specialized to the individual properties of the respective genomes.

The data presented in this study suggest that evolution has combined two independent gene regulatory functions into a single protein; whereas one of the regulatory elements was notably well conserved (JmjC/ZF), the other (CTD) was highly variable. This variability enabled the adaptation to the high genome flexibility that occurred in dictyostelid evolution (4). We predict that, similar to the CbfA CTDs, the CbfB CTDs have autonomous gene regulatory functions, but these functions are specialized to individual genomes and do not exhibit much overlap between species. However, a complete loss of function of this domain is also possible. This hypothesis can now be further tested experimentally on the basis of the data presented in this paper.

ACKNOWLEDGMENTS

We are grateful to H. Urushihara of the University of Tsukuba and A. Kuspa of the Baylor College of Medicine for discussing data from comparative dictyostelid genome projects prior to publication. We thank P.

Schaap for depositing many of the dictyostelid strains required for this study at the *Dictyostelium* Stock Center, and we thank the Stock Center staff for providing these strains.

This work was supported by the German Research Foundation DFG (WI 1142/6-1 and WI 1142/7-1).

REFERENCES

1. **Kessin RH.** 2001. *Dictyostelium*—evolution, cell biology, and the development of multicellularity. Cambridge University Press, Cambridge, United Kingdom.
2. **Bonner JT.** 2009. The social amoebae: the biology of cellular slime molds. Princeton University Press, Princeton, NJ.
3. **Eichinger L, Pachebat JA, Glöckner G, Rajandream M-A, 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.
4. **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.
5. **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, Fey P, Gaudet P, Anjard C, Mohan MB, Basu S, Bushmanova Y, van der Wel H, Katoh M, Coutinho PM, Saito T, Elias M, Schaap P, Kay RR, Henrissat B, Eichinger L, Rivero-Crespo F, Putnam NH, West CM, Loomis WF, Chisholm R, Shaulsky G, Strassmann JE, Queller DC, Kuspa A, Grigoriev I.** 2011. Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. Genome Biol. **12**: R20.
6. **Glöckner G, Szafranski K, Winckler T, Dingermann T, Quail M, Cox E, Eichinger L, Noegel AA, Rosenthal A.** 2001. The complex repeats of *Dictyostelium discoideum*. Genome Res. **11**:585–594.
7. **Hedges DJ, Deininger PL.** 2007. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat. Res. **616**:46–59.
8. **Huang CR, Burns KH, Boeke JD.** 2012. Active transposition in genomes. Annu. Rev. Genet. **46**:651–675.
9. **Winckler T, Schiefner J, Spaller T, Siol O.** 2011. *Dictyostelium* transfer RNA gene-targeting retrotransposons: studying mobile element-host interactions in a compact genome. Mob. Genet. Elements **1**:145–150.
10. **Beck P, Dingermann T, Winckler T.** 2002. Transfer RNA gene-targeted retrotransposition of *Dictyostelium* TRE5-A into a chromosomal UMP synthase gene trap. J. Mol. Biol. **318**:273–285.
11. **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.
12. **Schumann G, Zündorf I, Hofmann J, Marschalek R, Dingermann T.** 1994. Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the *Dictyostelium* repetitive element, from *Dictyostelium discoideum*. Mol. Cell. Biol. **14**:3074–3084.
13. **Bilzer A, Dölz H, Reinhardt A, Schmith A, Siol O, Winckler T.** 2011. The C-module-binding factor supports amplification of TRE5-A retrotransposons in the *Dictyostelium discoideum* genome. Eukaryot. Cell **10**: 81–86.
14. **Horn J, Dietz-Schmidt A, Zündorf I, Garin J, Dingermann T, Winckler T.** 1999. A *Dictyostelium* protein binds to distinct oligo(dA) x oligo(dT) DNA sequences in the C-module of the retrotransposable element DRE. Eur. J. Biochem. **265**:441–448.
15. **Winckler T, Trautwein C, Tschepke C, Neuhauser C, Zündorf I, Beck P, Vogel G, Dingermann T.** 2001. Gene function analysis by amber stop codon suppression: CMBF is a nuclear protein that supports growth and development of *Dictyostelium* amoebae. J. Mol. Biol. **305**:703–714.
16. **Dingermann T, Reindl N, Brechner T, Werner H, Nerke K.** 1990. Nonsense suppression in *Dictyostelium discoideum*. Dev. Genet. **11**:410–417.
17. **Winckler T, Iranfar N, Beck P, Jennes I, Siol O, Baik U, Loomis WF, Dingermann T.** 2004. CbfA, the C-module DNA-binding factor, plays an essential role in the initiation of *Dictyostelium discoideum* development. Eukaryot. Cell **3**:1349–1358.
18. **Lucas J, Bilzer A, Moll L, Zündorf I, Dingermann T, Eichinger L, Siol O, Winckler T.** 2009. The carboxy-terminal domain of *Dictyostelium* C-module-binding factor is an independent gene regulatory entity. PLoS One **4**:e5012. doi:10.1371/journal.pone.0005012.
19. **Raper KB.** 1984. The dictyostelids. Princeton University Press, Princeton, NJ.
20. **Papadopoulos JS, Agarwala R.** 2007. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics **23**:1073–1079.
21. **Levi S, Polyakov M, Egelhoff TT.** 2000. Green fluorescent protein and epitope tag fusion vectors for Dictyostelium discoideum. Plasmid **44**:231–238.
22. **Hentschel U, Zündorf I, Dingermann T, Winckler T.** 2001. On the problem of establishing the subcellular localization of *Dictyostelium* retrotransposon TRE5-A proteins by biochemical analysis of nuclear extracts. Anal. Biochem. **296**:83–91.
23. **Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara Catenazzi E M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ng BL, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Pinkard DC, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Rodriguez AC, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Sohna Sohna JE, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, vandeVondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ.** 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature **456**:53–59.
24. **Mortazavi A, Williams BA, McClue K, Schaeffer L, Wold B.** 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods **5**:621–628.
25. **Team RDC.** 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
26. **Robinson MD, McCarthy DJ, Smyth GK.** 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**:139–140.
27. **Anders S, Huber W.** 2010. Differential expression analysis for sequence count data. Genome Biol. **11**:R106. doi:10.1186/gb-2010-11-10-r106.

28. **Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, AmiGO Hub, Web Presence Working Group.** 2009. AmiGO: online access to ontology and annotation data. Bioinformatics **25**:288–289.
29. **Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG.** 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**:4876–4882.
30. **Schmidt HA, Strimmer K, Vingron M, von Haeseler A.** 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18**:502–504.
31. **Pfaffl MW.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. **29**:e45.
32. **Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y.** 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell **119**:941–953.
33. **Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y.** 2006. Histone demethylation by a family of JmjC domain-containing proteins. Nature **439**:811–816.
34. **Klose RJ, Kallin EM, Zhang Y.** 2006. JmjC-domain-containing proteins and histone demethylation. Nat. Rev. Genet. **7**:715–727.
35. **Agger K, Christensen J, Cloos PA, Helin K.** 2008. The emerging functions of histone demethylases. Curr. Opin. Genet. Dev. **18**:159–168.
36. **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.
37. **Taverna SD, Li H, Ruthenberg AJ, Allis CD, Patel DJ.** 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. **14**:1025–1040.
38. **Sanchez R, Zhou MM.** 2011. The PHD finger: a versatile epigenome reader. Trends Biochem. Sci. **36**:364–372.
39. **Honda S, Lewis ZA, Huarte M, Cho LY, David LL, Shi Y, Selker EU.** 2010. The DMM complex prevents spreading of DNA methylation from transposons to nearby genes in *Neurospora crassa*. Genes Dev. **24**:443–454.