Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Identifying the molecular basis of functions in the transcriptome of the social amoeba *Dictyostelium discoideum*

T.J. Whitney[1,2]*, D.G. Gardner[1]*, M.L. Mott[1,3] and M. Brandon[1,4]

[1]Department of Biological Sciences,
Idaho State University, Pocatello, ID, USA
[2]Current address: Department of Pathology,
University of Utah, Salt Lake City, UT, USA
[3]Current address: Pritzker School of Medicine,
University of Chicago, Chicago, IL, USA
[4]Current address: Fort Lewis College,
School of Natural and Behavioral Sciences, Durango, CO, USA

*These authors contributed equally to this study.
Corresponding author: M. Brandon
E-mail: Brandon_m@fortlewis.edu

Genet. Mol. Res. 9 (1): 394-415 (2010)
Received November 27, 2009
Accepted December 16, 2009
Published March 9, 2010

ABSTRACT. The unusual life cycle of *Dictyostelium discoideum*, in which an extra-cellular stressor such as starvation induces the development of a multicellular fruiting body consisting of stalk cells and spores from a culture of identical amoebae, provides an excellent model for investigating the molecular control of differentiation and the transition from single- to multi-cellular life, a key transition in development. We utilized serial analysis of gene expression (SAGE), a molecular method that is unbiased by dependence on previously identified genes, to obtain a transcriptome from a high-density culture of amoebae, in order to examine the transition to multi-cellular development. The SAGE method pro-

vides relative expression levels, which allows us to rank order the expressed genes. We found that a large number of ribosomal proteins were expressed at high levels, while various components of the proteosome were expressed at low levels. The only identifiable transmembrane signaling system components expressed in amoebae are related to quorum sensing, and their expression levels were relatively low. The most highly expressed gene in the amoeba transcriptome, *dut*A untranslated RNA, is a molecule with unknown function that may serve as an inhibitor of translation. These results suggest that high-density amoebae have not initiated development, and they also suggest a mechanism by which the transition into the development program is controlled.

Key words: *Dictyostelium discoideum*; Amoebae; Transcriptome; SAGE

INTRODUCTION

The social amoeba *Dictyostelium discoideum* has long been a productive model for basic eukaryotic cell processes, including cytoskeletal structure and motility, transmembrane signaling, and regulation of gene expression (Chung and Firtel, 2000). In addition, this organism's unusual life cycle, in which an extra-cellular stressor like starvation induces the development of a multicellular fruiting body consisting of stalk cells and spores from a culture of identical amoebae, provides an excellent model for investigating the molecular control of differentiation and the transition from single cell to multi-cellular life, a key transition in development (Firtel, 1996; Parent and Devreotes, 1996; Verkerke-van Wijk et al., 2001).

The advent of microarray technology has provided opportunities to efficiently compare the expression of known genes between various stages of development. For example, Urushihara et al. (2004) conducted a large-scale comparison of *Dictyostelium* cDNAs from the amoeba and slug stages of the life cycle to look for differences in gene expression patterns. Van Driessche et al. (2002) examined the expression patterns of genes across the life cycle to predict those whose expressions change during development, and therefore are expected to contribute to morphological changes.

Publication of the *Dictyostelium* genome sequence (Eichinger et al., 2005) made it possible to utilize molecular methods that are not biased by a dependence on previously identified genes. One of these methods is serial analysis of gene expression (SAGE), a technique that produces a library of cDNA tags from total RNA isolated from cells at a particular stage of the life cycle. After sequencing, the tags can be rank ordered relative to the expression levels of all other tags in the library (Velculescu et al., 1995). Comparison of the tag sequences to the *Dictyostelium* genome sequence identifies the genes, and produces a transcriptome that describes, in rank order, the activities of the cells that served as the source of RNA.

We report the production of an SAGE library and transcriptome from a high-density culture of *Dictyostelium* amoebae. We analyzed the transcriptome to predict the functional capabilities of amoebae and their ability to transition to multicellular development.

MATERIAL AND METHODS

Cell maintenance and preparation

We grew *Dictyostelium discoideum* AX2 wild-type cells in HL5 liquid media (0.06 M dextrose, 10 g/L proteose peptone, 5 g/L yeast extract, 0.004 M Na_2HPO_4, 0.004 M KH_2PO_4), shaking at 21°C, until they reached a concentration of 2.4 x 10^7 cells/mL. We washed cells twice in starvation buffer (100 mL 5 mM Na_2HPO_4, 5 mM NaH_2PO_4, 0.02 mM $CaCl_2$, 2 mM $MgSO_4$), and finally resuspended the cells in 1 mL starvation buffer.

RNA isolation

We extracted total RNA from amoebae prepared as described above using Ambion's RNAqueous RNA Isolation Kit (Austin, TX, USA). In order to test the integrity of the RNA, we separated 2 µL of each sample on a 1% (w/v) agarose gel. Visualization of distinct 18S and 28S ribosomal RNA bands indicated that the RNA was intact. We measured the concentration and purity of the RNA using a Pharmacia Biotech Gene Quant II spectrophotometer, and used 8.7 µg of total RNA as a starting point for SAGE.

Serial analysis of gene expression

We obtained an I-SAGE Kit from Invitrogen (San Diego, CA, USA) and followed the protocol and recommendations contained in the I-SAGE manual, with the modifications described below. In brief, we synthesized cDNA from the amoebae mRNA; then, through a series of restriction digests, we produced a 14-bp tag that was unique to each expressed gene. We ligated the tags into ditags, then into concatemers, and cloned and sequenced 30,830 tags.

In order to optimize our results, we modified several of the steps of the SAGE procedure. At various points in the SAGE procedure, wash steps, generally followed by brief (2-3 s) centrifugation, are recommended in order to recover all of the cDNA bound to magnetic beads. However, during cDNA synthesis we observed significant clumping of the beads after centrifugation, which severely limited cDNA recovery. Therefore, we eliminated the centrifugation steps. The final step in the preparation of 26 bp ditags is ethanol precipitation and air-drying the DNA pellet. We replaced the air-drying step with removal of ethanol by pipetting and blotting the tubes on Kim-wipes as air-drying causes degradation of the 26 bp ditags (Kenzelmann and Muhlemann, 1999).

We found that high ditag DNA concentrations inhibit ligation of ditags into concatemers. In order to maximize ligation of the 26 bp ditags, we set up four different ligation reactions using 25% of the DNA in each tube, maintaining the manufacturer's suggested volumes of the other enzymes and reagents. Additionally, we added a second aliquot of T4 DNA ligase after 3 h of ligation, and increased the total ligation reaction time to 15 h. Finally, we vortexed and briefly (5 s) microfuged the ligation reaction tubes periodically during the ligation reaction.

Prior to gel purification of the concatemers, we heated the concatemer ligation reaction mixture to 65°C for 20 min to release the partially hybridized concatemers. Finally, we lengthened the reaction time for ligation of concatemers into pZErO® from 1 to 6 h at 16°C.

Data analysis

We used polymerase chain reaction (PCR) with M13 forward and reverse primers to amplify the cloned concatemers. We purified the PCR products using 40% (w/v) polyethylene glycol, and submitted them for sequencing to the Idaho State University Molecular Research Core Facility, which employs an Applied Biosystems 3100 Genetic Analyzer (Foster City, CA, USA).

We entered the sequence files into the SAGE software that is included in the I-SAGE kit. This software extracts individual tag sequences from the concatemer sequences, and compiles a list of each independent tag, and the number of times that tag is represented in the library. In order to identify the gene represented by each tag, we compared the specific tag sequence to the *D. discoideum* genome database (Chisholm et al., 2006) as well as the GenBank database, using BLAST (Basic Local Alignment Search Tool). All SAGE sequence tags were converted into FASTA format and a batch BLAST search was run against mRNA sequences downloaded from Genbank. Additionally, the SAGE sequence tags, in FASTA format, were individually imported into the query field of the nucleotide BLAST program located at www.ncbi.nlm.nih.gov. A mega BLAST search was then performed on the SAGE sequence tags using a *D. discoideum* (taxid:44689) curated, refseq mRNA, database. The BLAST program automatically adjusted the search parameters to search for a short input sequence. The best *D. discoideum* mRNA sequence alignment for each individual SAGE tag was then chosen only from those responses that had 100% alignment with the search tag. Due to the small size of the search tag, 14 bp, the e-value, or statistically significant value, was used only when two or more mRNA's had 100% alignment with the search tag; the mRNA with the lowest e-value was selected.

RESULTS

General properties of the SAGE library

The amoeba-stage SAGE library is based on the sequences of 30,830 individual tags. Of these, 27,412 tags appeared in the library in only one copy. Because we cannot distinguish between true extremely low copy number tags and sequencing errors, we did not report the single copy tags in the library. We discontinued tag sequencing at this point for several reasons. First, no tags shifted from the single copy category into a higher expression level category. Second, the relative expression levels of the most highly expressed genes (i.e., those expressed at the level of at least 13 copies) did not change with the identification of additional tags, suggesting that the library accurately represents the relationships between the expressed genes. These data suggest that amoebae express about 3418 genes at the level of at least two RNA molecules. Since the *Dictyostelium* genome predicts about 12,500 genes, the SAGE library predicts that about 27% of the genome is required for amoeba structure and functions (Eichinger et al., 2005). The complete list of SAGE tags and their relative expression levels is available at SAGEmap (accession number GSE12095). The amoeba-stage transcriptome described here consists of the 316 most highly expressed genes, that is, all genes expressed at the level of at least 13 copies (Table 1).

Table 1. Amoeba-stage complete transcriptome of *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
GGACTTAGAG	EAL64828	773	*dut*A untranslated RNA
ATTAGAGGAC	XM_629433.1, XM_629434.1, XM_635973.1, XM_635972.1, XM_635971.1	440	Ubiquitin (*ubqN, ubqO, ubqM, ubqL, ubqK*)
TGCAGACGTT	XM_640038.1	308	40S ribosomal protein S29 (*rps*29)
CAAAGACTTT	XM_640564.1	297	AX4 putative glycoside hydrolase/ *Mus musculus* zinc finger protein 119 (*Zfp*119)
ATTTTTATTA	XM_633616.1	258	CHD gene family protein containing chromodomain, helicase domain, and DNA-binding domain
GGATAACAAA	EAL69803	256	G beta-like protein (*gpbB*)
GGGTATGGAT	XM_636628.2	195	cysteine proteinase (*CP*4)
TTTGGTTGGT	XM_632956.1	175	calcium-dependent cell adhesion molecule-1 (*cadA*)
CTGGTTACTC	XM_64088.1	168	protein synthesis elongation factor 1-alpha (*tef*2)
GAAAGACGTC	XM_15384.1	168	V12 mRNA/ribosomal protein 1024 (*rps*9)
GAAAAGAGAA	XM_635906.1	155	40S ribosomal protein S6 (*rps*6)
CTCGTCCTTA	XM_636364.1	154	ribosomal protein L18 (*rpl*18)
TCGCTTTAAT	XM_634269.1	154	hypothetical protein
GATATTGAAG	XM_637875.1	151	ribosomal protein L10a (*rpl*10a)
TGGTTATCCA	XM_637389.2	142	*Dictyostelium discoideum* AX4 putative transport protein (sec24l) mRNA, complete cds
GCAATCAAAA	XM_630714.1	140	hypothetical protein
GGGTATCGAA	XM_639885.1	140	cysteine proteinase CP5 (*cprE*)
TGTAGAGGTG	XM_637392.1	139	60S ribosomal protein L4 (*rpl*4)
CTACTTTCGA	XM_635251.1, U13671.1	138	Actin binding protein (*hat B*)/hisactophilin II (*hsII*)
CAAGTTGAAC	AB000109.1	133	mitochondrial DNA
GTATAGCTTA	EAL64828	130	*dutA* RNA
CGTAAACCAG	X15383.1	122	Ribosomal protein (*rp*119)vegetative specific V14 gene
GTGGTTTGAA	XM_639670.1	112	40S ribosomal protein S30 (*rps*30)
CACTAATCAA	XM_635539.1	111	hypothetical protein
GGGTAAGATT	XM_641702.1	110	ribosomal protein L35a (*rpl*35a)
TGAAATATAT	XM_001732911.1	104	putative polyketide synthase (*pks*)
GGGTGATAAT	XM_633636.1	103	histidine kinase (dhkG) mRNA, complete cds
GCCACTTTCT	XM_628965.1	100	ribosomal protein s2 (*rps*2)/LLRep3
GGGTGCCGAT	XM_630282.1	95	hypothetical protein
TGCTATCCAC	XM_635481.1	95	40S ribosomal protein S26 (*rps*26)
AAAATTAAAA	XM_641596.1	94	actin (*act*8)
AAAATTAAAA	XM_640170.1	94	actin 16 (*act*16)
GCCGCTCAAG	XM_630185.1	94	40S ribosomal protein S14 (*rps*14)
TGGTTACAAG	XM_641274.1	94	ribosomal protein L36a (*rpl*36a)
AGAGGTCTCA	XM_636637.1	93	vegetative specific gene V18 gene for ribosomal protein/ ribosomal protein L11 (*rpl*11)
GTCGTCCAGG	XM_633034.1	92	40S ribosomal protein S15 (*rps*15)
TAACAACCGT	XM_641608.1	90	40S ribosomal protein SA (*rpsA*)
GTCGTTCTTA	X00134.1	88	17S small subunit ribosomal RNA gene
GTGTTAACGA	XM_639972.1	84	ribosomal protein L15 (*rpl*15)
GGTACCAAAG	XM_636854.1, XM_642089.1	82	ribosomal protein L31 (*rpl*31)/putative myb transcription factor
CCACTACAAC	X15710.1, XM_639100.1	80	ribosomal protein L2/60S ribosomal protein L8 (*rpl*8)
TGAAATAAAA	XM_637148.1	79	extracellular matrix protein ST430 (13/13)
TGCTGTCTCT	XM_633409.1	79	glycyl-tRNA synthetase (*glys*) (12/12)
GTGTAATCCA	EAL63602	77	small aggregate formation (*smlA*)
TTTAAGTGGT	XM_642699.1	76	P17 protein
AAATATAAAA	XM_635590.1	72	hypothetical protein
CGCTGGTGCA	XM_631726.1	69	ribosomal protein L10E (*rpl*10)
ATCAAATCTG	XM_629345.1, U78756.1	68	40S ribosomal protein s3 (*rps*3), (*rpgG*)

Continued on next page

Table 1. Continued.

Dictyostelium tag sequence	ID	Count	Description
GAGAATCATC	XM_632378.1	68	hypothetical protein
GAAGTCGGTA	AY171066.1	64	extrachromosomal palindromic ribosomal RNA
GATGTCATCT	XM_639821.1	64	40S ribosomal protein S4 (*RPS*4)
CCTCTGGTAT	XM_640073.1	63	40S ribosomal protein S23 (*rps*23)
GAGAAAACCA	XM_640259.1	63	ribosomal protein L32 (*rpl*32)
AAAAATCAAA	XM_639860.2	62	putative polyketide synthase (*pks*15)
GATAAAAAAG	XM_637269.1	62	40S ribosomal protein S20 (*rps*20)
GTAAACCAAT	XM_638427.1	61	AX4 RNA polymerase I, largest subunit (*rpa*1)
CCAAAAGATA	XM_642485.1	59	histone H3 (*ddH*3)
GAAAAAGGTT	AF095929.1	59	actin related protein 2 (*arpB*)
GGTAATGGTA	XM_634924.1	59	cyclophilin (*ppiA*)
TCCCTATTAA	XM_632008.1	59	hypothetical protein
GCTGCTCACT	XM_636632.1	57	ribosomal protein L21 (*rpl*21)
GGGTGCTGAC	XM_631551.1	57	hypothetical protein
GTTGTAACGG	X00134.1	54	18S rRNA gene
ATGAAAGTGC	M25217.1	53	actin (*act*)
AATCACCCAA	XM_638028.1	52	40S ribosomal protein S18 (*rps*18)
GGCCACCGAA	XM_629870.1	52	40S ribosomal protein S8 (*rps*8)
TAAGATTGTT	X14970.1	52	translation initiation factor *eIF-4D*
CTTGTGAGTT	XM_642699.1	51	17-kDa protein (p17)
TTCAACTGTT	XM_637485.1	51	40S ribosomal protein S27 (*rps*27)
GTATGATAAT	XM_638765.1	50	glyceraldehyde-3-phosphate dehydrogenase (*gpdA*)
GTGAATCAAA	XM_635424.1	49	putative alpha-amylase (*amyA*)
TGTCGATCCA	XM_641310.1	49	hypothetical protein
TGCTACTATC	XM_642726.1	48	60S ribosomal protein L9 (*rpl*9)
TGGTGATGGT	XM_635474.1	46	PhoPQ-activated pathogenicity-related protein (*aprA*)
AAAGAAGATG	XM_642543.1	45	S-adenosyl-L-homocysteine hydrolase (*sahA*)
GCAGGAAAAA	XM_636572.1	45	hypothetical czl protein
AAAAATATAG	AB000109.1	44	mitochondrial DNA
CAGTAGCTAA	AY007805.1	44	ribosomal protein L36
CAGTTGATCC	XM_641577.1	44	gelation factor (*abpcC*)
CGAATGGAAT	XM_642251.1	44	actin binding protein (coronin) (*corA*)
AATTATTGAA	AF309947.1	43	Rac1A (*rac1A*)
AATTCTTAAT	XM_629919.1, XM_629706.1	43	ribosomal protein (L3) gene/AX4 aminomethyltransferase (*rpl*3)
CTGATAATGA	XM_642071.1, U23957.1	42	acetylornithine deacetylase (*argE*)/P52D
TACTCTTTCT	AF203735.1	42	culmination specific protein 45D (45D)
TTGACATTGA	M91383.1	42	thioredoxin (*trx*3)
AACTCTGGTC	XM_632836.1	41	40S ribosomal protein s5 (*rps*5)
GAACCAGCCT	XM_638827.1	41	hypothetical protein
GGATCCAGTG	XM_641936.1	41	hypothetical protein DDB_0190094 partial mRNA
AGTGTCGCTG	XM_631261.1	39	ADP-ribosylation factor 1
CCAACCCGTT	XM_640574.1	39	ribosomal protein L27 (*rpl*27)
GGGTAATAAT	XM_629547.1	39	discoidin II (*dscE*)
GTTATTGTAG	XM_638733.1	39	conditioned medium factor (*cmfA*)
GTTGGGAAGA	AY0440851.1	39	Rho GDP-dissociation inhibitor (*rdiA*)
CAAAACCCAG	U78759.1	38	IfdA translation initiation factor (*ifdA*)
GAAAGATGGA	XM_639860	38	putative polyketide synthase (*pks*15)
TCCATCATCA	XM_640805.1	37	EGF-like domain
TCCGCCTTAG	XM_632431.1	37	putative GATA-binding transcription factor (DDB_0216328) mRNA, complete cds
AAAATAATAA	XM_636161.1	36	YELA translation initiation factor (*yelA*)
AAAATTTACA	XM_001732912.1	36	putative polyketide synthase
AAAGTTAGAG	XM_629874.1	36	ribosomal protein L13 (*rpl*13)
GCATTTGGAA	XM_631261.1	36	ADP-ribosylation factor (*arfA*)
TGTAACAACA	XM_632548.1	36	putative actin binding protein
GTGTGGTGAA	M92992.1, XM_628883.1	35	cyclic AMP-regulated protein (p16) /actin binding protein

Continued on next page

Table 1. Continued.

Dictyostelium tag sequence	ID	Count	Description
TAAGAGTTTT	XM_631880.1	35	26S proteasome regulatory subunit S1 (*psmD1*)
AGTTTCAGCC	XM_640335.2	34	hypothetical protein
CCATCTTACA	XM_632592.1	34	histone H2B domain containing protein (*H2Bv3*)
GGGTCTTGAC	XM_636633.1	34	cysteine proteinase (*cprF*)/cathepsin O?
TACTTTACCC	XM_639670.1	34	40S ribosomal protein s30 (*rps*30)
GGAAATTTCT	XM_636553.1	33	cathepsin D (*ctsD*)
GGGTACATAT	XM_636202.1	33	hypothetical protein
GGTATGATGT	XM_638942.1	33	hypothetical protein
AAGTCATCGA	XM_630012.1	32	hypothetical protein (29C)
ACGAACTTTA	XM_633918.1	32	calreticlin (*crtA*)/*grp*94 mRNA for glucose-regulated protein 94
AGAGGTATCC	XM_631629.1	32	elongation factor 2
ATGAGCAGTT	EAL63335	32	mitochondrial DNA
CACACCAGTA	XM_634256.1	32	p34-Arc (*arpE*)
TAGATCAGAA	XM_634139.1	32	phosphopyruvate hydratase (*enoA*)
TCCCCGTACA	XM_633420.1	32	NOT2/NOT3/NOT5 family protein
TGTTCACCAT	XM_633215.1	32	ribosomal protein L17 (*spl*17)
GACCACAACG	XM_637999.1	31	40S ribosomal protein s15 (*rps*15a)/40S ribosomal protein S24
TTTGTCAAGA	XM_631966.1	31	alpha tubulin (*tuba*)
CAAATGTAGT	XM_633133.1	30	RING Zn finger-containing protein (DDB-0232321)
TAAAAGCTCT	EAL66985	30	p41-Arc (*arpD*)
TTTGGTTATG	XM_630291.1	30	S-adenosylmethionine synthetase (*metK*)
AAGAACAAGC	XM_641070.1	29	beta tubulin gene (*tubB*)
CAATTCGATA	XM_641945.1	29	vacuolar protein sorting-associated protein (*vps*26)
GGATGATGAA	XM_628865.1, U15926.1	29	putative polyketide synthase/elongation factor 1b (*pks*1)
GGATGGTGGT	XM_638832.1	29	acetyl-coA C-acyltransferase
GTCATTTCTT	XM_629407.1	29	putative protein tyrosine phosphatase, dual specificity (DDB_0238566) mRNA, complete cds
TGGATGTGAC	XM_001732949.1	29	putative polyketide synthase (*pks*)
AGCTGGGTTT	EAL63335	28	mitochondrial DNA
TGTTGCCCTC	XM_637474.1	28	60S ribosomal protein L7a (*rpl*7a)
AAGAAAACTC	XM_638053.1	27	ribosomal protein L7 (*rpl*7)
AGATTCAAAT	XM_631705.1	27	ribosomal protiein L12 (*rpl*12)
ATATGACCCA	XM_632900.1	27	type I phosphodiesterase nucleotide phosphatase family protein/ PkiA (*pkiA*)
CTAATCGTCA	XM_637723.1	27	ubiquitin (*ubqC*)
GATAAACAAA	XM_640344.1	27	porin (*porA*)
TTGAAACTCA	XM_640947.1	27	actinobindin (*abnB*)
AAATGTGATG	XM_641596.1, XM_640170.1	26	actin (*act*8)/actin (*act*16)
CAATGTTGAA	EAL71934	26	*PepA*
CTTGGACTAA	XM_633462.1	26	40S ribosomal protein s13 (*rps*13)
AAGAAAGCTG	XM_632648.1	25	myosin II heavy chain (*mhcA*)
GCGAATGCAC	XM_642513.1	25	hypothetical protein
TAAACGAATC	XM_638093.1	25	cytoplasmic dynein heavy chain (*dhcA*)
TCTCAAGATC	AY007804.1	25	ribosomal protein L28 (*rpl*28) 13/13
CTCGCTCAAA	XM_636770.1	24	30,000 dalton actin bundling protein (*abpB*)
GAGATAATCT	XM_631114.1	24	hypothetical protein
GAGTAAAGAT	XM_639850.1	24	CsbA (*csbA*) and CsbB (*csbB*)
GGCTCTCTCT	XM_638782.1	24	fructose-bisphosphate adolase (*fba*)
GTTTGTTTGG	XM_640382.1	24	EGF-like domain containing protein (*canC*)
TAATAAACTC	AC116986.1	24	chromosome 2 map
AAATCAAAAA	XM_001732910.1	23	putative polyketide synthase
CATTCGTAAA	XM_640374.1	23	proteosomal alpha-subunit 7-1 (*pntB*)
GCTACCGTCA	XM_630705.1	23	ribosomal protein L23 (*rpl*23)
GTTTCTAAAA	XM_641945.1	23	vacuolar protein sorting-associated protein
GTTTTGGTGG	XM_640757.1	23	hypothetical protein
TAAATAGATC	XM_630528.1	23	hypothetical protein
TCATTCGTTT	XM_638104.1	23	glutamine amidotransferase (*pyr*1-3)
TCATTCGTTT	XM_638093.1	23	cytoplasmic dynein heavy chain (*dhcA*)

Continued on next page

Table 1. Continued.

Dictyostelium tag sequence	ID	Count	Description
AAATCGCTTT	XM_636325.1	22	nucleoside diphosphate kinase (*ndkM*)
AAGATGATGA	XM_638093.1	22	heat shock protein
CAGCACTAAA	M91382.1	22	thioredoxin (*trx*2)
CGAATTTATG	XM_640151.1	22	actin capping protein CAP34 protein
GATTATCTTA	XM_632335.1	22	polyketide synthase (*pks*25)
GCTTTGGATG	XM_632592.1	22	histone H2B domain containing protein (*H2B*v3)
GTCACTTCTT	XM_639759.1	22	hypothetical protein
GTGAAGATAA	XM_641945.1	22	vacuolar protein sorting-associated protein (DDB_0234203) mRNA, complete cds
AATGCTTTCA	XM_632600.1	21	60S ribosomal acidic phosphoprotein P0
ATCATCAATA	XM_637516.1	21	vacuolar H+ ATPase B subunit (*vatB*)
CGCAAATGGT	XM_633239.1	21	NADH dehydrogenase (ubiquinone) (DDB_0233208)
CTGATCAAGA	XM_642037.1	21	superoxide dismutase (*sodA*)
GCTGTCGTCA	EAL63335	21	mitochondrial DNA
GGATCTGCTG	XM_630177.1	21	hypothetical protein
GTCAAGCAAA	XM_632222.2	21	hypothetical protein
GTGGTAGTAG	XM_642014.1	21	diaminopimelate epimerase (*dafE*)
GTTAGAGCCA	XM_641467.1	21	succinate dehydrogenase (ubiquinone)
TAAAAAAAAA	AC117072.3	21	chromosome 2 map
TAAATTGGAC	XM_637030.1	21	6-phosphogluconate dehydrogenase (*gnd*)
TAGAAGAAGA	XM_642384.1	21	non-transporter ABC protein (*abcF*4)
TATTGGTCAT	EAL63838	21	V-ATPase A subunit (*vatA*)
TGCTCTGTAG	XM_639875.1	21	major facilitator superfamily protein DDB_0238776
TGCTCTGTAG	XM_631290.1	21	myosin IB (*myoB*) mRNA
TTGCAGAAAT	XM_630145.1	21	hypothetical protein DDBdraft_0183916
AATTTGATGT	XM_639860.2	20	putative poyketide synthase (*pks*15)
CTAATGGGTC	XM_638179.1	20	hypothetical protein (DDB_0232153)
GCTAGAAATG	XM_632378.1	20	hypothetical protein (DDBdraft_0187197)
GTAATATTGT	XM_640523.2	20	BEACH domain containing protein (*lvsB*)
GTCTTCAAGA	XM_639159.1	20	colossin B mRNA (*colB*)
TTATTGATTA	XM_641945.1	20	vacuolar protein sorting-associated protein (DDB _0234203)
TTTTCAATAA	XM_639838.2	20	putative polyketide synthase (*pks*14)
AAAAACACAT	XM_642213.1	19	hyothetical protein DDBdraft_0216555
AAAAACACAT	XM_641971.1	19	hypothetical protein DDBdraft_0216640
AAAAACACAT	XM_642177.1	19	hypthetical protein DDBdraft_0216547
AAAAAGAAAT	EAL68664	19	clathrin light chain (*clcA*)
AGCAGAAGCA	XM_640867.2	19	hypothetical protein DDBdraft_0215165
ATGCTCCAAT	XM_001733034.1	19	hypothetical protein DDBdraft_0215064
ATGCTCCAAT	XM_628847.1	19	leucine-rich repeat-containing protein (*lrrA*)
GGAATTACCA	XM_632762.1	19	hypothetical protein DDBdraft_0215883
GTACTGAATT	EAL64547	19	crystal protein (*crysS*)
TAAAATTTTG	EAL71482	19	*rtoA*
TGGACCAAAG	XM_633814.1	19	hypothetical protein DDB_0231746
CTCGTGTCTA	XM_629958.1	18	cystatin A1 (*cpiA*) mRNA, complete cds
CTGCTATTTC	XM_632737.1	18	hypothetical protein DDBdraft_0186902
GAGAACTTAT	XM_632346.1	18	SNF7 family protein (*vps*20) mRNA, complete cds
GGTTGGTAGT	XM_631716.1	18	leucine-rich repeat-containing protein (*lrr*) (*roco*4)
GTGATCTGTA	XM_637067.1	18	hypothetical protein DDBdraft_0204455
GTGATCTGTA	XM_635945.1	18	hypothetical protein DDBdraft_0204740
GTGGTGGGAA	XM_640047.1	18	putative small nuclear ribonucleoparticle-associated protein DDB_0233178
GTTGGTACTC	XM_631286.1	18	ankyrin repeat-containing protein DDB_0232969
TAAAACAATT	XM_639838.2	18	putative polyketide synthase (*pks*15)
TACCAAAATG	XM_640812.1	18	hypothetical protein DDBdraft_0201696
TGCTAACCTT	XM_632744.1	18	alkyl-dihydroxyacetonephosphate synthase (*agps*)
TGCTAACCTT	XM_636773.1	18	hypothetical protein DDBdraft_0205035
TTGCTTCTGA	XM_639860.2	18	putative polyketide synthase (*pks*15)
AAGACACCAG	XM_639730.1	17	heat shock protein (*hspE*-1) mRNA, complete cds
AAGACACCAG	XM_639336.1	17	heat shock protein (*hspE*-2) mRNA, complete cds

Continued on next page

Table 1. Continued.

Dictyostelium tag sequence	ID	Count	Description
AAGATTAAAG	XM_641717.1	17	putative protein serine/threonine kinase DDB_0230126 partial mRNA
CACAAGAAAC	XM_633628.1	17	P-type ATPase (*atp*7a)
CACACCAACA	XM_638837.1	17	hypothetical protein DDBdraft_0217482
GGAAACCTTT	XM_639002.1	17	hypothetical protein DDBdraft_0217527
GGAGGTATGC	XM_641119.1	17	WH2 domain-containing protein DDB_0231673
GGTGCCATCG	XM_638855.1	17	amidophosphoribosyltransferase (*purF*)
GTAAATCATT	XM_001134471.1	17	annexin VII (*nxnA*)
TAACAAAGAA	XM_640523.2	17	hypothetical protein BEACH domain-containing protein (*lvsB*) mRNA
TCAAATGCTG	XM_641218.1	17	major vault protein alpha (*mvp*-alpha)
TGAAAAAAAA	AF482390.1	17	transporter AbcG11 (*abcG*11)
TGCAGTTTTA	XM_632827.2	17	adenosine kinase (*adk*)
TGGACTAAAT	XM_638093.1	17	hypothetical protein cytoplasmic dynein heavy chain (*dhcA*)
AAGCACAGGT	XM_630148.1	16	hypothetical protein RING Zn finger-containing protein (*vps*8) mRNA, complete
AATGGATCAA	AY171066.1	16	extrachromosomal palindromic ribosomal RNA
ACTAATCAAT	XM_638814.1	16	hypothetical protein
ATCAATGGGT	XM_639599.1	16	discoidin-I D chain (*dscD*)
ATGATGATGA	XM_633148.1	16	nucleotide binding protein 1-like protein
CAACCTCTGG	XM_629574.1	16	hypothetical protein colossin A (*colA*) mRNA, complete cds
CAACTCAATA	XM_001134515.1	16	severin (*sevA*)
CGATTCAATT	XM_632550.1	16	putative actin binding protein
CGTAGAAAGC	XM_635431.1	16	hypothetical protein
CTTTCCGTGC	L46371	16	cortexillin II
GATGATGTAC	XM_636206.1	16	hypothetical protein
GTCTCTCCAG	XM_632040.1	16	RasGTPase-activating protein
GTGAAAACAA	XM_635685.1	16	hypothetical protein
GTTCACTCCT	XM_633237.1	16	fatty acid desaturase (*fadA*)
TCAATTCGAT	XM_633521.1	16	ERG4/ERG24 ergosterol biosynthesis protein family protein (DDB_0232079)
TGAAGATGGT	XM_633121.1	16	hypothetical protein
TGGAAATGGT	XM_637070.1	16	EGF-like domain-containing protein
AAAACCATTG	XM_638265.1	15	protein disulfide isomerase (*dis*)
AAACCGTCAA	XM_633574.1	15	40S ribosomal protein S12 (*rps*12)
AAGAGTGGTA	XM_641660.1	15	ribosomal protein L30 (*rpl*30)
ATAGAATGAG	XM_630305.1	15	penta EF hand calcium binding protein (*pefA*)/ apoptosis-linked gene 2 (*alg*2A)
ATGATGAAGC	XM_629890.1	15	vacuolar proton ATPase 100-kDa subunit (*vatM*)
ATGTTTGGTG	XM_639226.1	15	hypothetical protein
ATTGAGAGAG	XM_642726.1	15	60S ribosomal protein L9 (*rpl*9)
ATTGCTTCTT	XM_634949.1	15	putative actin binding protein (DDB_0232396)
CCGGATCACT	AY171065S2	15	extrachromosomal palindromic ribosomal RNA
GAAAATGGCC	XM_638890.1	15	hypothetical protein peptidase C19 family protein (DDB_0237738) mRNA
GAATTTTAAA	XM_639838.2	15	hypothetical protein putative polyketide synthase (*pks*14) mRNA, complete cds
GGAAAATACA	XM_641146.2	15	hypothetical protein
GTCACCACTA	XM_633438.1	15	pleckstrin homology (PH) domain-containing protein (*gxcQ*) mRNA, complete cds
GTGGTAGTTA	AC114263	15	chromosome 2 map
GTTGATGGGG	XM_641596.1, XM_640170.1	15	actin (*act*8)/actin (*act*16)
TAAAAACAAC	XM_638317.1	15	13/13 natural resistance-associated macrophage protein (*nramp*)
TAAAGTTAGA	XM_632583.1	15	hypothetical protein vacuolar protein sorting-associated protein mRNA
TATCTCAAGA	XM_629630.1	15	26S proteosome subunit ATPase 5 (*psmC*5)/HIV1 TAT-binding protein homologue (*tbp*1O)
TATTTGATAC	XM_633831.1	15	mMob1-like protein
AGATCATCTC	XM_642439.1	14	rab GTPase (*rab*6)

Continued on next page

Table 1. Continued.

Dictyostelium tag sequence	ID	Count	Description
ATGATCGTGG	XM_635251.1, U13671.1	14	actin binding protein (*hatB*)/hisactophilin II (*hsII*)
ATGCTGTCAT	XM_001134481.1	14	hypothetical protein centrosomal protein CP224 (*mtaA*) mRNA, complete cds
CAAGTCGATA	XM_630022.1	14	GlcNAc transferase (*gnt*2)/putative phosphatidylinositol 4-phosphate 5-kinase
CTGGTCTTGT	XM_631976.1	14	hypothetical protein
GAGAAGAAGA	XM_638794.1	14	pleckstrin homology (PH) domain-containing protein
GATTGTCCAG	XM_635847.1	14	ribonuclease T2 (*ddiA*)
GCCGTGATGA	XM_632857.2	14	hypothetical protein
GGCCGTACAA	XM_641229.1	14	acyl-CoA binding protein (*acbA*)
GGGTATCACT	XM_633934.1	14	hypothetical protein
GGGTATCGCA	EAL73180	14	ADP/ATP translocase (*ancA*)
GTGCAAATTG	XM_641171.1	14	putative countin receptor Cnr9 (*cnrI*)
GTTGACACTC	XM_633034.1	14	40S ribosomal protein S15 (*rps*15)
TAAAAATATT	XM_631437.1	14	putative protein serine/threonine kinase (*irlE*)/LTR-retrotransposon Skipper
TCATTCACTA	XM_633173.1	14	hypothetical protein
TCCGGTGTCA	XM_638802.1	14	catalase (*catA*)
TGAAATATAA	XM_640620.1	14	13/13 LISK family protein kinase
TGATGGTGGT	XM_633438.1	14	hypothetical protein pleckstrin homology (PH) domain containing protein (*GxcQ*) mRNA
TGATGTTGCC	XM_631158.1, U67940	14	P14 domain-containing protein (*rcdGG*)ORFveg106
TGGACCTGGA	XM_631420.1	14	cyclase associated protein (*cap*)
TGGATAGATA	XM_633869.1	14	proteosome subunit beta type 7 (*psmB*7)
TGGTTCACCA	XM_633383.1	14	mannose-6-phosphate isomerase (*mpi*)/countin
TGTGGTCCAC	XM_638093.1	14	hypothetical protein cytoplasmic dynein heavy chain (*dhcA*)
TTCAATGTAT	XM_001134613.1	14	ubiquinol-cytochrome-c reductase hinge protein (*clec6A*)
TTGCCAATGG	XM_641284.1	14	CD9 (*cdp*9)
AAAATGGTAT	EAL71195	13	5'AMP activated gamma subunit (*prkag*)
AACTAAATTA	AC116956.2	13	chromosome 2 map
AAGATACTGA	XM_633862.1	13	phosphoprotein phosphatase A (*pppA*)
AATAATGAGT	XM_639860.2	13	hypothetical protein putative polyketide synthesis (*pks*15)
AATTACCTAC	XM_631966.1	13	alpha tubulin (*tubA*)
AATTTCAGAC	AY953941.1, AY953940.1	13	U1B snRNA/U1A snRNA
AGAGATTGGA	XM_629514.1	13	malate dehydrogenase (*mdhB*)
ATCATAACCA	XM_641793.1	13	hypothetical protein
ATGACATCGG	V00192.1	13	5.8S ribosomal RNA
ATGGATGAAT	XM_640496.1	13	13/13 RING Zn finger-containing protein (*dg*1106)
ATGGTTTCGG	XM_001134587.1	13	4-hydroxyphenylpyruvate dioxygenase
CAAAAATATA	XM_633368.1	13	hypothetical protein
CATATAATAT	AB007024.1	13	DAPS-1 mRNA for proteosome subunit
CGCATAATCA	XM_640523.2	13	BEACH domain-containing protein (*lvsB*)
GAGAACTGCA	XM_638794.1	13	pleckstrin homology (PH) domain-containing protein
GTAAACCACC	EAL63602	13	small aggregate formation (*smlA*)
GTAGAAAAAT	XM_637377.1	13	valyl tRNA synthetase (*valS*1)
GTAGAACCTC	XM_630153.1	13	13/13 hydroxyacylglutathione hydrolase (*gloB*2)
GTCATCCAAC	AF482388.1	13	ABC transporter AbcG9 (*abcG*9)
GTCGTGATAC	XM_639901.1	13	seryl-tRNA synthetase (*serS*)
GTGAAAAGGA	XM_641077.1	13	alpha-glucosidase II (*modA*)
GTGTTGGTTC	XM_633060.1	13	hypothetical protein
GTTAGGTGAT	XM_630798.3	13	FNIP repeat-containing protein DDB_0238003
GTTGTGTGTA	XM_637503.1	13	hypothetical protein DDBdraft_0203368
GTTTCGTTTC	XM_630703.1	13	hypothetical protein DDBdraft_0188832
TGAAAGATTA	XM_632907.1	13	LIM domain-containing protein DDB_0233145
TTAATTGGTA	EAL61182	13	*rab*4

Ribosome-related genes and other RNAs

Genes that produce components of the ribosome comprise 20% of the amoeba transcriptome (Table 2). This category includes the third most highly expressed gene, 40S ribosomal protein S29, and 11 of the 28 genes expressed at the level of 100 copies or higher. Forty-nine different ribosomal proteins, initiation and elongation factors, several rRNA genes, and a potential regulator of RNA synthesis (NOT2/NOT3/NOT5 family protein) were also present at moderate to high expression levels. These data suggest that protein synthesis is a major function of amoebae.

Table 2. Ribosome-related genes and other RNA molecules of *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
TGCAGACGTT	XM_640038.1	308	40S ribosomal protein S29 (*rps*29)
CTGGTTACTC	XM_64088.1	168	protein synthesis elongation factor 1-alpha (*tef*2)
GAAAGACGTC	XM_15384.1	168	V12 mRNA/ribosomal protein 1024 (*rps*9)
GAAAAGAGAA	XM_635906.1	155	40S ribosomal protein S6 (*rps*6)
CTCGTCCTTA	XM_636364.1	154	ribosomal protein L18 (*rpl*18)
GATATTGAAG	XM_637875.1	151	ribosomal protein L10a (*rpl*10a)
TGTAGAGGTG	XM_637392.1	139	60S ribosomal protein L4 (*rpl*4)
CGTAAACCAG	X15383.1	122	Ribosomal protein (*rp*119) vegetative specific V14 gene
GTGGTTTGAA	XM_639670.1	112	40S ribosomal protein S30 (*rps*30)
GGGTAAGATT	XM_641702.1	110	ribosomal protein L35a (*rpl*35a)
GCCACTTTCT	XM_628965.1	100	ribosomal protein s2 (*rps*2)/LLRep3
TGCTATCCAC	XM_635481.1	95	40S ribosomal protein S26 (*rps*26)
GCCGCTCAAG	XM_630185.1	94	40S ribosomal protein S14 (*rps*14)
TGGTTACAAG	XM_641274.1	94	ribosomal protein L36a (*rpl*36a)
AGAGGTCTCA	XM_636637.1	93	vegetative specific gene V18 gene for ribosomal protein/ribosomal protein L11 (*rpl*11)
GTCGTCCAGG	XM_633034.1	92	40S ribosomal protein S15 (*rps*15)
TAACAACCGT	XM_641608.1	90	40S ribosomal protein SA (*rps*A)
GTCGTTCTTA	X00134.1	88	17S small subunit ribosomal RNA gene
GTGTTAACGA	XM_639972.1	84	ribosomal protein L15 (*rpl*15)
GGTACCAAAG	XM_636854.1, XM_642089.1	82	ribosomal protein L31 (*rpl*31)/putative myb transcription factor
CCACTACAAC	X15710.1, XM_639100.1	80	ribosomal protein L2/60S ribosomal protein L8 (*rpl8*)
CGCTGGTGCA	XM_631726.1	69	ribosomal protein L10E (*rpl*10)
ATCAAATCTG	XM_629345.1, U78756.1	68	40S ribosomal protein s3 (*rps*3), (*rpgG*)
GAAGTCGGTA	AY171066.1	64	extrachromosomal palindromic ribosomal RNA
GATGTCATCT	XM_639821.1	64	40S ribosomal protein S4 (*RPS*4)
CCTCTGGTAT	XM_640073.1	63	40S ribosomal protein S23 (*rps*23)
GAGAAAACCA	XM_640259.1	63	ribosomal protein L32 (*rpl*32)
GATAAAAAAG	XM_637269.1	62	40S ribosomal protein S20 (*rps*20)
GTAAACCAAT	XM_638427.1	61	AX4 RNA polymerase I, largest subunit (*rpa*1)
GCTGCTCACT	XM_636632.1	57	ribosomal protein L21 (*rpl*21)
GTTGTAACGG	X00134.1	54	18S rRNA gene
AATCACCCAA	XM_638028.1	52	40S ribosomal protein S18 (*rps*18)
GGCCACCGAA	XM_629870.1	52	40S ribosomal protein S8 (*rps*8)
TAAGATTGTT	X14970.1	52	translation initiation factor *eIF-4D*
TTCAACTGTT	XM_637485.1	51	40S ribosomal protein S27 (*rps*27)
TGCTACTATC	XM_642726.1	48	60S ribosomal protein L9 (*rpl*9)
CAGTAGCTAA	AY007805.1	44	ribosomal protein L36
AATTCTTAAT	XM_629919.1, XM_629706.1	43	ribosomal protein (L3) gene/AX4 aminomethyltransferase (*rpl*3)
AACTCTGGTC	XM_632836.1	41	40S ribosomal protein s5 (*rps*5)
CCAACCCGTT	XM_640574.1	39	ribosomal protein L27 (*rpl*27)
CAAAACCCAG	U78759.1	38	IfdA translation initiation factor (*ifdA*)
TCCGCCTTAG	XM_632431.1	37	putative GATA-binding transcription factor (DDB_0216328) mRNA, complete cds
AAAATAATAA	XM_636161.1	36	YELA translation initiation factor (*yelA*)
AAAGTTAGAG	XM_629874.1	36	ribosomal protein L13 (*rpl*13)
TACTTTACCC	XM_639670.1	34	40S ribosomal protein s30 (*rps*30)
AGAGGTATCC	XM_631629.1	32	elongation factor 2
TCCCCGTACA	XM_633420.1	32	NOT2/NOT3/NOT5 family protein

Continued on next page

Table 2. Continued.

Dictyostelium tag sequence	ID	Count	Description
TGTTCACCAT	XM_633215.1	32	ribosomal protein L17 (*spl*17)
GACCACAACG	XM_637999.1	31	40S ribosomal protein s15 (*rps*15a)/40S ribosomal protein S24
TGTTGCCCTC	XM_637474.1	28	60S ribosomal protein L7a (*rpl*7a)
AAGAAAACTC	XM_638053.1	27	ribosomal protein L7 (*rpl*7)
AGATTCAAAT	XM_631705.1	27	ribosomal protiein L12 (*rpl*12)
CTTGGACTAA	XM_633462.1	26	40S ribosomal protein s13 (*rps*13)
TCTCAAGATC	AY007804.1	25	ribosomal protein L28 (*rpl*28) 13/13
GCTACCGTCA	XM_630705.1	23	ribosomal protein L23 (*rpl*23)
AATGCTTTCA	XM_632600.1	21	60S ribosomal acidic phosphoprotein P0
GTGGTGGGAA	XM_640047.1	18	putative small nuclear ribonucleoparticle-associated protein DDB_0233178
AATGGATCAA	AY171066.1	16	extrachromosomal palindromic ribosomal RNA
AAACCGTCAA	XM_633574.1	15	40S ribosomal protein S12 (*rps*12)
AAGAGTGGTA	XM_641660.1	15	ribosomal protein L30 (*rpl*30)
ATTGAGAGAG	XM_642726.1	15	60S ribosomal protein L9 (*rpl*9)
CCGGATCACT	AY171065S2	15	extrachromosomal palindromic ribosomal RNA
GTTGACACTC	XM_633034.1	14	40S ribosomal protein S15 (*rps*15)
AATTTCAGAC	AY953941.1, AY953940.1	13	U1B snRNA/U1A snRNA
ATGACATCGG	V00192.1	13	5.8S ribosomal RNA

Table 2 includes the RNA polymerase I large subunit, expressed at moderate levels (61 copies), suggesting an active transcription rate. Table 2 also includes a few other types of RNA molecules such as extrachromosomal palindromic ribosomal RNA and the spliceosome components U1B/U1A snRNA, which are expressed at low levels.

Enzymes and components of basic metabolism

Numerous genes in the amoeba-stage transcriptome (18%) produce the enzymes, regulatory molecules, modifying enzymes, and structural components of basic metabolism (Table 3). This group includes three of the 28 genes expressed at more than 100 copies, including the fourth-highest expressed gene in the transcriptome, a putative glycoside hydrolase.

The remainder of the genes in this group were expressed at moderate to low levels. Components of glycolysis, the tricarboxylic acid cycle, fatty acid oxidation, and the electron transport chain were found, as well as six different kinases and three different phosphatases. While the electron transport chain components reside in the mitochondria, we included them in the metabolism table for clarity.

Several biosynthetic enzymes were also expressed, including 13 different polyketide synthetases, and three t-RNA synthetases. While basic-metabolism proteins comprised a large group of expressed genes, the moderate level of expression suggests that these are basic maintenance functions of amoebae.

Actin, actin-binding, and cytoskeleton-related genes

With the exception of the hisactophilin II gene, the 26 actin-related genes included in Table 4 were expressed at moderate to low levels. They represent 8% of the transcriptome. These genes represent all of the major categories of cytoskeleton components, including actin-binding, capping, severing, cross-linking, and membrane associating, essential for maintaining amoeba shape, and allowing motility (Eichinger et al., 2005). Two myosin genes, mhcA and myoB, were included in this category due to their predicted role in chemotaxis (Postma et al., 2004).

Table 3. Enzymes and components of basic metabolism in *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
CAAAGACTTT	XM_640564.1	297	AX4 putative glycoside hydrolase/*Mus musculus* zinc finger protein 119 (*Zfp*119)
TGAAATATAT	XM_001732911.1	104	putative polyketide synthase (*pks*)
GGGTGATAAT	XM_633636.1	103	histidine kinase (dhkG) mRNA, complete cds
TGCTGTCTCT	XM_633409.1	79	glycyl-tRNA synthetase (*glys*) (12/12)
AAAAATCAAA	XM_639860.2	62	putative polyketide synthase (*pks*15)
GGTAATGGTA	XM_634924.1	59	cyclophilin (*ppiA*)
GTATGATAAT	XM_638765.1	50	glyceraldehyde-3-phosphate dehydrogenase (*gpdA*)
GTGAATCAAA	XM_635424.1	49	putative alpha-amylase (*amyA*)
AAAGAAGATG	XM_642543.1	45	S-adenosyl-L-homocysteine hydrolase (*sahA*)
CTGATAATGA	XM_642071.1, U23957.1	42	acetylornithine deacetylase (*argE*)/P52D
TTGACATTGA	M91383.1	42	thioredoxin (*trx*3)
GAAAGATGGA	XM_639860	38	putative polyketide synthase (*pks*15)
AAAATTTACA	XM_001732912.1	36	putative polyketide synthase
TAGATCAGAA	XM_634139.1	32	phosphopyruvate hydratase (*enoA*)
TTTGGTTATG	XM_630291.1	30	S-adenosylmethionine synthetase (*metK*)
GGATGATGAA	XM_628865.1, U15926.1	29	putative polyketide synthase/elongation factor 1b (*pks*1)
GGATGGTGGT	XM_638832.1	29	acetyl-CoA C-acyltransferase
GTCATTTCTT	XM_629407.1	29	putative protein tyrosine phosphatase, dual specificity (DDB_0238566) mRNA, complete cds
TGGATGTGAC	XM_001732949.1	29	putative polyketide synthase (*pks*)
ATATGACCCA	XM_632900.1	27	type I phosphodiesterase nucleotide phosphatase family protein/PkiA (*pkiA*)
GGCTCTCTCT	XM_638782.1	24	fructose-bisphosphate adolase (*fba*)
AAATCAAAAA	XM_001732910.1	23	putative polyketide synthase
TCATTCGTTT	XM_638104.1	23	Glutamine amidotransferase (*pyr*1-3)
AAATCGCTTT	XM_636325.1	22	nucleoside diphosphate kinase (*ndkM*)
GATTATCTTA	XM_632335.1	22	polyketide synthase (*pks*25)
CAGCACTAAA	M91382.1	22	thioredoxin (*trx*2)
CGCAAATGGT	XM_633239.1	21	NADH dehydrogenase (ubiquinone) (DDB_0233208)
GTGGTAGTAG	XM_642014.1	21	diaminopimelate epimerase (*dafE*)
GTTAGAGCCA	XM_641467.1	21	succinate dehydrogenase (ubiquinone)
TAAATTGGAC	XM_637030.1	21	6-phosphogluconate dehydrogenase (*gnd*)
AATTTGATGT	XM_639860.2	20	putative poyketide synthase (*pks*15)
TTTTCAATAA	XM_639838.2	20	putative polyketide synthase (*pks*14)
TAAAACAATT	XM_639838.2	18	putative polyketide synthase (*pks*15)
TGCTAACCTT	XM_632744.1	18	alkyl-dihydroxyacetonephosphate synthase (*agps*)
TTGCTTCTGA	XM_639860.2	18	putative polyketide synthase (*pks*15)
AAGATTAAAG	XM_641717.1	17	putative protein serine/threonine kinase DDB_0230126 partial mRNA
GGTGCCATCG	XM_638855.1	17	amidophosphoribosyltransferase (*purF*)
TGCAGTTTTA	XM_632827.2	17	adenosine kinase (*adk*)
GTTCACTCCT	XM_633237.1	16	fatty acid desaturase (*fadA*)
TCAATTCGAT	XM_633521.1	16	ERG4/ERG24 ergosterol biosynthesis protein family protein (DDB_0232079)
AAAACCATTG	XM_638265.1	15	protein disulfide isomerase (*dis*)
GAATTTTAAA	XM_639838.2	15	hypothetical protein putative polyketide synthase (*pks*14) mRNA, complete cds
CAAGTCGATA	XM_630022.1	14	GlcNAc transferase (*gnt2*)/putative phosphatidylinositol 4-phosphate 5-kinase
GATTGTCCAG	XM_635847.1	14	ribonuclease T2 (*ddiA*)
GGCCGTACAA	XM_641229.1	14	acyl-CoA binding protein (*acbA*)
TAAAAATATT	XM_631437.1	14	putative protein serine/threonine kinase (*irlE*)/LTR-retrotransposon Skipper
TCCGGTGTCA	XM_638802.1	14	catalase (*catA*)
TGAAATATAA	XM_640620.1	14	13/13 LISK family protein kinase
TTCAATGTAT	XM_001134613.1	14	ubiquinol-cytochrome-c reductase hinge protein (*clec6A*)
AAGATACTGA	XM_633862.1	13	phosphoprotein phosphatase A (*pppA*)
AATAATGAGT	XM_639860.2	13	hypothetical protein putative polyketide synthesis (*pks*15)
AGAGATTGGA	XM_629514.1	13	malate dehydrogenase (*mdhB*)
ATGGTTTCGG	XM_001134587.1	13	4-hydroxyphenylpyruvate dioxygenase
GTAGAAAAAT	XM_637377.1	13	valyl tRNA synthetase (*valS*1)
GTAGAACCTC	XM_630153.1	13	13/13 hydroxyacylglutathione hydrolase (*gloB*2)
GTCGTGATAC	XM_639901.1	13	seryl-tRNA synthetase (*serS*)
GTGAAAAGGA	XM_641077.1	13	alpha-glucosidase II (*modA*)

Table 4. Actin, actin-binding, and cytoskeleton-related genes in *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
CTACTTTCGA	XM_635251.1, U13671.1	138	actin binding protein (*hat B*)/hisactophilin II (*hsI*I)
AAAATTAAAA	XM_641596.1	94	actin (*act*8)
AAAATTAAAA	XM_640170.1	94	actin 16 (*act*16)
TTTAAGTGGT	XM_642699.1	76	P17 protein
GAAAAAGGTT	AF095929.1	59	actin related protein 2 (*arpB*)
ATGAAAGTGC	M25217.1	53	actin (*act*)
CTTGTGAGTT	XM_642699.1	51	17-kDa protein (p17)
CAGTTGATCC	XM_641577.1	44	gelation factor (*abpcC*)
CGAATGGAAT	XM_642251.1	44	actin binding protein (coronin) (*corA*)
TGTAACAACA	XM_632548.1	36	putative actin binding protein
GTGTGGTGAA	M92992.1, XM_628883.1	35	cyclic AMP-regulated protein (p16)/actin binding protein
CACACCAGTA	XM_634256.1	32	p34-Arc (*arpE*)
TAAAAGCTCT	EAL66985	30	p41-Arc (*arpD*)
TTGAAACTCA	XM_640947.1	27	actinobindin (*abnB*)
AAATGTGATG	XM_641596.1, XM_640170.1	26	actin (*act*8)/actin (*act*16)
AAGAAAGCTG	XM_632648.1	25	myosin II heavy chain (*mhcA*)
CTCGCTCAAA	XM_636770.1	24	30,000 dalton actin bundling protein (*abpB*)
CGAATTTATG	XM_640151.1	22	actin capping protein CAP34 protein
TGCTCTGTAG	XM_631290.1	21	myosin IB (*myoB*) mRNA
GGAGGTATGC	XM_641119.1	17	WH2 domain-containing protein DDB_0231673
GTAAATCATT	XM_001134471.1	17	annexin VII (*nxnA*)
CAACTCAATA	XM_001134515.1	16	severin (*sevA*)
CGATTCAATT	XM_632550.1	16	putative actin binding protein
CTTTCCGTGC	L46371	16	cortexillin II
ATTGCTTCTT	XM_634949.1	15	putative actin binding protein (DDB_0232396)
GTTGATGGGG	XM_641596.1, XM_640170.1	15	actin (*act*8)/actin (*act*16)
ATGATCGTGG	XM_635251.1, U13671.1	14	actin binding protein (*hatB*)/hisactophilin II (*hsII*)
TGGACCTGGA	XM_631420.1	14	cyclase associated protein (*cap*)

The endosomal system and membrane transport

Components of the endosomal system, including receptor mediated endocytosis, lysosomal enzymes, trafficking and sorting molecules, protein folding systems, heat shock proteins, and several membrane transport systems, are described in Table 5. This group of genes represents 9% of the transcriptome, and includes three of the 28 genes expressed at more than 100 copies, two of which are lysosomal enzymes (cysteine proteinases CP4 and CP5). The remainder of the genes in this category were expressed at relatively low levels. Six of the genes are cysteine proteinases, reflecting the phagocytic nature of amoeba nutrition. One of the genes, cystatinA1, is an inhibitor of cysteine proteases. ADP-ribosylation factor 1 is included in this list because it is most likely to be associated with the Golgi apparatus and vesicle trafficking (Venkateswarlu et al., 2007; Weimer et al., 2008). The membrane transporters expressed in amoebae include two ABC transporters (abcG9 and abcG11) and one P-type ATPase that move ions and molecules across the plasma membrane, and components for one V-type ATPase (vatA and vatB), which probably maintains acidity in vacuoles. All of these transporters were expressed at low levels in the amoebae.

Table 5. The endosomal system components and membrane transport genes in *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
GGGTATGGAT	XM_636628.2	195	cysteine proteinase (*CP*4)
TGGTTATCCA	XM_637389.2	142	*D. discoideum* AX4 putative transport protein (sec24l) mRNA, complete cds
GGGTATCGAA	XM_639885.1	140	cysteine proteinase CP5 (*cprE*)
AGTGTCGCTG	XM_631261.1	39	ADP-ribosylation factor 1
GGGTCTTGAC	XM_636633.1	34	cysteine proteinase (*cprF*)/cathepsin O?
GGAAATTTCT	XM_636553.1	33	cathepsin D (*ctsD*)
ACGAACTTTA	XM_633918.1	32	calreticlin (*crtA*)/*grp*94 mRNA for glucose-regulated protein 94
CAATTCGATA	XM_641945.1	29	vacuolar protein sorting-associated protein (*vps*26)
CAATGTTGAA	EAL71934	26	*PepA*
GTTTCTAAAA	XM_641945.1	23	vacuolar protein sorting-associated protein
AAGATGATGA	XM_638093.1	22	heat shock protein
GTGAAGATAA	XM_641945.1	22	vacuolar protein sorting-associated protein (DDB_0234203) mRNA, complete cds
ATCATCAATA	XM_637516.1	21	vacuolar H+ ATPase B subunit (*vatB*)
TATTGGTCAT	EAL63838	21	V-ATPase A subunit (*vatA*)
GTAATATTGT	XM_640523.2	20	BEACH domain containing protein (*lvsB*)
TTATTGATTA	XM_641945.1	20	vacuolar protein sorting-associated protein (DDB _0234203)
AAAAAGAAAT	EAL68664	19	clathrin light chain (*clcA*)
GTACTGAATT	EAL64547	19	crystal protein (*cryS*)
TAAAATTTTG	EAL71482	19	*rtoA*
CTCGTGTCTA	XM_629958.1	18	cystatin A1 (*cpiA*) mRNA, complete cds
GAGAACTTAT	XM_632346.1	18	SNF7 family protein (*vps*20) mRNA, complete cds
AAGACACCAG	XM_639730.1	17	heat shock protein (*hspE*-1) mRNA, complete cds
AAGACACCAG	XM_639336.1	17	heat shock protein (*hspE*-2) mRNA, complete cds
TAACAAAGAA	XM_640523.2	17	hypothetical protein BEACH domain-containing protein (*lvsB*) mRNA
TGAAAAAAAA	AF482390.1	17	transporter AbcG11 (*abcG*11)
AAGCACAGGT	XM_630148.1	16	hypothetical protein RING Zn finger-containing protein (*vps*8) mRNA, complete cds
ATGATGAAGC	XM_629890.1	15	vacuolar proton ATPase 100-kDa subunit (*vatM*)
GAAAATGGCC	XM_638890.1	15	hypothetical protein peptidase C19 family protein (DDB_0237738) mRNA
TAAAAACAAC	XM_638317.1	15	13/13 natural resistance-associated macrophage protein (*nramp*)
TAAAGTTAGA	XM_632583.1	15	hypothetical protein vacuolar protein sorting-associated protein mRNA
CGCATAATCA	XM_640523.2	13	BEACH domain-containing protein (*lvsB*)
GTCATCCAAC	AF482388.1	13	ABC transporter AbcG9 (*abcG*9)

Small G-proteins

Thirteen genes in the transcriptome (4%) are G-proteins and their interacting partners (Table 6). The moderately high expression genes are small G-proteins that are most likely involved in motility via linkage with the actin cytoskeleton, such as Rac1A, and endosomal vesicle traffic, such as ADP-ribosylation factor, rab GTPase and rho GDP-dissociation inhibitor (Postma et al., 2004). The ras GTPase, which may be involved in transmembrane signal transduction, was expressed at relatively low levels.

Table 6. Small G-proteins in *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
GGATAACAAA	EAL69803	256	G beta-like protein (*gpbB*)
AATTATTGAA	AF309947.1	43	Rac1A (*rac1A*)
GTTGGGAAGA	AY0440851.1	39	Rho GDP-dissociation inhibitor (*rdiA*)
GCATTTGGAA	XM_631261.1	36	ADP-ribosylation factor (*arfA*)
ATGCTCCAAT	XM_628847.1	19	leucine-rich repeat-containing protein (*lrrA*)
GGTTGGTAGT	XM_631716.1	18	leucine-rich repeat-containing protein (*lrr*) (*roco*4)
GTCTCTCCAG	XM_632040.1	16	RasGTPase-activating protein
GTCACCACTA	XM_633438.1	15	pleckstrin homology (PH) domain-containing protein (*gxcQ*) mRNA, complete cds
AGATCATCTC	XM_642439.1	14	rab GTPase (*rab*6)
GAGAAGAAGA	XM_638794.1	14	pleckstrin homology (PH) domain-containing protein
TGATGGTGGT	XM_633438.1	14	hypothetical protein pleckstrin homology (PH) domain containing protein (*GxcQ*) mRNA
GAGAACTGCA	XM_638794.1	13	pleckstrin homology (PH) domain-containing protein
TTAATTGGTA	EAL61182	13	*rab*4

Four pleckstrin homology (PH) domain genes, all expressed at low levels, are included in this category. Originally identified as proteins that bind to phosphatidylinositols, PH

domains can also bind small G-proteins, such as Rho, Rac and ARF (Lemmon, 2004). Since the amoeba transcriptome does not contain any phosphatidylinositol metabolizing enzymes, it seems more likely that the PH domain proteins are associated with the small G-proteins in amoebae that are associated with cytoskeleton remodeling or vesicle traffic.

The most unusual member of this category is a G-beta-like protein, which is the sixth-most-highly expressed gene in the transcriptome. No other potential heterotrimeric G-protein component is expressed in the transcriptome, so the function of this gene product in amoebae is difficult to ascertain.

Ubiquitin and proteosome-related

Five components of the proteosome and two ubiquitin ligases were expressed at moderate to low levels (Table 7). Ubiquitin was the second-highest expressed gene in the transcriptome. A single tag sequence is contained within five different predicted ubiquitin genes in *Dictyostelium*. We could not distinguish which ubiquitin gene(s) is (are) the source of this RNA. The wide range of processes that require ubiquitinated proteins explains the high level of expression.

Table 7. Ubiquitin and proteosome-related genes of *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
ATTAGAGGAC	XM_629433.1, XM_629434.1, XM_635973.1, XM_635972.1, XM_635971.1	440	ubiquitin (*ubqN, ubqO, ubqM, ubqL, ubqK*)
TAAGAGTTTT	XM_631880.1	35	26S proteosome regulatory subunit S1 (*psmD1*)
CAAATGTAGT	XM_633133.1	30	RING Zn finger-containing protein (DDB-0232321)
CTAATCGTCA	XM_637723.1	27	ubiquitin (*ubqC*)
CATTCGTAAA	XM_640374.1	23	proteosomal alpha-subunit 7-1 (*pntB*)
TATCTCAAGA	XM_629630.1	15	26S proteosome subunit ATPase 5 (*psmC*5)/HIV1 TAT-binding protein homologue (*tbp*1O)
TGGATAGATA	XM_633869.1	14	proteosome subunit beta type 7 (*psmB*7)
ATGGATGAAT	XM_640496.1	13	13/13 RING Zn finger-containing protein (*dg*1106)
CATATAATAT	AB007024.1	13	DAPS-1 mRNA for proteosome subunit

Tubulins and dyneines

The 11 genes in this category produce the components of microtubules and molecular motors (Table 8). They participate in a variety of processes in amoebae, such as formation of the mitotic spindle, and provide structural support. All of these genes were expressed at moderate to low levels in the transcriptome.

Table 8. Tubulin and dyneine genes of *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
TTTGTCAAGA	XM_631966.1	31	alpha tubulin (*tuba*)
AAGAACAAGC	XM_641070.1	29	beta tubulin gene (*tubB*)
TAAACGAATC	XM_638093.1	25	cytoplasmic dynein heavy chain (*dhcA*)
TCATTCGTTT	XM_638093.1	23	cytoplasmic dynein heavy chain (*dhcA*)
GTCTTCAAGA	XM_639159.1	20	colossin B mRNA (*colB*)
TCAAATGCTG	XM_641218.1	17	major vault protein alpha (*mvp*-alpha)
TGGACTAAAT	XM_638093.1	17	hypothetical protein cytoplasmic dynein heavy chain (*dhcA*)
CAACCTCTGG	XM_629574.1	16	hypothetical protein colossin A (*colA*) mRNA, complete cds
ATGCTGTCAT	XM_001134481.1	14	hypothetical protein centrosomal protein CP224 (*mtaA*) mRNA, complete cds
TGTGGTCCAC	XM_638093.1	14	hypothetical protein cytoplasmic dynein heavy chain (*dhcA*)
AATTACCTAC	XM_631966.1	13	alpha tubulin (*tubA*)

DNA binding

Three different histone genes were expressed at moderate to low levels in the amoebae (Table 9). The most striking gene in this category is a member of the CHD gene family, it had the fifth-highest level of expression in the transcriptome. This recently identified gene family activates or represses RNA transcription, depending on the component partners that are available (Hall and Georgel, 2007).

Table 9. DNA binding molecules of *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
ATTTTTATTA	XM_633616.1	258	CHD gene family protein containing chromodomain, helicase domain, and DNA-binding domain
CCAAAAGATA	XM_642485.1	59	histone H3 (*ddH*3)
CCATCTTACA	XM_632592.1	34	histone H2B domain containing protein (*H2Bv3*)
GCTTTGGATG	XM_632592.1	22	histone H2B domain containing protein (*H2B*v3)
ATGATGATGA	XM_633148.1	16	nucleotide binding protein 1-like protein

Cell adhesion

The major cell adhesion molecule of amoebae, the *cad*A gene, was highly expressed, while the well-characterized contact site B gene was expressed at relatively low levels (Table 10). The two discoidin genes and the single extracellular matrix gene were originally predicted to be expressed only during later developmental stages. Their presence in the transcriptome indicates that they have a function in amoebae.

Table 10. Cell adhesion molecules of *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
TTTGGTTGGT	XM_632956.1	175	calcium-dependent cell adhesion molecule-1 (*cadA*)
TGAAATAAAA	XM_637148.1	79	extracellular matrix protein ST430 (13/13)
GGGTAATAAT	XM_629547.1	39	discoidin II (*dscE*)
GAGTAAAGAT	XM_639850.1	24	CsbA (*csbA*) and CsbB (*csbB*)
ATCAATGGGT	XM_639599.1	16	discoidin-I D chain (*dscD*)

Mitochondrium-related

Several mitochondrial DNA sequences were detected by the SAGE method. Table 11 also includes nuclear encoded, mitochondria-specific proteins other than the electron transport chain components. The porin, superoxide dismutase, ADP/ATP translocase, and 5'AMP-activated gamma subunit were all expressed at relatively low levels.

Table 11. Mitochondrium-related genes of *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
CAAGTTGAAC	AB000109.1	133	mitochondrial DNA
AAAAATATAG	AB000109.1	44	mitochondrial DNA
ATGAGCAGTT	EAL63335	32	mitochondrial DNA
AGCTGGGTTT	EAL63335	28	mitochondrial DNA
GATAAACAAA	XM_640344.1	27	porin (*porA*)
CTGATCAAGA	XM_642037.1	21	superoxide dismutase (*sodA*)
GCTGTCGTCA	EAL63335	21	mitochondrial DNA
GGGTATCGCA	EAL73180	14	ADP/ATP translocase (*ancA*)
AAAATGGTAT	EAL71195	13	5'AMP activated gamma subunit (*prkag*)

Quorum-sensing

Six genes with possible quorum-sensing functions were expressed at moderate to low levels (Table 12). The small aggregate formation, countin and putative countin receptor are components of the countin factor complex, while conditioned medium factor and *rcdGG* are two additional proposed quorum-sensing systems (Mahadeo and Parent, 2006).

Table 12. Quorum-sensing genes of *Dictyostelium discoideum*.

Dictyostelium tag sequence	ID	Count	Description
GTGTAATCCA	EAL63602	77	small aggregate formation (*smlA*)
GTTATTGTAG	XM_638733.1	39	conditioned medium factor (*cmfA*)
GTGCAAATTG	XM_641171.1	14	putative countin receptor Cnr9 (*cnrI*)
TGATGTTGCC	XM_631158.1, U67940	14	P14 domain-containing protein (*rcdGG*) ORFveg106
TGGTTCACCA	XM_633383.1	14	mannose-6-phosphate isomerase (*mpi*)/countin
GTAAACCACC	EAL63602	13	small aggregate formation (*smlA*)

Hypothetical and genes of unknown function

The largest group of genes in the transcriptome (76 tags) involves hypothetical proteins and genes with unknown functions (Table 13). We have defined hypothetical as those genes with recognized open reading frames in the *Dictyostelium* database. This group includes the most highly expressed gene in the transcriptome, *dut*A untranslated RNA, and four other genes expressed at greater than 100 copies. Several genes have structural motifs, such as four genes with EGF-like domains, but remain uncharacterized in any organism. Other genes have been identified in *Dictyostelium*, but their functions are unknown, such as the culmination-specific protein. Conversely, some genes have been identified in other organisms; but their function in *Dictyostelium* is unknown, such as the PhoPQ-activated pathogenicity-related protein.

Table 13. Hypothetical and genes of *Dictyostelium discoideum* of unknown function.

Dictyostelium tag sequence	ID	Count	Description
GGACTTAGAG	EAL64828	773	*dutA* untranslated RNA
TCGCTTTAAT	XM_634269.1	154	hypothetical protein
GCAATCAAAA	XM_630714.1	140	hypothetical protein
GTATAGCTTA	EAL64828	130	*dutA* RNA
CACTAATCAA	XM_635539.1	111	hypothetical protein
GGGTGCCGAT	XM_630282.1	95	hypothetical protein
AAATATAAAA	XM_635590.1	72	hypothetical protein
GAGAATCATC	XM_632378.1	68	hypothetical protein
GGTAATGGTA	XM_634924.1	59	cyclophilin (*ppiA*)
TCCCTATTAA	XM_632008.1	59	hypothetical protein
GGGTGCTGAC	XM_631551.1	57	hypothetical protein
TGTCGATCCA	XM_641310.1	49	hypothetical protein
TGGTGATGGT	XM_635474.1	46	PhoPQ-activated pathogenicity-related protein (*aprA*)
GCAGGAAAAA	XM_636572.1	45	hypothetical czl protein
TACTCTTTCT	AF203735.1	42	culmination specific protein 45D (45D)
GAACCAGCCT	XM_638827.1	41	hypothetical protein
GGATCCAGTG	XM_641936.1	41	hypothetical protein DDB_0190094 partial mRNA
TCCATCATCA	XM_640805.1	37	EGF-like domain
AGTTTCAGCC	XM_640335.2	34	hypothetical protein
GGGTACATAT	XM_636202.1	33	hypothetical protein
GGTATGATGT	XM_638942.1	33	hypothetical protein

Continued on next page

Table 13. Continued.

Dictyostelium tag sequence	ID	Count	Description
AAGTCATCGA	XM_630012.1	32	hypothetical protein (29C)
GCGAATGCAC	XM_642513.1	25	hypothetical protein
GAGATAATCT	XM_631114.1	24	hypothetical protein
GTTTGTTTGG	XM_640382.1	24	EGF-like domain containing protein (*canC*)
TAATAAACTC	AC116986.1	24	chromosome 2 map
GTTTTGGTGG	XM_640757.1	23	hypothetical protein
TAAATAGATC	XM_630528.1	23	hypothetical protein
GTCACTTCTT	XM_639759.1	22	hypothetical protein
GGATCTGCTG	XM_630177.1	21	hypothetical protein
GTCAAGCAAA	XM_632222.2	21	hypothetical protein
TAAAAAAAAA	AC117072.3	21	chromosome 2 map
TAGAAGAAGA	XM_642384.1	21	non-transporter ABC protein (*abcF4*)
TGCTCTGTAG	XM_639875.1	21	major facilitator superfamily protein DDB_0238776
TTGCAGAAAT	XM_630145.1	21	hypothetical protein DDBdraft_0183916
CTAATGGGTC	XM_638179.1	20	hypothetical protein DDB_0232153
GCTAGAAATG	XM_632378.1	20	hypothetical protein DDBdraft_0187197
AAAAACACAT	XM_642213.1	19	hypothetical protein DDBdraft_0216555
AAAAACACAT	XM_641971.1	19	hypothetical protein DDBdraft_0216640
AAAAACACAT	XM_642177.1	19	hypothetical protein DDBdraft_0216547
AGCAGAAGCA	XM_640867.2	19	hypothetical protein DDBdraft_0215165
ATGCTCCAAT	XM_001733034.1	19	hypothetical protein DDBdraft_0215064
GGAATTACCA	XM_632762.1	19	hypothetical protein DDBdraft_0215883
TGGACCAAAG	XM_633814.1	19	hypothetical protein DDB_0231746
CTGCTATTTC	XM_632737.1	18	hypothetical protein DDBdraft_0186902
GTGATCTGTA	XM_637067.1	18	hypothetical protein DDBdraft_0204455
GTGATCTGTA	XM_635945.1	18	hypothetical protein DDBdraft_0204740
GTTGGTACTC	XM_631286.1	18	ankyrin repeat-containing protein DDB_0232969
TACCAAAATG	XM_640812.1	18	hypothetical protein DDBdraft_0201696
TGCTAACCTT	XM_636773.1	18	hypothetical protein DDBdraft_0205035
CACACCAACA	XM_638837.1	17	hypothetical protein DDBdraft_0217482
GGAAACCTTT	XM_639002.1	17	hypothetical protein DDBdraft_0217527
ACTAATCAAT	XM_638814.1	16	hypothetical protein
CGTAGAAAGC	XM_635431.1	16	hypothetical protein
GATGATGTAC	XM_636206.1	16	hypothetical protein
GTGAAAACAA	XM_635685.1	16	hypothetical protein
TGAAGATGGT	XM_633121.1	16	hypothetical protein
TGGAAATGGT	XM_637070.1	16	EGF-like domain-containing protein
ATAGAATGAG	XM_630305.1/	15	penta EF hand calcium binding protein (*pefA*)/apoptosis-linked gene 2 (*alg*2A)
ATGTTTGGTG	XM_639226.1	15	hypothetical protein
GGAAAATACA	XM_641146.2	15	hypothetical protein
GTGGTAGTTA	AC114263	15	chromosome 2 map
TATTTGATAC	XM_633831.1	15	mMob1-like protein
CTGGTCTTGT	XM_631976.1	14	hypothetical protein
GCCGTGATGA	XM_632857.2	14	hypothetical protein
GGGTATCACT	XM_633934.1	14	hypothetical protein
TCATTCACTA	XM_633173.1	14	hypothetical protein
TTGCCAATGG	XM_641284.1	14	CD9 (*cdp*9)
AACTAAATTA	AC116956.2	13	chromosome 2 map
ATCATAACCA	XM_641793.1	13	hypothetical protein
CAAAAATATA	XM_633368.1	13	hypothetical protein
GTGTTGGTTC	XM_633060.1	13	hypothetical protein
GTTAGGTGAT	XM_630798.3	13	FNIP repeat-containing protein DDB_0238003
GTTGTGTGTA	XM_637503.1	13	hypothetical protein DDBdraft_0203368
GTTTCGTTTC	XM_630703.1	13	hypothetical protein DDBdraft_0188832
TGAAAGATTA	XM_632907.1	13	LIM domain-containing protein DDB_0233145

DISCUSSION

The amoeba-stage transcriptome provides the information necessary to analyze the relative expression levels of all expressed genes, and make predictions about the functional capabilities of *Dictyostelium* amoebae. Our results, in some cases, confirm the results of

previous studies that utilized molecular methods that depend on previously identified gene sequences, such as microarray analysis (Van Driessche et al., 2002; Urushihara et al., 2004). However, the value of an SAGE-produced transcriptome is that the entire array of expressed genes can be rank-ordered relative to each other, and identified based upon homology with the genome, rather than inclusion of specified sequences within a mixed probe. The result is a compilation of all gene products that are present within an amoeba. With this information, components of a process or pathway can be linked to this developmental stage, the functional capabilities identified, and their relative importance inferred.

Protein synthesis is a major function of amoebae, as previously reported (Van Driessche et al., 2002). Numerous ribosomal components are present and their expression levels vary from very high (300 copies) to relatively low (13 copies) levels. Conversely, the detectable proteosome components are few and not highly expressed (35 copies or less). Together these data suggest that the expressed genes are relatively stable, because the expression levels of the protein degradation machinery are about 10 times less than those of the protein synthesis machinery.

The moderate to high expression levels of the endosomal system components, transporters and lysosomal enzymes are indicative of the phagocytic/pinocytic nature of amoeba nutrition. The many metabolic enzymes expressed indicate that amoebae are utilizing nutrients for energy production and synthesis of necessary structural components. However, the relatively low expression levels of metabolic enzymes, particularly those of glucose oxidation and oxidative phosphorylation, suggest that this level is simply representative of basal metabolic rates for amoebae. Given the large amount of ATP needed for early development processes (aggregation), it appears that amoebae are not producing and storing ATP for later use.

The only identifiable transmembrane receptor expressed in the amoebae was the countin receptor, which, along with its cognate ligand countin was expressed at low levels (14 copies). These two gene products are components of a quorum-sensing system that also includes the small A gene product (Mahadeo and Parent, 2006). The conditioned medium factor (CMF) represents a second quorum-sensing system and was expressed at moderate to low levels, although its receptor CMFR1 is not present in the transcriptome (Clarke and Gomer, 1995). Other potential cell-signaling components include regulators of monomeric G-protein function, such as Rac1A or Rab6. However, in the absence of ras or receptor tyrosine kinases, these gene products are more likely involved in linking the cytoskeleton to the membrane or in vesicle trafficking. Therefore, quorum sensing appears to be the only form of transmembrane signaling that occurs in amoebae. These data support the paradigm that quorum-sensing systems activate the cAMP synthesis and detection systems required for aggregation, the earliest stage of multicellular development (Van Haastert et al., 1996). Furthermore, the absence of cAMP receptors or heterotrimeric G-protein α-subunits indicates that amoebae at a density of 2.4×10^7 cells/mL (the concentration used for this study) have not initiated development. While an amoeba culture of this density is likely approaching the stationary phase, the gene expression patterns in these amoebae demonstrate that they have not made the transition to multicellular development (Franck et al., 2008).

One of the surprising findings of the *Dictyostelium* genome sequence and analysis is the large number (43) of polyketide synthase genes present in the genome (Eichinger et al., 2005). The amoeba transcriptome includes 13 different polyketide synthases with a wide range of expression levels from high (104 copies) to low (13 copies). To date, the only clear function for polyketide synthase is in the biosynthetic pathway for differentiation-inducing factor-1 (Dif-1) (Kay, 1998; Ghosh et al., 2008). However, Dif-1 functions much later in

development to induce the differentiation of stalk cells. Therefore, the products of the 13 amoeba-stage polyketide synthases are unknown.

The function of the most highly expressed gene in amoebae, *dut*A untranslated RNA (773 copies), is unknown (Yoshida et al., 1994; Hinas and Soderborn, 2007). Obviously, a gene expressed at a level almost twice that of the next highest gene has an important function, and identifying that function should produce important insights into what defines an amoeba. Non-coding RNAs, like *dut*A, function in a variety of ways in other organisms, such as chromatin remodeling or transcriptional or translational inhibition (Hinas and Soderbom, 2007). A recent report suggests that *dut*A untranslated RNA can function as a multicopy suppressor of STATa, a gene required for the last stage of development, culmination (Shimada and Kawata, 2007). A similar situation exists for the fifth-most-highly expressed gene in the amoeba transcriptome, chromodomain/helicase/DNA-binding domain or CHD protein (258 copies) (Hall and Georgel, 2007). CHD proteins in other organisms can function to increase or decrease transcription, depending on the available partner proteins. Possibly, amoebae remain undifferentiated because development genes are repressed at the transcription level by CHD and at the translation level by *dut*A. Understanding the mechanisms by which both of these gene products function should produce fruitful insights into developmental control of gene expression.

The large number of hypothetical gene products or genes of unknown function (76 genes), particularly those expressed at high levels (>100 copies), indicates that there is much we do not know about the genes necessary to be an amoeba. The amoeba transcriptome provides a basis for selective characterization of genes that can fill the knowledge gaps, and enhance our understanding of the transition point from single cells to multicellular development.

ACKNOWLEDGMENTS

Research supported by NIH Grant #P20 RR16454 from the BRIN/INBRE Program of the National Center for Research Resources, and an NSF EPSCoR Ph.D. fellowship to T.J. Whitney.

REFERENCES

Chisholm RL, Gaudet P, Just EM, Pilcher KE, et al. (2006). dictyBase, the model organism database for *Dictyostelium discoideum*. *Nucleic Acids Res.* 34: D423-D427.

Chung CY and Firtel RA (2000). *Dictyostelium*: A Model Experimental System for Elucidating the Pathways and Mechanisms Controlling Chemotaxis. In: Principles of Molecular Regulation (Conn PM and Means AR, eds.). Springer, New York, 99-114.

Clarke M and Gomer RH (1995). PSF and CMF, autocrine factors that regulate gene expression during growth and early development of *Dictyostelium*. *Experientia* 51: 1124-1134.

Eichinger L, Pachebat JA, Glockner G, Rajandream MA, et al. (2005). The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 435: 43-57.

Firtel RA (1996). Interacting signaling pathways controlling multicellular development in *Dictyostelium*. *Curr. Opin. Genet. Dev.* 6: 545-554.

Franck C, Ip W, Bae A, Franck N, et al. (2008). Contact-mediated cell-assisted cell proliferation in a model eukaryotic single-cell organism: an explanation for the lag phase in shaken cell culture. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 77: 041905.

Ghosh R, Chhabra A, Phatale PA, Samrat SK, et al. (2008). Dissecting the functional role of polyketide synthases in *Dictyostelium discoideum*: biosynthesis of the differentiation regulating factor 4-methyl-5-pentylbenzene-1,3-diol. *J. Biol. Chem.* 283: 11348-11354.

Hall JA and Georgel PT (2007). CHD proteins: a diverse family with strong ties. *Biochem. Cell Biol.* 85: 463-476.

Hinas A and Soderbom F (2007). Treasure hunt in an amoeba: non-coding RNAs in *Dictyostelium discoideum*. *Curr. Genet.* 51: 141-159.

Kay RR (1998). The biosynthesis of differentiation-inducing factor, a chlorinated signal molecule regulating *Dictyostelium* development. *J. Biol. Chem.* 273: 2669-2675.

Kenzelmann M and Muhlemann K (1999). Substantially enhanced cloning efficiency of SAGE (Serial Analysis of Gene Expression) by adding a heating step to the original protocol. *Nucleic Acids Res.* 27: 917-918.

Lemmon MA (2004). Pleckstrin homology domains: not just for phosphoinositides. *Biochem. Soc. Trans.* 32: 707-711.

Mahadeo DC and Parent CA (2006). Signal relay during the life cycle of *Dictyostelium*. *Curr. Top. Dev. Biol.* 73: 115-140.

Parent CA and Devreotes PN (1996). Molecular genetics of signal transduction in *Dictyostelium*. *Annu. Rev. Biochem.* 65: 411-440.

Postma M, Bosgraaf L, Loovers HM and Van Haastert PJ (2004). Chemotaxis: signalling modules join hands at front and tail. *EMBO Rep.* 5: 35-40.

Shimada N and Kawata T (2007). Evidence that noncoding RNA *dut*A is a multicopy suppressor of *Dictyostelium discoideum* STAT protein Dd-STATa. *Eukaryot. Cell* 6: 1030-1040.

Urushihara H, Morio T, Saito T, Kohara Y, et al. (2004). Analyses of cDNAs from growth and slug stages of *Dictyostelium discoideum*. *Nucleic Acids Res.* 32: 1647-1653.

Van Driessche N, Shaw C, Katoh M, Morio T, et al. (2002). A transcriptional profile of multicellular development in *Dictyostelium discoideum*. *Development* 129: 1543-1552.

Van Haastert PJ, Bishop JD and Gomer RH (1996). The cell density factor CMF regulates the chemoattractant receptor cAR1 in *Dictyostelium*. *J. Cell Biol.* 134: 1543-1549.

Velculescu VE, Zhang L, Vogelstein B and Kinzler KW (1995). Serial analysis of gene expression. *Science* 270: 484-487.

Venkateswarlu K, Brandom KG and Yun H (2007). PI-3-kinase-dependent membrane recruitment of centaurin-α2 is essential for its effect on ARF6-mediated actin cytoskeleton reorganisation. *J. Cell Sci.* 120: 792-801.

Verkerke-van Wijk I, Fukuzawa M, Devreotes PN and Schaap P (2001). Adenylyl cyclase A expression is tip-specific in *Dictyostelium* slugs and directs StatA nuclear translocation and CudA gene expression. *Dev. Biol.* 234: 151-160.

Weimer C, Beck R, Eckert P, Reckmann I, et al. (2008). Differential roles of ArfGAP1, ArfGAP2, and ArfGAP3 in COPI trafficking. *J. Cell Biol.* 183: 725-735.

Yoshida H, Kumimoto H and Okamoto K (1994). *dut*A RNA functions as an untranslatable RNA in the development of *Dictyostelium discoideum*. *Nucleic Acids Res.* 22: 41-46.