Published online March 9, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 5 ***1647–1653***
DOI: 10.1093/nar/gkh262

Analyses of cDNAs from growth and slug stages of *Dictyostelium discoideum*

Hideko Urushihara*, Takahiro Morio, Tamao Saito[1], Yuji Kohara[2], Eiko Koriki, Hiroshi Ochiai[1], Mineko Maeda[3], Jeffrey G. Williams[4], Ikuo Takeuchi[5] and Yoshimasa Tanaka

Institute of Biological Sciences, University of Tsukuba, Tsukuba-shi 305-8572, Japan, [1]Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan, [2]National Institute of Genetics, Mishima 411-8540, Japan, [3]Department of Biology, Graduate School of Science, Osaka University, Osaka 560-0043, Japan, [4]School of Life Sciences, University of Dundee, Welcome Trust Biocentre, Dow Street, Dundee DD1 5EH, UK and [5]Novartis Foundation for the Promotion of Science, Minato-ku, Tokyo 106-0032, Japan

Received October 26, 2003; Revised December 25, 2003; Accepted January 13, 2004 DDBJ/EMBL/GenBank accession nos+

ABSTRACT

Dictyostelium **is a favored model for studying problems in cell and developmental biology. To comprehend the genetic potential and networks that direct growth and multicellular development, we are performing a large-scale analysis of** ***Dictyostelium*** **cDNAs. Here, we newly determine 7720 nucleotide sequences of cDNAs from the multicellular, slug stage (S) and 10 439 from the unicellular, vegetative stage (V). The combined 26 954 redundant ESTs were computer assembled using the PHRAP program to yield 5381 independent sequences. These 5381 predicted genes represent about half of the estimated coding potential of the organism. One-third of them were classified into 12 functional categories. Although the overall classification patterns of the V and S libraries were very similar, stage-specific genes exist in every category. The majority of V-specific genes function in some aspect of protein translation, while such genes are in a minority in the S-specific and common populations. Instead, genes for signal transduction and multicellular organization are enriched in the population of S-specific genes. Genes encoding the enzymes of basic metabolism are mainly found in the common gene population. These results therefore suggest major differences between growing and developing** ***Dictyostelium*** **cells in the nature of the genes transcribed.**

INTRODUCTION

The cellular slime mold *Dictyostelium discoideum* is a simple eukaryote that feeds on bacteria and proliferates by fission. Upon starvation, typically ~100 000 amoebae aggregate together, migrate as a slug and form a fruiting body composed of a spore mass and a stalk (1). Since this process can, under appropriate conditions, be completed within 24 h, and since cells differentiate into only two major cell types, spores and stalk cells, it serves as a very good model of multicellular development. The growing cells are also an excellent system with which to study those processes, such as phagocytosis and cell movement, that other lower eukaryotes such as yeast cells do not undertake. The small, haploid 34 Mb genome (2) is advantageous for molecular genetic approaches and it allows many of the genetic approaches used in yeast. In order to facilitate all aspects of *Dictyostelium* research, a transcriptome analysis by the *Dictyostelium* cDNA Project in Japan (3,4) and a genome analysis by the *Dictyostelium* Genome Sequencing Consortium (5–8) are being carried out.

One highly significant, simplifying factor in transcriptome analysis is that, by definition, the sequences represent expressed genes. Gene prediction therefore becomes a much less arbitrary process and, if the genome sequence is available for a sequence, it is possible to use EST data to define a complete gene structure, including exon–intron boundaries. An EST set will also reveal undiscovered genes and, by microarray analysis, gene expression profiles can be obtained from the entire EST set. The microarray set will, however, normally be selective because ESTs prepared from a non-normalized library will predominantly be derived from moderately to highly expressed genes. This means that genes expressed at a very low level will be under-represented and they are best obtained by gene prediction from the genomic DNA sequence. It is not, however, at all clear that

*To whom correspondence should be addressed. Tel: +81 29 853 4664; Fax: +81 29 853 6614; Email: hideko@biol.tsukuba.ac.jp

+AU034148–AU035032, AU051786–AU052005, AU052113–AU053815, AU053822–AU054189, AU060072–AU060433, AU060444–AU062282, AU071274–AU075122, AU075916–AU075965, AU076312–AU076395, AU261276–AU271296

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

sequences of very low abundance can be detected using current microarray technology. For the effective use of EST data, the correct clustering and assembly of ESTs and their subsequent annotation are of critical importance. Transcriptome analyses in various organisms have therefore paid substantial attention to these issues (9,10).

In the present study, we completed the structural analysis of cDNAs in the slug stage libraries generated in a previous work (3) and extended the analysis to the growth phase cDNAs.

MATERIALS AND METHODS

Organism and cDNA libraries

The axenic strain of *D.discoideum*, AX4, was used. This is the same strain as used in the analysis of genomic DNA (8). The oligo(dT)-primed cDNA libraries from the slug stage cells, the SL and SS libraries, have been described (3). The SL library was constructed by ligating cDNA inserts to the vector pSPORT1 after size fractionating the SS cDNAs. A cDNA library of the unicellular stage (VS library) was constructed in the present study as follows. Poly(A)$^+$ RNAs were prepared from AX4 cells in logarithmic growth phase (3×10^6 cells/ml) in HL5 medium. The RNAs were then reverse transcribed using an oligo(dT)$_{15}$ primer carrying a NotI adapter. After the second strand synthesis, SalI adapters were ligated to both ends, followed by NotI digestion to yield a NotI end at the 3′-terminus. Primers and small DNA fragments were removed with a Chroma Spin-400 column (Clontech, USA) and the recovered cDNAs were ligated into pBluescript II KS$^-$ for the transformation of *Escherichia coli* DH5α cells.

Determination of nucleotide sequences

Plasmid DNA was prepared using a plasmid miniprep kit (QIAprep 8; Qiagen, Germany). Sequencing reactions were performed using the DYEnamic™ Direct Cycle Sequencing Kit (Amersham, USA) or PRISM Ready Reaction Dye Terminator Cycle Sequencing Kit (Applied Biosystems, USA). The nucleotide sequences were determined by a single pass from both ends of the inserts using an automated DNA sequencer, Model 377 or 3700 (Perkin-Elmer, USA) and the data were edited manually or using the PHRED program (11,12).

Sequence assembly and cDNA clustering

The read pair of each clone was automatically merged into a single sequence if their ends matched over 50 nt or longer at identities not less than 90% in the lfasta (13) alignment. Slightly shorter overlaps were accepted only if the length of merged sequences fitted the insert size of the clones. This process will be called 'Automerge'. Sequence assembly was performed using the PHRAP program (14). The short sequences (<100 nt) and *dutA*-containing sequences were eliminated prior to assembly. The cDNA copies of the *dutA* RNA, a non-coding polyadenylated RNA sequence (15), quite frequently appeared as portions of the inserts of otherwise unrelated clones mainly in the SS library and caused chimera-like contigs. Short repeat sequences such as (TAA)$_n$ and (TCC)$_n$ were masked using the Repeat-masker program included in the PHRED–PHRAP package. Then, the whole sequences in the separate or combined libraries were assembled. The PHRAP conditions we used were: stringency 0.85, vector bound 0, and pre-assembly.

The PHRAP output was further processed as follows. First of all, singlets and singletons normally eliminated for genomic assembly were integrated as contigs, because low prevalence cDNAs are inevitably unique but also valuable. When two sequences derived from the same clones (read pairs) were found in independent contigs, the relevant contigs were artificially joined with an affixed gap of 10 'N's and are termed 'gapped contigs'. This avoids an overestimation of gene numbers. If more than two contigs were to be joined but their actual positional relationships were unclear, the third and any additional contigs were left separate and are termed 'orphan contigs'. The singlets, which represent the sequences of unique cDNA clones, and the remainder of the PHRAP contigs are termed 'ordinary contigs'. The sum of the ordinary contigs and gapped contigs corresponds to the number of independent genes.

Functional classification

The cDNAs and their contigs were assigned to either multiple functional categories or to single major ones in the present study. The functional categories are based on those of *Saccharomyces cerevisiae* obtained from MIPS (http://mips.gsf.de/proj/yeast/CYGD/db/index.html) and supplemented with those characteristic to multicellular organisms (Supplementary Material 1). To assign genes to appropriate functional categories, we first made a list of keywords extracted from descriptions in BLASTX results and assigned them to appropriate functional categories referring to MIPS. The smallest number of categories was automatically chosen for each keyword to be employed for single category assignment, which was manually modified where necessary. We then searched the highest hit (E value < e – 4) BLASTX description of each sequence with keywords and assigned it to the corresponding category. When ambiguous descriptions, such as 'probable membrane protein' or 'unknown sequence of', were found in the highest hit from BLASTX, the next highest hit was searched up to the tenth. Our classification correlates mainly to the cellular process of Gene Ontology (GO) (http://www.geneontology.org/).

RESULTS

Collection and assembly of ESTs

The previously unanalyzed clones in the slug stage libraries (SS and SL) and all clones in the vegetative stage (VS) library were sequenced from both ends to generate a total of 26 954 ESTs (Table 1). The PHRAP assembly of these ESTs generated 7680 sequence clusters composed of contigs, singletons and singlets. We processed these sequence clusters as described in Materials and Methods and finally obtained 3450 ordinary contigs and 1931 gapped contigs, together corresponding to the sequences of 5381 independent genes and 368 orphan contigs.

The appearance of new genes during the sequencing of each library is shown in Figure 1A. The efficiency of finding new genes decreased most rapidly in the SL library. The high redundancy in this library can be attributed to the size fractionation step included in the library construction. New

Table 1. Summary of EST collection and assembly

	Library SL	SS	All S	VS	Total
Clones	5247	9176	14 423	6716	21 139
Reads[a]	6420	10 095	16 515	10 439	26 954
Previous work	1132	7663	8795	0	8795
Present work	5288	2432	7720	10 439	18 159
PHRAP results	2660	4212	5811	2674	7680
Contigs	1728	3182	4406	2381	5749
Ordinary contig	677	2032	2456	2075	3450
Gapped contig	932	1030	1531	293	1931
Orphan contig	119	120	419	13	368
Independent gene	1609	3062	3987	2368	5381
Average redundancy	3.3	3.0	3.6	2.8	3.9

[a]DDBJ accession numbers: AU033318–AU035032, AU036913–AU037370, AU037392–AU040094, AU051786–AU052005, AU052113–AU053815, AU053822–AU054189, AU060072–AU060433, AU060444–AU062282, AU071274–AU075122, AU075916–AU075965, AU076312–AU076395, AU261276–AU271296, C24606–C24688, C25499–C25687, C83826–C84252, C84677–C84905, C85032–C85043, C89614–C91581, C91900–C93423, C93683–C94532.

genes were still emerging at a rate of ~1 gene/5 clones in all the libraries at the end of sequencing, suggesting the existence of many low copy number mRNAs in both the growing and in the developing cells. The sequencing proceeded from SS to SL and then to VS. The rate of gene discovery increases markedly upon the inclusion of the VS library. This is a clear indicator of stage-dependent gene expression. The clone redundancy patterns for low redundancy (<5 representatives) genes are similar among the three libraries, the fraction of unique clones being ~60% (Fig. 1B). For the entire sequences, unique clones decreased to ~35%.

Gene identification and functional classification

Among the 5381 independent genes described above, 153 (2.8%) are identical to individually characterized genes of *D.discoideum*. By BLASTX searches, ~40% showed homology to database entries high enough to predict their functions (E value < e – 9), 30% were weakly homologous and the next 30% had no hits at all. Detailed annotations can be found in the catalog of S+V genes (Supplementary Material 2). When BLASTN searches were performed against the published chromosome 2 sequences (8), 1063 (19.8%) showed a very high homology (E value < e – 99), indicating their location on this chromosome. Most of those genes coincided with the predicted ones on it (http://genome.imb-jena.de/cgi-bin/dicty/chr2html.pl), although detailed matching was not complete due to fragmentary cDNA sequences and/or absence of untranslated regions in the predicted gene sequences.

Comparison of cDNAs from vegetative and slug stages

The slug stage (S) cDNA clones were assembled to yield 3987 independent genes and the vegetative stage (V) clones yielded 2368 genes (Table 1). Among the 5381 non-redundant gene sequences obtained from all cDNA reads, only 1060 genes contain both slug phase and growth phase cDNAs. Although the rest of the sequences were wholly V-specific or S-specific, only those gene sequences generated by the assembly of five or more cDNA clones were utilized for specificity analysis. The numbers of such redundant clones in the SL, SS and VS libraries are 3816, 6019 and 4811, respectively. These SL and

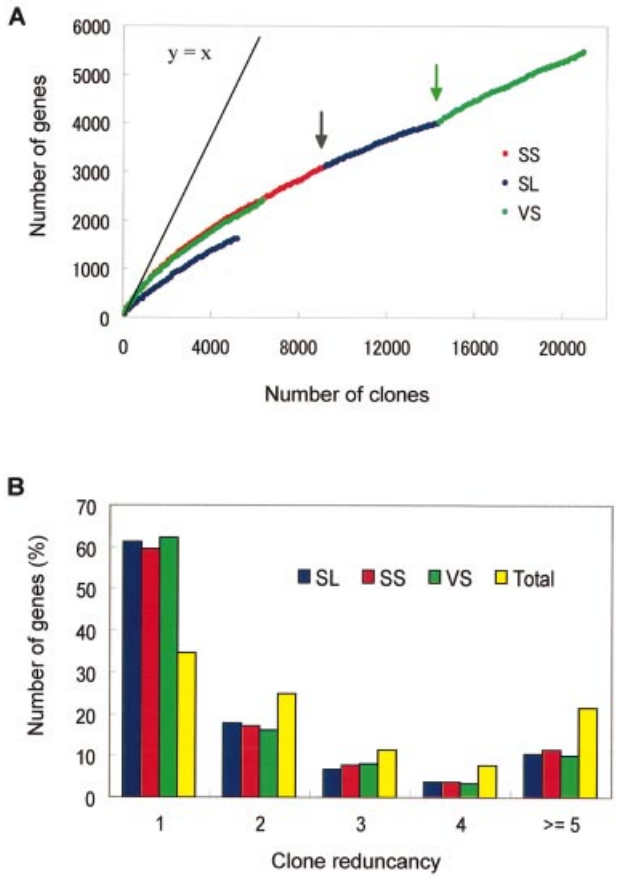

Figure 1. Clone redundancy features. (**A**) Number of independent genes plotted against number of clones sequenced for SS (red), SL (blue) and VS (green) libraries. The curve for the SS library is extended by adding SL and VS clones (indicated by blue and green arrows, respectively). The line Y = X shows the case where every clone is independent. (**B**) Number of genes for clone redundancy indicated on the abscissa is shown for the assembly of each and total libraries.

SS clones were combined and termed S clones, and the clone numbers were normalized for comparison between the S and VS (V) libraries. Genes with 90% or more V clones were regarded as V-specific genes and those with 90% or more S clones were S-specific genes. Only those with 30–70% V (or S) clones were regarded as common genes. The numbers of V-specific, common and S-specific genes were 182, 299 and 398, respectively.

When the genes in each specificity category were functionally classified, clear differences were observed among them (Fig. 2A). Genes in category 4 (translation) occupy the largest fraction of V-specific genes, while the common and S-specific genes were extremely under-represented in this category. S-specific genes are abundant in categories 11 (signal transduction) and 12 (multicellular organization). Common genes were, as expected, mostly classified into category 1 (basic metabolism). If the clones, instead of genes, were classified, the stage-dependent characteristics in classification patterns were even clearer (Fig. 2B).

Gene switching for multicellular development

The high level (top 20) genes specific to S stage cells are shown in Table 2. In addition to the genes in categories 11 and

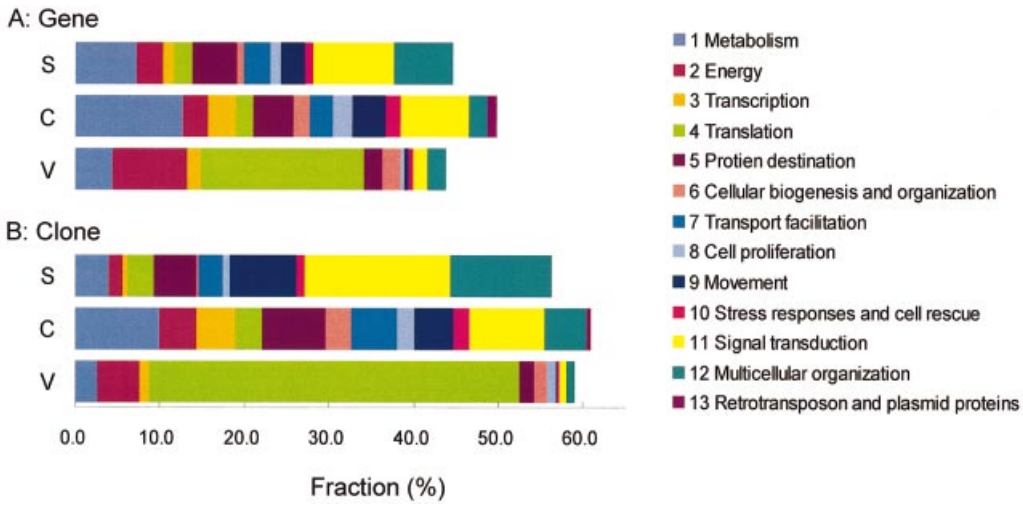

Figure 2. Comparison of growth phase and slug stage cDNAs. Genes containing five clones or more are functionally classified. S, V and C indicate slug stage-specific (90% or more S clones), growth phase-specific (90% or more V clones) and common (70–30% of either stage clones), respectively. Categories 14 (classification uncertain) and 15 (no or faint hits) are not shown.

Table 2. High level genes turned on for development

Contig[a]	Clone no.[b]	Identity/homology description (E value)	Category
3460	454	*Dictyostelium* sp.Gα4 (5.10E – 66)	11.3
3505	387	*Dictyostelium discoideum* gene for actin A8	9.1
4978	227	No hits	15.1
4072	149	*Dictyostelium discoideum* Dp87 protein	12.2
368	147	*Mus musculus* J1 protein (1.00E – 113)	4.1
5143	137	*Dictyostelium discoideum* sp96 gene for spore coat protein SP96	12.2
4312	127	*Dictyostelium discoideum* calcium-binding protein	11.7
4663	94	*Dictyostelium discoideum* calcium-binding protein 4a	11.7
4512	92	*Drosophila melanogaster* opt1 long protein (7.10E – 32)	7.9
3837	86	*Dictyostelium discoideum* spore coat protein SP60	12.2
339	81	*Dictyostelium discoideum* cysteine proteinase 2	5.5
3872	67	*Dictyostelium discoideum* SP85 protein	12.2
4235	66	No hits	15.1
4454	61	*Dictyostelium discoideum* ORFveg106 mRNA	14.1
4128	60	*Arabidopsis thaliana* T29M8.9 protein (4.8)	15.2
3908	58	*Dictyostelium discoideum* spore coat protein SP87	12.2
665	57	No hits	15.1
4337	55	No hits	15.1
4381	52	No hits	15.1
1769	46	No hits	15.1

[a]Bold numbers indicate that the contig corresponds to a known *Dictyostelium* gene.
[b]Underlining indicates that the contig contains S clones only and is exclusively S-specific.

12, which are as expected, others were found in other categories, such as 4 (translation), 5 (protein destination), 7 (transport facilitation) and 9 (movement). Since these categories are related to basic cellular processes, the existence of slug-specific genes in them indicates that different sets of genes, not just developmental genes relevant to terminal differentiation, are transcribed in developing cells, even for the same cellular processes as those in growing cells. It is noteworthy that nearly one-third of these high level genes have no or faint hits to the database entries.

The top 20 genes specific to V stage cells (turned off for development) are all assigned to category 4.1 (ribosomal proteins), except for gene V4b, of uncertain function, and one gene without any hits to the database (data not shown). Thus, switching gene expression on and off in development is very extensive. It should be pointed out, however, that genes for transcriptional control (category 3.5) are low in frequency, suggesting that transcriptome switching is not operated by a few key genes but by a complicated network involving many interacting genes. Probable transcription factors with five or more clones are shown in Table 3.

DISCUSSION

In the present study, we performed a large-scale analysis of cDNAs from *D.discoideum* cells at the unicellular (vegetative) and multicellular (slug) stages. PHRAP-based assembly of the EST reads generated sequences for 5381 independent genes. This clustering may still have inaccuracies due to the presence of many singletons, but it is obviously better than in-house

Table 3. Possible transcription factor genes appeared frequently[a]

Contig[b]	Clones	Identity/homology description (E value)	Specificity[c]
4508	11	*Dictyostelium discoideum* homeobox-containing protein Wariai (1.00E – 104)	S
3614	5	*Dictyostelium discoideum* G-box-binding factor (6.00E – 13)	S
3556	5	*Dictyostelium discoideum* G-box-binding factor	S
4029	11	*Dictyostelium discoideum* transcriptional repressor TUP1	(S)
3785	8	*Xenopus laevis* xMSS1 protein (1.00E – 100)	C
555	7	*Schizosaccharomyces pombe* putative VBP1/PAC10 family protein (4.00E – 18)	C
5237	5	*Dictyostelium discoideum* LimC protein	C
3345	5	*Danio rerio* cytoskeleton-associated LIM domain (2.00E – 12)	(V)
3176	6	L*eishmania major* HEXBP DNA binding protein (3.10E – 41)	(V)
2403	14	*Drosophila melanogaster* vig protein (8.00E – 08)	V
3195	13	*Dictyostelium discoideum* LIM domain protein	V

[a]Only those containing five clones or more are listed.
[b]Bold numbers indicate that the contig corresponds to a known *Dictyostelium* gene.
[c](S) indicates that the specificity is between S and C, and (V) between V and C.

BLAST searches; where it is difficult to exclude false groupings while securely maintaining real groups. The gene number in *D.discoideum* has been estimated to be ~8000 by Rot analysis (16), 10 000 from the average gene density in the total genome (17) and 11 000 from chromosome 2 sequencing (8). Thus, genes described here should cover about half of the coding potential in this organism. The sequence and other information from our cDNA clones is available from the Project web site (see the section Web Navigation) and constitutes a valuable resource for *Dictyostelium* research (7). In addition to the analysis of independent genes, our cDNA clone set has enhanced the published microarray studies for temporal and spatial expression analyses (18–20). Evidently, the assembled sequences, relevant cDNA clones and their annotations described here are prerequisites for those microarray works.

The prevalence of particular cDNAs in a library can give some approximation of gene expression levels, even if it is not a faithful reflection of an mRNA population without a filtering or fractionation process, as, for example, shorter cDNAs tend to be cloned more efficiently than longer ones and therefore tend to be more prevalent. In fact, the already published genes found in Table 2, actin A8 (21), cysteine proteinase 2 (22) and SP87 (23), are all known as development-specific genes. Several other developmental genes, such as *lagC* (24) and *carA* (25), are also found exclusively in the slug libraries. The most remarkable transcriptome changes associated with the onset of development are an extensive suppression of genes related to translation, which is also reported by microarray analyses to occur a few hours after the initiation of development (20; our preliminary data), and induction of those relevant to signal transduction and to multicellular development. Thus, strong regulation at the transcriptional level operates for the cellular transition from growth to development, in agreement with the results of microarray analysis of gene expression patterns (18). For greater accuracy, quantitative methods such as real time RT–PCR should be employed, but its application to entire gene sets is currently difficult to perform. The microarray technique is more direct than library analysis, but it still has problems of specificity and sensitivity of hybridization besides preparation of high quality arrays.

We have identified at least some 50% of the genes in *Dictyostelium*. To obtain a complete gene list, the use of cDNA libraries from cells of different physiological status and/or normalization will be necessary. Thus, recently, we efficiently isolated new genes using a subtraction library of sexual phase cells (26). For complete structural information, we are continuing our sequencing efforts using full-length cDNA libraries constructed by the oligo-capping method (27), which estimates inclusion of transcription initiation sites to be 50–80% (28,29). Increasing the number of sequences will improve the accuracy of assembly and also help to fill the sequence gaps. Finally, the necessity for integration of genome sequencing results should be pointed out. Many of the cDNA sequences had hits to *Dictyostelium* genomic sequences publicly available (http://genome.imb-jena.de/dictyostelium/). These sequence data are of enormous help in verifying sequence assembly and finding genes for low copy number mRNAs. Moreover, alignment of cDNA and genome sequences provides important information concerning untranscribed regions that control gene activities. These matters will be the most intriguing issues upon finishing the cDNA analyses.

WEB NAVIGATION

The Dicty_cDB top page (URL1) is an entry point to all our cDNA data. The sequence and related information for each cDNA clone are collected in the Clone-index page (URL2), which can be accessed via Clone-overview (URL3). Alternatively, the index page can be found using Search-master (URL4), which also provides information on relevant contigs (of the newest version) and DDBJ accession nos for query clones, as well as keyword (URL5), pattern (URL6) and BLAST homology searches (URL7). The library-based catalogs and contig catalog are obtained from Catalog-index (URL8). The Contig-index pages (URL9) are accessed either via Contig-overview (URL10) or those catalogs. Since the sequencing project is still continuing and the contig set is subject to revision, the contig names in the present work do not necessarily coincide with the Web. The guide for requesting clones or clone sets can be seen in the Request page (URL11). The addresses of the above mentioned web pages are summarized in Table 4.

Table 4. List of URLs for *Dictyostelium* cDNA information

No.	Title	URL
1	Dicty_cDB top	http://www.csm.biol.tsukuba.ac.jp/cDNA/database.html
2[a]	SLA101 index	http://www.csm.biol.tsukuba.ac.jp/CSM/SL/SLA1-A/SLA101Q.Seq.d/
3	Clone Overview	http://www.csm.biol.tsukuba.ac.jp/CSM/
4	Search_master	http://www.csm.biol.tsukuba.ac.jp/~tools/html/search_master.html
5	Keyword-search	http://www.csm.biol.tsukuba.ac.jp/WAIS/
6	Pattern-search	http://www.csm.biol.tsukuba.ac.jp/~tools/html/FindPattern.html
7	BLAST search	http://dicty.gene.tsukuba.ac.jp/~blast/blast/CSM-contig.html
8	Catalog Index	http://www.csm.biol.tsukuba.ac.jp/catalogue/Catalogue.html
9[a]	Contig 00009 index	http://www.csm.biol.tsukuba.ac.jp/CSM_Contig/Contig-U00009–1Q.Seq.d/
10	Contig Overview	http://www.csm.biol.tsukuba.ac.jp/CSM_Contig/
11	Request	http://www.csm.biol.tsukuba.ac.jp/cDNA/req_clones.html

[a]Only one example is shown.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully thank Professor Naotake Ogasawara of the Nara Institute of Science and Technology for valuable comments on and encouragement of our project. This work was supported by research grants from Research for the Future (JSPS-RFTF96L00105) of the Japan Society for the Promotion of Science and by a Grant-in-Aid for Scientific Research on Priority Areas C from Ministry of Education, Culture, Sports, Science and Technology of Japan (0823105).

REFERENCES

1. Kessin,R.H. (2001) *Dictyostelium: Evolutional Cell Biology and the Development of Multicellularity*. Cambridge University Press, New York, NY.
2. Cox,E.C., Vocke,C.D., Walter,S., Gregg,K.Y. and Bain,E.S. (1990) Electrophoretic karyotype for *Dictyostelium discoideum*. *Proc. Natl Acad. Sci. USA*, **87**, 8247–8251.
3. Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I. and Tanaka,Y. (1998) The *Dictyostelium* Developmental cDNA Project: generation and analysis of expressed sequence tags from the first-finger stage of development. *DNA Res.*, **5**, 335–340.
4. Urushihara,H. (2002) Functional genomics in *Dictyostelium discoideum*. *Mol. Cells*, **13**, 1–4.
5. Sucgang,R., Shaulsky,G. and Kuspa,A. (2000) Toward the functional analysis of the *Dictyostelium discoideum* genome. *J. Eukaryot. Microbiol.*, **47**, 334–339.
6. Kuspa,A., Sucgang,R. and Shaulsky,G. (2001) The promise of a protist: the *Dictyostelium* genome project. *Funct. Integr. Genomics*, **1**, 279–293.
7. Kreppel,L. and Kimmel,A.R. (2002) Genomic database resources for *Dictyostelium discoideum*. *Nucleic Acids Res.*, **30**, 84–86.
8. Glockner,G., Eichinger,L., Szafrnski,K., Pachebat.,J.A, Bankier,A.T., Dear,P.H., Lehmann,R., Baumgart,C., Parra,G., Abril,J.F., Gulgo,R., Kumpf,K., Tunggal,B., the *Dictyostelium* Sequencing Consortium, Cox,E., Quall,M.A., Platzer,M., Rosenthal,A. and Noegel,A.A. (2002) Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature*, **418**, 79–85.
9. Smith,T.P.L., Crosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keele,J.W. (2001) Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle. *Genome Res.*, **11**, 626–630.
10. The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*, **489**, 685–690.
11. Ewing,B., Hillier,L., Wendl,M.C. and Green,P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.*, **8**, 175–185.
12. Ewing,B. and Green,P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.*, **8**, 186–194.
13. Pearson,W.R. and Miller,W. (1992) Dynamic programming algorithm for biological sequence comparison. *Methods Enzymol.*, **210**, 575–601.
14. Gordon,D., Abajian,C. and Green,P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.*, **8**, 195–202.
15. Yoshida,H., Kumimoto,H. and Okamoto,K. (1994) *dutA* RNA functions as an untranslatable RNA in the development of *Dictyostelium discoideum*. *Nucleic Acids Res.*, **11**, 41–46.
16. Blumberg,D.D. and Lodish,H.F. (1980) Changes in the messenger RNA population during differentiation of *Dictyostelium discoideum*. *Dev. Biol.*, **78**, 285–300.
17. Loomis,W.F., Welker,D. Hughes,J., Maghakian,D. and Kuspa,A. (1995) Integrated maps of the chromosomes in *Dictyostelium discoideum*. *Genetics*, **141**, 147–157.
18. Van Driessche,N., Shaw,C., Katoh,M., Morio,T., Sucgang,R., Ibarra,M., Kuwayama,H., Saito,T., Urushihara,H., Maeda,M., Takeuchi,I., Ochiai,H., Eaton,W., Tollett,J., Halter,J., Kuspa,A., Tanaka,Y. and Shaulsky,G. (2002) A transcriptional profile of multicellular development in *Dictyostelium discoideum*. *Development*, **129**, 1543–1552.
19. Maeda,M., Sakamoto,H., Iranfar,N., Fuller,D., Maruo,T., Ogihara,S., Morio,T., Urushihara,H., Tanaka,Y. and Loomis,W.F. (2003) Changing patterns of gene expression in prestalk cell subtypes of *Dictyostelium* recognized by *in situ* hybridization with genes from microarray analyses. *Eukaryot. Cell*, **2**, 638–645.
20. Iranfar,N., Fuller,D. and Loomis,W.F. (2003) Genome-wide expression analyses of gene regulation during early development of *Dictyostelium discoideum*. *Eukaryot. Cell*, **2**, 664–670.
21. Romans,P. and Firtel,R.A. (1985) Organization of the actin multigene family of *Dictyostelium discoideum* and analysis of variability in the protein coding regions. *Mol. Biol.*, **186**, 321–335.
22. Pears,C.J. Mahbubani,H.M. and Williams,J.G. (1985) Characterization of two highly diverged but developmentally co-regulated cysteine proteinase genes in *Dictyostelium discoideum*. *Nucleic Acids Res.*, **13**, 8853–8866.
23. Yoder,B.K., Mao,J., Erdos,G.W., West,C.M. and Blumberg,D.D. (1994) Identification of a new spore coat protein gene in the cellular slime mold *Dictyostelium discoideum*. *Dev. Biol.*, **163**, 49–65.
24. Dynes,J., Clark,A.M., Shaulsky,G., Kuspa,A., Loomis,W.F. and Firtel,R.A. (1994) *LagC* is required for cell–cell interactions that are essential for cell-type differentiation in *Dictyostelium*. *Genes Dev.*, **8**, 948–958.
25. Klein,P.S., Sun,T.J., Saxe,C.L.,III, Kimmel,A.R., Johnson,R.L. and Devreotes,P.N. (1988) A chemoattractant receptor controls development in *Dictyostelium discoideum*. *Science*, **241**, 1467–1472.

26. Muramoto,T., Suzuki,K., Shimizu,H., Kohara,Y., Koriki,E., Obara,S., Tanaka,Y. and Urushihara,H. (2003) Construction of a gamete-enriched gene pool and RNAi-mediated functional analysis in *Dictyostelium discoideum*. *Mech. Dev.*, **120**, 967–975.
27. Maruyama,K. and Sugano,S. (1994) Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene*, **138**, 171–174.
28. Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S. (1997) Construction and characterization of a full length-enriched and a 5′-end-enriched cDNA library. *Gene*, **200**, 149–156.
29. Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki,Y., Nakamura,Y., Suyama,A. and Sugano,S. (2001) *EMBO Rep.*, **2**, 388–393.