Published in final edited form as:

Methods Mol Biol. 2013 ; 983: 59–92. doi:10.1007/978-1-62703-302-2_4.

One stop shop for everything *Dictyostelium*: dictyBase and the Dicty Stock Center in 2012

Petra Fey, Robert J. Dodson, Siddhartha Basu, and Rex L. Chisholm

dictyBase and the Dicty Stock Center, Northwestern University, Center for Genetic Medicine, Lurie 7-125, 303 E Superior Street, Chicago, IL 60611, USA

Abstract

dictyBase (<http://dictybase.org>), the model organism database for *Dictyostelium discoideum*, includes the complete genome sequence and expression data for this organism. Relevant literature is integrated into the database, and gene models and functional annotation are manually curated from experimental results and comparative multigenome analyses. dictyBase has recently expanded to include the genome sequences of three additional Dictyostelids, and has added new software tools to facilitate multigenome comparisons. The Dicty Stock Center, a strain and plasmid repository for *Dictyostelium* research has relocated to Northwestern University in 2009. This allowed us integrating all *Dictyostelium* resources to better serve the research community. In this chapter, we will describe how to navigate the website and highlight some of our newer improvements.

Keywords

Dictyostelium discoideum; database; genomic sequence; multigenome; genome browser; Blast; gene page; functional annotation; strains; phenotypes

1. Introduction

Established in 2003, dictyBase (<http://dictybase.org>) is the central repository of genome sequence data for *Dictyostelium discoideum* [1, 2]. It is the single portal to the most comprehensive, current, and highly curated database available online for this important model organism. The primary goals of dictyBase are to facilitate and promote the use of *Dictyostelium* as an experimental system, and to serve the needs of the *Dictyostelium* research community. Accuracy, usability, and service are our highest priorities. To accomplish these aims, dictyBase integrates the genome data with published research, provides research tools to analyze and retrieve data, and maintains a forum for collaboration within the research community, including an archived ListServ, a colleague database, and a history environment for which users can submit content.

dictyBase houses the entire *Dictyostelium* genome, which consists of an approximately 34 Mbp nuclear genome [3], the 55-kb mitochondrial genome [4], the extrachromosomal ribosomal RNA genes [5], and over 163,000 EST sequences [6, 7]. Nearly 13,000 genes have been identified, and these serve as the central focus of dictyBase. Each gene has an individual gene page consisting of all relevant data and information pertaining to the gene and, if applicable, its protein product, including sequence, function, orthologs, phenotypes, literature references, and gene ontology terms. Automated processes were initially used to

assign gene function, membership within a gene family, and GO terms, but experienced curators refine these preliminary assignments by examining experimental results from the literature. Most recently, curators have completed a manual review and where appropriate, refinement of all *D. discoideum* gene models. On an ongoing basis, curators focus on literature curation to improve functional annotation, add strains and mutant phenotypes, and associate gene ontology terms with gene products.

dictyBase has undergone several rounds of upgrades to improve service and introduce new tools, and is now taking the initial steps toward becoming a genome portal for the Amoebozoa clade. Originally designed to house a single genome sequence, dictyBase has since expanded to accommodate additional Dictyostelid genomes, including *Dictyostelium purpureum* [8], *Dictyostelium fasciculatum*, and *Polysphondylium pallidum* [9]. These additions prompted the development of a 'unified' BLAST server and an updated Genome Browser in which all four organisms are available for comparison. In addition, the Dicty Stock Center, a warehouse for *Dictyostelium* strains and plasmids, has relocated to Northwestern University in April 2009, providing a unique opportunity to completely integrate these two resources. In this chapter, we will elaborate on these points to give both the new and seasoned user insights on how to make optimal use of the abundant information in dictyBase.

2. dictyBase Access and Text Search

dictyBase contains a wealth of data and information including genomic sequence data, gene annotations, technical protocols, literature, the Dicty Stock Center repository, and colleague profiles. All information can be accessed from the dictyBase front page, either through direct links and dropdowns, or through the search box. The front page contains announcements and news in the **center**. The **left** side contains direct links to pages of high interest, often added or changed in close discussion with our users. On the **right** side, the most recent *Dictyostelium* literature is displayed on a weekly basis.

2.1. The dictyBase top bar

The **blue top bar** can be accessed from all dictyBase web pages, and contains seven categorized links to different sections of the database. These include websites for related genomes (for more information see **Subheading 7**), educational resources and training protocols, tools, such as the BLAST server (**Subheading 6**) and the genome browser (**Subheading 5**), the Stock Center pages (**Subheading 8**), a download section (**Subheading 3.3**), as well as forms for abstract submission and colleague entries. For a detailed listing see Table 1.

2.2. General Search

Nearly every dictyBase web page contains a simple search box in the upper right where any search term (e.g. cAMP, myo, STE, DDB_G0289129) can be entered to search the entire database. See Table 2 for a complete list of searchable fields.

TIP: To **widen the search**, e.g. for gene families with the same name stem or if unsure of the gene name, or Gene ontology term, use the wildcard character (*).

2.3. Gene name search

There is an extra button to restrict the search to gene names, for a faster return of the search output. For distinct gene names, the search output yields the specific gene page. However, if the gene name is a synonym of another gene, an intermediate screen appears displaying the search results.

TIP: To open a gene page directly without going through search, type or paste a gene name or gene ID directly into the simple link ‘dictybase.org/gene/NAME/ID’. This works for primary gene names and identifiers.

3. Contents of dictyBase

3.1. Data and Annotations—dictyBase is the central repository for information related to the species *D. discoideum*, and increasingly including other Dictyostelids (see **Subheading 7**). Data for *Dictyostelium* has been both imported from external resources, or actively and continuously annotated by curators at dictyBase. The data and annotations currently available at dictyBase are listed in Table 3.

3.2. The Curated Model—Each **curated model** is a curator reviewed gene model derived from careful inspection of an automated gene prediction. It may be an exact copy of the prediction or altered in one or several aspects, based on supporting evidence such as ESTs, RNAseq expression, or sequence similarity. Once a curated model has been created, it replaces the automated gene prediction as the sequence available from the gene page (see **Subheading 4**). A curated gene model is indicated by a green check mark and the supporting evidence is listed. An explanatory note may be added if needed (e.g., see Fig. 1).

As of June 2011, all gene predictions have been individually inspected and a curated model has been added where evidence allowed. This resulted in **11,987 curated models**, which includes all gene models annotated for 33 genes with multiple transcripts and 650 curated pseudogenes. Annotated pseudogenes in dictyBase are genes that have significant homology to existing *D. discoideum* protein coding genes, as well as a confirmed frame shift and/or deletion. When a gene model was incorrectly predicted due to a frame shift in the genomic sequence and there was evidence that the genomic sequence was incorrect, e.g. from ESTs, curators added ‘artificial gaps’ in gene models to create the best possible open reading frame. Artificial gaps restored the correct protein sequence when the underlying genomic sequence had insertions, but in case of deletions it resulted in a loss of one or more amino acids. Extensive curator notes explain the situation, and in cases where the complete protein sequence could not be restored in the database, the correct sequence has been entered on the wiki page (see **Subheading 3.5**). Two examples of large groups of genes for which a curated model could not be determined are the transposable and retrotransposable elements where gene structure does not follow consensus exon/intron organization and there is little, if any, support available (see Table 3).

TIP: Pseudogenes are indicated by addition of a “_ps” after the name, which in turn stems from the gene they are most similar to, for example pks4_ps. **Transposable and retrotransposable elements** contain a _TE or _RTE appendix, such as DDB_G0282919_TE, DDB_G0273327_RTE respectively.

3.3. Downloading Data—dictyBase provides a central **download page**, linked from the top bar, from which users can obtain sequence data and continuously updated annotation files including gene ontology and mutant phenotypes, dictyBase ontologies, and protein domains. For a list of downloadable items see Table 4.

TIP: If you need to convert identifiers from one type to another, an **ID Converter** tool is available under Tools from the top bar (<http://dictybase.org/tools/convert>). You may convert former DDB into current DDB identifiers, current DDB into DDB_G and UniProt identifiers, and any combination thereof.

3.4. Other dictyBase Pages—In addition to the data and annotations, dictyBase contains many resources such as a *Dictyostelium* tutorial “Learn about *Dictyostelium*”, techniques, teaching protocols, nomenclature guidelines, pictures and videos, the dictyNews, the Dicty ListServ, links to Dicty laboratories, and information concerning the annual international *Dictyostelium* conference. These pages are accessible from the top bar under **Explore** and **Community**. About Us, Contact, and Help pages are available from links to the right of the dark blue top bar.

3.5. Community Annotations—Each gene page in dictyBase links to a corresponding wiki page where all users have an account through which they may add information about the gene, including figures or photos (e.g. http://wiki.dictybase.org/dictywiki/index.php/DDB_G0275559). Curators recognize when new content has been added to the community wiki, and create a new link “View annotation for [gene name]” on the gene page (Fig. 2A). In addition, curators add a summary of the community annotation to the description field on the gene page, which enters the essence of the annotation into the database, thus making it searchable in dictyBase (Fig. 2B).

4. The dictyBase Gene Page

Each gene in dictyBase has its own **Gene Page**, the central resource for all information about the *Dictyostelium* gene and its product(s). Figure 3 depicts a typical **Gene Summary Page** that is divided into sections described in detail in **Subheadings 4.1-4.9**. Most sections link to separate details pages (Fig. 4), which are also accessible from browser tabs on top of the gene page (Fig. 3A). The page is also customizable, as sections can be collapsed by clicking on the “–” button in the navigation tools on each dark blue horizontal bar (Fig. 3B). This information is stored in browser cookies and the section will remain collapsed on every subsequently visited gene page, until the section is opened again on any gene page. The small arrow button on navigation tools returns the display to the top of the page. (Fig. 3B).

TIP: Many sections are only available on a gene page if data exists for these sections, e.g., if there are annotations for gene ontology, or if there are phenotypes curated. If the section is **not visible, no data** is available!

TIP: Each section has a button with a question mark to the far right next to the navigation tools on the dark blue horizontal bar (Fig. 3B), which links to a pop-up **Help** page. In addition, each tab at the top of the page has a link to **Help** pages.

4.1. General Information—The General Information section (Fig. 3C) on top contains the primary gene name, a name description, the unique gene ID, the gene product, any alternative gene or protein names, a short description, and links to the community annotation wiki page. Note the red link to an existing community annotation (see also **Subheading 3.5**).

TIP: To ensure finding the same gene page on the next visit, make a note of the unique and stable gene ID (DDB_G). In the very rare case that the gene ID becomes obsolete, you will be redirected to the new ID.

4.2. Genomic Information—This section (Fig. 3D) displays location, chromosome number, coordinates, and orientation of the gene, as well as a gene map that links to the Genome Browser (**Subheading 5**).

TIP: Gene Models in the **Watson** orientation are in red with the arrow pointing to the right, as the gene is located on the top DNA strand (depicted also in Fig. 1). Gene tracks in the

Crick orientation are depicted in blue with an arrow pointing to the left (as shown in the Gene Page example, Fig. 3 D)

4.3. Gene Product Information—The Gene Product Information section (Fig. 3E), displays and links to numerous data and tools. First, the curation status of a gene model (see also **Subheading 3.2** and Fig. 1), tells the user if the gene model has been curated, and what evidence was available to support the call. Next, there are protein length, molecular weight, and a link to the Protein Page (Fig. 4A) on the left, and the exon coordinates are listed on the right. This section also provides access to **BLAST**, and in a dropdown the sequence one wishes to BLAST - protein, cDNA, or genomic sequence - can be preselected and auto-filled. The BLAST Server (see **Subheading 6**) can also be accessed from a tab in the top bar of each gene page. In addition, the sequence of the gene can be accessed from here, by preselecting the sequence type and clicking the **Get Fasta** button, with the sequence opening in a new window.

TIP: Many **Details Pages** such as Protein, Gene Ontology, Phenotypes, BLAST, or References are linked from their respective sections, but can also be accessed from the tabs at the top of the page (Fig. 3A).

TIP: For genes annotated with multiple transcripts, this section features sub-tabs to view the different **Splice Variants** separately. The same applies for the Protein Page when the sequence is different, as protein domains may be affected (e.g. DDB_G0284327; see also **Subheading 4.4**, below).

4.4. Protein Information—Limited protein data including the sequence is available from the Gene Summary Page (Fig. 3E, F), however, in-depth protein information is available from the link in the corresponding section or from the **Protein Tab** on top of the page (Fig. 3A). On top of the **Protein Page** (Fig. 4A), the general info focuses on information including the dictyBase sequence ID (e.g. DDB0238349), the protein length, the molecular weight, and a link to the amino acid composition. In addition, this section displays annotations imported from UniProt such as ‘subcellular location’ and ‘protein existence’. Next, external links to protein sources in UniProt and GenBank [10] are listed, followed by a section with **Protein Domains** displayed in a graphical as well as tabular form obtained from InterPro [11]. On mouse-over of the individual domains a link-out to the source of the domain (e.g. InterPro, Pfam [12]) appears. Click the link ‘Table view’ below the graphical domain image to retrieve a table of the protein domains associated to the gene. The final section shows the **Protein Sequence**.

TIP: Search for a protein domain (e.g. search for ‘Myosinheavy’) in the general search box on top of every page and retrieve the list of genes that have this domain. When the search term hits in several searchable fields (e.g. searching for the domain ‘MIP’ (Major Intrinsic Protein)), click on the link ‘3 protein domains’ on the intermediary search result output to retrieve the gene lists for each of the 3 MIP domains (PF00230, PF09140, PS00221).

4.5. Associated Sequences and Regulatory Elements—This section contains links to GenBank records and ESTs (Fig. 3G). Note, this section is only available when associated sequences are present. The same applies to the ‘Regulatory Elements’ section, which is not available in the gene depicted in Figure 3 (as an example for Regulatory Elements see the *sfrA* gene).

4.6. The Gene Ontology (GO)—The GO (www.geneontology.org; [13]) is a project to produce a controlled vocabulary for annotating gene products that can be applied across all organisms. For many users, the GO provides a quick overview of the cellular role of a gene,

however, GO can also be used for analysis of high-throughput proteomics or expression experiments [14-16]. The GO consists of three categories: molecular function, biological process, and cellular component. GO annotations are displayed on the Gene Page (Fig. 3H). The listed GO terms also include **Evidence Codes** indicating the type of supporting information for a given annotation (<http://www.geneontology.org/GO.evidence.shtml>). A link at the top of the GO section 'View evidence and references' (see Fig. 3H) as well as the **Gene Ontology Tab** on top of the page lead to a detailed **Gene Ontology Page** (see Fig. 4B) where all GO annotations for that gene are listed with their evidence and reference. The reference given for a GO annotation describes an experiment or analysis as the source of the annotation; this reference is often a published paper but can also be an unpublished method, for example, tools to assess sequence similarity. Each GO term links to a page listing all genes that are annotated to that GO term including the reference, and an external link leads to AmiGO, the GO consortium's term and annotation browser [17].

Tip: Annotations with the evidence code **IEA** (inferred from electronic annotation) are purely automated and, although often correct, may contain inaccurate information. Evidence codes other than IEA, e.g. **IDA** (inferred from direct assay), or **IMP** (inferred from mutant phenotype) have been reviewed by a curator and come from a peer-reviewed publication and therefore are indicative of higher-quality annotations.

TIP: The detailed GO page also contains other proteins that have been shown to **interact** with the protein being annotated. For example, with the evidence codes **IGI** (inferred from genetic interaction) or **IPI** (inferred from protein interaction), the annotations include the protein(s) that interact.

4.7. Orthologs—To maximize the knowledge gained using *Dictyostelium*, it is very important to be able to compare known functions of genes with their counterparts from other species and vice versa. To facilitate those analyses, dictyBase gene pages include an **Orthologs Tab**, which lists orthologs of 8 different species (if data is available): *Dictyostelium purpureum*, Homo sapiens, Mus musculus, Drosophila melanogaster, Caenorhabditis elegans, Saccharomyces cerevisiae, Arabidopsis thaliana, and *Escherichia coli* (see Fig. 4C; note that for this gene available ortholog data is from human and mouse only). Ortholog data was obtained from InParanoid [18] and OrthoMCL [19], and in the case of *D. purpureum* from A. Kuspa (private communication). The data is shown in the table containing the species name, a link to the sequence used to calculate the orthologs (usually the model organism database for the species, or Ensembl [20]), a link to UniProt (when available), and the gene product name. Note that InParanoid and OrthoMCL calculate both orthologs and paralogs.

4.8. Phenotypes—Phenotype analysis resulting from mutations is a widely used and informative method to understand gene function. To improve consistency and searchability, we have developed standardized methods for strain curation, including a 'strain descriptor' to address the lack of uniformity in strain nomenclature; a list of useful strain characteristics (such as 'overexpressor', 'drug resistant', or 'null mutant'), and a list of the genetic modifications a strain might have (endogenous deletion, insertion, etc). Phenotypes are also captured with a controlled vocabulary that is constantly being expanded as new phenotypes are described, consisting of a phenotype ontology. The phenotype ontology consists of two composite terms: (1) the anatomical part [21] or the biological process changed in the mutant, and (2) a quality describing that modification. For example, a 'delayed aggregation' phenotype qualifies the 'aggregation' (biological process) as being 'delayed', or the 'decreased spore size' qualifies the 'spore' to be of 'decreased size'. Curation at dictyBase and the Dicty Stock Center (see **Subheading 8**) are now completely integrated [22].

Phenotypes are listed next to their respective **Strains** on the Gene Page (Fig. 3I). This section and the **Phenotypes Tab** on top link to a separate details page (Fig. 4D), where strains are listed along with their phenotypes, strain characteristics, and the reference that describes the mutation. Each strain links to a detailed strain page (described in detail in **Subheading 8.1**) and each phenotype links to a page listing all genes annotated with that phenotype.

TIP: Phenotype terms can be searched using the dictyBase search box on top of every web page, and to improve searchability, most terms have useful synonyms, for example ‘small slug’ for ‘decreased slug size, or ‘agg-’ for ‘abolished aggregation’.

4.9. Links—The next section contains internal as well as external **Links** (Fig. 3J). The page links to expression data from microarrays and RNAseq analyses [23], to *Dictyostelium* Researchers working on the gene (linked by curators), and reciprocal links to external resources such as Inparanoid [18], GenBank Protein [10], UniProtKB (also displaying ID) [24, 25], and ENA, the European Nucleotide Archive [26].

4.10. Summary—The summary section consists of two parts: a curator-composed **Gene Summary** (Fig. 3K), and a semi-automated **Curation Status** note (Fig. 3L). The manual summary has been added to 647 genes as of February 2012. Curators summarize the content of curated papers, and update the summary when a new paper is published (see also **Subheading 9**). The separate curation status note below the summary informs the user when the gene page has last been updated and to what extent. There are 4 different notes:

- Genes that are not curated contain the automatic note: “*This gene has not been manually annotated*”
- Genes that have a curated model but no other manual annotations: “*A curated model has been added, Date, Curator Initials*”
- Genes with basic annotations: “*Basic annotations have been added to this gene, Date, Curator Initials*”
- Genes comprehensively annotated: “*Gene has been comprehensively annotated, Date, Curator Initials*”

TIP: A gene may be ‘**comprehensively**’ annotated although there is limited data available. This means **all publicly available data** has been annotated by the date noted. If any new paper has since been published, the date in the note will be updated when a curator annotates that new publication. Finally, a gene does not need to have a manual summary paragraph to be considered ‘comprehensively’ annotated.

4.11. References—The five most recent articles are displayed on the Gene Page. The full list of references for a gene can be accessed from the Gene Page through the **References Tab** on top or by clicking on View Complete List of References (see Fig. 3M, reference list truncated). On the References Page, the publications are listed with authors, title, a link to dictyBase Curated Paper (see below), PubMed and when available ‘Full Text’ at the journal site. Listed to the right of those links are other genes addressed in the paper. On the left side of the page is a list of **Literature Topics**. The Literature Topics are general categories such as Disease Related, Development/Morphogenesis, and Endocytosis, and attempt to provide a quick overview of the focus of the paper. For example, to view references that discuss mutations in the gene, click on the Mutants/Phenotypes category on the References Page to receive a list of relevant papers. The table can also be filtered by any keyword using the filter box on the upper right.

The Literature Topics are assigned by curators for every gene discussed in the article. Therefore, each gene described in a paper may have a different combination of topics. This information can be viewed on the **dictyBase Curated Paper** page, which is accessible by clicking on the **dictyBase paper** icon (the left of three icons shown next to each reference in Figures 3M 4B, D, and E). The Curated Paper page contains the abstract and a table displaying the genes addressed as well as their Literature Topics.

TIP: Type a name or term into the filter box on the References Page, which might be useful when a gene has a long list of papers and you want to **filter** for a specific author, year of publication or major topic annotated.

5. The Genome Browser

dictyBase genomic annotations are graphically displayed in the Generic Genome Browser (GBrowse), a versatile and customizable tool developed by the Generic Model Organism Database Project (GMOD; www.gmod.org). The genomic position of genes and gene models, both curated and automatically predicted, is based on their chromosomal coordinates on the genome sequence [3]. Other annotations such as GenBank records, ESTs [6, 7], and interspecies Blast hits (TBLASTN) between Dictyostelids are shown as alignments to the genome sequence. RNA sequence data [27] represents individual nucleotide reads from multiple developmental time points that are available as single tracks or compiled into one track. Gbrowse is accessible from the top bar of each Dictyostelid database. (see also Table 1).

5.1. GBrowse Search Functions—The search box **Landmark or Region** in the upper left corner of the display (Fig. 5C) allows the user to display any desired location within the genome. Searchable terms include coordinates, any gene name or sequence ID. For example, entering the *D. discoideum* coordinates '3:20,000..35,000' displays the 15 kb region on chromosome 3 between the coordinates 12,000 and 35,000, or '*mhka*' goes directly to the gene. The retrieved sequence, gene model or EST will appear highlighted in your browser window. A search using a wildcard (*) is also possible and allows searching for a whole gene family. For example, a search for *arc** returns a page with a list of each gene identified with that name stem, the coordinates, and a small map indicating the location on the chromosome (or contig).

Below the Landmark or Region search is a dropdown to select the organism to view (Fig. 5D), as dictyBase now hosts *D. discoideum* and three additional species (see **Subheading 7**). The display is identical for all organisms; however, the availability of tracks differs by organism.

TIP: The yellow highlight after searching for a specific sequence is persistent. To **delete the highlight** click the link at the bottom of the screen (Fig. 5M), or in the preference pane (Fig. 5B).

5.2. The GBrowse Main Display—The Gbrowse display shows a selected chromosomal region up to 200 kilobases (kb). At the time of writing the described GBrowse version (V 2.4) for *D. discoideum* was still under development, therefore Figure 5 depicts the *D. purpureum* GBrowse instance and shows 20 kb on sequence scaffold 12 ('scaffold_12:98,644..118,643'). The first line on top of the display indicates the organism viewed, the chromosome, and the coordinates (Fig. 5A). Immediately below are tabs leading to the Browser window, the track selection interface (see **Subheading 5.3**), and preferences (Fig. 5B) followed by the search box and data source dropdown described in **Subheading 5.1** (Fig. 5C, D, respectively).

The graphical display making up the main portion of this page is divided into 3 sections: chromosome Overview, gene Region, and gene Details.

- **Overview:** displays the highest order of genome assembly in a gray panel, with the current location enclosed within a red box with a darker shade on top for navigation (Fig. 5E). The genome assembly in *D. discoideum* is on the chromosomal level and the overview displays contigs assembled on a chromosome. *D. fasciculatum* and *P. pallidum* contigs are assembled into ‘super-contigs’ and displayed in the overview, while the *D. purpureum* genomic sequence is assembled in scaffolds (super-contigs), but no lower level contigs are available. Thus, Figure 5 shows the overview panel of the *D. purpureum* browser displaying individual genes on a scaffold and not contigs on a higher-level assembly as the other three genomes.
- **Region:** typically features the coordinates and genes selected in the chromosomal region as described above, again highlighted in red with a darker navigation bar on top. As is the case in the Overview, the darker top of the highlighted box can be moved with your mouse from side to side in a defined range –tripling the region that can easily be viewed, indicated by dotted lines (e.g. 20 kb has a range of 60 kb) (Fig. 5F).
- **Details:** displays each chromosomal feature, represented by individual tracks (Fig. 5G-J). The Gene and the Gene Prediction tracks (Curated Models in the annotated *D. discoideum* database) are turned on by default (see more about track selection in Subheading 5.3).

TIP: In the upper left corner of the Details section (Fig. 5G) you find a small **Ruler** – double click to open and grab it with your mouse to move it to any desired location. This tool allows carefully checking how well any tracks align.

5.3. Track Selections and Configuration—To select the chromosomal features you wish to view, click either the ‘Select Tracks’ tab on top or the button at the bottom of the page. A window opens from which tracks can be selected by clicking on the check boxes (see Table 6A for a full list of available tracks). Clicking on ‘Back to Browser’ or on the Browser tab returns to the browser display. The newly selected track(s) will appear below the currently displayed tracks. To change the track order, move your mouse over the highlighted track name and drag the track up or down. These track settings are then saved in your browser by default. Note that the number of tracks to be displayed and the size of the region determine the loading speed of the browser window.

To the left of each track name are buttons to show/hide, turn off, share, customize, or download the track (Fig. 5L, also Fig. 6A). To customize the track click on the ‘tool’ button (Fig. 5Le) and a pop-up window opens to configure the track. For example, the EST track can be changed from showing all ESTs on one line (‘compact’), showing each EST separately (‘expand’; as depicted in Fig. 5), or adding the name to each individual EST (‘expand and label’). By default the EST setting is on auto for an automatic adjustment based on the amount of ESTs available. See Figure 5L for a description of each track option.

TIP: Choose your **Favorite Tracks** by clicking on the star symbol on their left (Fig. 5La), either in the Select Tracks tab or in the Main Display. Once set, after you have changed the settings temporarily the favorite settings can be reactivated with one click in the Track Selection interface. One example of the benefit of displaying different tracks in GBrowse is depicted in Figure 6A, which shows all available developmental time points of RNAseq expression for *D. purpureum* [27]. The data shows that this gene, the ortholog of *D. discoideum* calcineurin A, is strongly expressed during development (Fig. 6A3-6) and in prestalk cells (Fig. 6A9).

TIP: There is a scale in the center of the RNAseq tracks (Fig. 6A12), which is generated automatically by comparing tracks and by taking neighboring gene expression levels into account. Thus, the scale has to be considered when evaluating expression levels.

5.4. GBrowse Navigation—The Scroll/Zoom navigation tools in the upper right of the display (Fig. 5O) allows moving left or right; by selecting the arrows or double arrows the view shifts by either 50% or 100% of the currently displayed genome segment, respectively; the plus (+) and minus (–) buttons zoom in or out by 10%. The drop-down option displays preselected sizes between 100 bp and 200 kb, remaining centered on the currently displayed sequence. In the web browser, the Scroll/Zoom tools are easily recognizable by their yellow buttons.

Navigation is even more intuitive directly in the Overview, Region, or Details areas (described above, **Subheading 5.2**), in which you may simply draw any size window with your mouse up to the allowed size of 200 kb. The zoomed window size in kb is visible and updates as you draw the window. Furthermore, when you draw the zoom window in the Details area you have options to ‘Zoom in’, ‘Recenter on this region’, and ‘Dump selection as FASTA’. Click the box next to the Scroll/Zoom navigation tools and **Flip** the image to change the orientation of the genes.

TIP: Zoom in to 100 bp or less and view the DNA sequence when the DNA/GC Content track is turned on. Zooming out will show a GC-content plot indicating possible coding regions.

5.5. Sequence Downloads and Other Configurable Operations—Above the Scroll/Zoom navigation tool additional operations are available (Fig. 5N). It is recommended that each operation first be configured. The **Configure...** and **Go** buttons will execute the respective commands on the operation selected in the pull-down menu. Currently it is possible to choose between the following operations.

5.5.1 Annotate Restriction Sites: The position of restriction sites of interest may be displayed in GBrowse. Clicking on **Configure...** when **Annotate Restriction Sites** is chosen in the pull-down leads to a list of restriction sites from which any number can be selected, and the configuration will be stored in your browser. Note that to view the restriction sites the **Restriction Sites** track (see Table 6A) must be turned on.

5.5.2. Download Decorated FASTA File: This is a very useful tool for visualizing intron/exon boundaries, EST and RNAseq alignments, and more. There are numerous decoration options: upper/lowercase letters, font style and colors as well as background colors, all of which first need to be configured. Fig. 6B shows an example of how a decorated FASTA file helps to distinguish the first exon of a Curated Model, supported by ESTs and RNA sequence, from the Gene Predictions. Your browser will remember these configuration settings.

TIP: **Flip** your sequence when downloading a gene located on the Crick strand to retrieve the reverse complement of the sequence.

5.5.3. Download Sequence File: When not specifically configured, this option opens a new window with the DNA sequence that is currently covered in the browser. Different sequence formats, such as GFF3, FASTA, GenBank, or Raw Sequence, can be chosen by configuring and the output can be chosen between HTML, Text, or Save to Disk.

5.5.4. Download Track Data: This function allows downloading the data of the current view, including selected tracks, in GFF format and subsequent import into another GBrowse instance. The GFF version can be configured (choose version 2, 2.5 and 3), the region to download, and the tracks.

5.6. Bookmarks, Image Downloads and other Functions—In the upper left on the very top of the GBrowse window is a ‘File’ dropdown and next to it ‘Help’ (Fig. 5P). Under File you will find:

- **Bookmark this.** Clicking on this link renders the web address of the current window into a unique link that can be bookmarked.
- **Share/ these tracks.** Export all currently selected tracks to another GBrowse genome browser by first copying the given URL, then going to another GBrowse instance selecting the "Upload and Share Tracks" tab and pasting the URL into the "Import tracks".
- **Export as:**
 - *Low resolution image.* Download the current view as an image and save in the simple .png format.
 - *High resolution image.* This image can be saved as a high quality SVG (Scalable Vector Graphics) file. The SVG image is resizable without any loss of resolution, and can be opened and edited in any vector graphics application such as Adobe Illustrator, from which it can easily be saved as a raster-based image (ESP, TIFF, JPEG).
 - *GFF annotation table.* Download the current data in GFF3 format or *FASTA sequence file.* Another place from where a FASTA DNA file of the current view can be downloaded (compare *Subheading 5.3*)
- **Get Chrom Sizes.** Downloads a text file listing all chromosomes (*D. discoideum*) or contigs (all other species) with their length in base pairs.
- **Reset to defaults.** Resets all parameters to dictyBase default.

Help is available for the Genome browser and information is available about Gbrowse and the currently used database. Finally, ‘Show my user ID’ provides session IDs if you wish to use a script to upload or download browser data from the current session.

6. The Unified Blast Server

dictyBase features a BLAST sever (<http://dictybase.org/tools/blast>) that includes data from all 4 genome sequences contained in dictyBase, and thus, like the new GBrowse, (see **Subheading 5**), serves as a ‘unifying element’ between the different genomes (see also **Subheading 7**). The BLAST tool is accessible from every Gene Page, in which case the selected sequence auto-fills (Fig. 3F), or from the Tools drop-down on the top bar. The BLAST Server offers the choice of different BLAST programs, several different datasets, and configurable parameters. BLAST databases contain all Dictyostelid species currently available at dictyBase. BLAST search results display alignments and provide links to the Gene Page (see **Subheading 4**). The BLAST server also links out to BLAST at NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

6.1. Blast Databases

From a single page, users can blast against all 4 currently available genomes individually, or against all genomes simultaneously (see Table 5). The datasets for all genomes contain

proteins and coding sequences derived from their primary sequence set, which is the curated model (where available, otherwise it is the Sequencing Center gene prediction) in *D. discoideum*, the JGI gene prediction for *D. purpureum*, and the GenBank submitted gene prediction for *D. fasciculatum* and *P. pallidum*. Other common datasets are the chromosomal or assembled DNA sequences. There are databases for expressed sequence tags (ESTs) for *D. discoideum* and *D. purpureum* while for *D. discoideum* two additional databases are provided: the genomic sequence (coding sequence $\pm$ 1000 bp flanking), and the non-coding sequence, which includes annotated pseudogenes and non-coding RNAs.

6.2. Optimizing Blast Options

Because the database is restricted to Dictyostelid sequences, the dictyBase BLAST Server returns results relatively quickly, even when blasting against ‘All’. However, users may optimize BLAST results for their specific needs; this can be accomplished by changing the **E Value**, the **Number of alignments to show**, choosing between different **Word sizes** and five different **Matrices**, or turning **Gapped alignment** and **Filtering** off (on by default). Detailed information about BLAST in dictyBase can be found in the Help documentation (<http://dictybase.org/db/html/help/blast.html>) and in [28].

Tip: Because *Dictyostelium* proteins often contain repetitive (low complexity) regions, it may be useful to turn off the **Filtering** to get the most complete alignment. Note that this slows the search, so it is advisable to either lower the E value (default = 0.1) or decrease the number of sequences in the output (default = 50), or both.

7. Other Genomes

Additional genomes are available from the Genomes drop-down in the top bar. The *Dictyostelium purpureum* genome was the first additional genome added to our new multi-genome environment (<http://genomes.dictybase.org/purpureum>). The *D. purpureum* genome was sequenced as a collaboration between the Joint Genome Institute (JGI) and the Baylor College of Medicine (<http://genome.jgi-psf.org/Dicpu1/Dicpu1.home.html>) and provided directly to dictyBase with the JGI set of gene predictions and including ESTs [8]. The *Dictyostelium fasciculatum* and *Polysphondylium pallidum* genomes have been sequenced by a European consortium [9] and submitted to GenBank (accession numbers ADHC00000000 and ADBJ00000000, respectively), from which the data was imported into dictyBase (<http://genomes.dictybase.org/fasciculatum>, <http://genomes.dictybase.org/pallidum>). *D. purpureum* contains an additional set of gene predictions obtained in-house using the *D. discoideum*-trained geneid server (<http://genome.crg.es/geneid.html>). Like the model organism *D. discoideum*, all additional genomes contain a gene page (see **Subheading 4**) for every predicted gene, and a Genome Browser that displays all sequencing data (see **Subheading 5**). Each organism also has a download page accessible from the top bar (e. g. *D. fasciculatum*: <http://genomes.dictybase.org/fasciculatum/downloads>); for downloadable items see Table 4B. The left side bar on the front page of these organisms contains links to chromosomal sequences (super-contigs for *D. purpureum*, both super-contigs and contigs for *D. fasciculatum* and *P. pallidum*). A list of genes is also available that features a search box; any gene names or gene IDs contained in the table are searchable. These lists can be viewed up to 100 records at a time and each record has a link to its Gbrowse location. Each front page features tables with Genome Statistics; it shows either “Counts” for number of sequences and genes, or “Feature lengths” for minimum, maximum and median lengths.

TIP: The *P. pallidum* and *D. fasciculatum* genomes have been annotated with gene names for gene homologs to named *D. discoideum* genes before submission to GenBank, e.g. *shkD*, *gskA* [9]. These names are of course searchable from the List of Genes for these organisms.

8. The Dicty Stock Center

In spring of 2009, the Dicty Stock Center moved from Columbia University to Northwestern University. This has allowed the integration of dictyBase and the Stock Center by streamlining the strain collection and improving curation consistency. As of March 2012, the Dicty Stock Center collection has grown to over 1,850 strains and more than 700 plasmids. The strain collection is diverse, including natural isolates of different Dictyostelid strains, a large collection of axenic strains including null mutants, REMI (restriction enzyme-mediated integration) mutants, labeled strains for cell biological studies, chemical mutants, tester strains for asexual genetic analysis, and bacterial strains serving as *Dictyostelium* food source. Other materials such as a cDNA library and several antibodies are also available. The collection of these biological materials in a central repository ensures that they will always be readily available to the research community.

8.1. Search the Stock Center—The Stock Center has its own **Drop-down** from the top bar on every page in dictyBase (see also Table 1). The third link in the drop-down leads to the Stock Center search interface. To search the contents of the Stock Center, you must first choose to search either the **Strains** or the **Plasmids** database. By default, strains are searched in ‘All’ fields; choosing a specific field narrows the search. Strain and Plasmid search fields are shown in Table 7. A broad strain search may be restricted using the **Mutagenesis Method** (e.g. Homologous Recombination, Knockdown) or **Strain Characteristics** (e.g. GFP marked, null mutant, hygromycin resistant). These are controlled vocabularies and are added when curators annotate strains from papers. For these two filters a popup window shows the list of terms from which to choose. When possible, plasmids are now linked to the genes that they harbor. Thus many plasmid pages can be reached through the link on the gene page found at the bottom of the Strains and Phenotypes section instead of going through search.

TIP: To restrict your **plasmid** search use **keywords** such as ‘RFP’, ‘expression vector’, or ‘ecmB promoter’. The complete plasmid keyword list is available on request.

8.2. The Strain Descriptor and Nomenclature—Through work on linking strains with phenotype curation we recognized the need for consistent nomenclature to describe strains. Consulting with the community, we devised a Strain Descriptor to provide a quick overview of the key genetic modifications that produced the strain, including the gene name, the promoter, the mutations, and tags or reporter genes. Examples of strain descriptors are:

- acbA-/[ecmA]:GFP: An acbA null mutant expressing GFP under control of the prestalk promoter ecmA.
- [act15]:cdk5:GFP: wild type strain overexpressing (using an actin 15 promoter) GFP fused at the N-terminus to the cdk5 gene.

We encourage researchers to name strains systematically consisting of 2 or 3 capital letters plus a unique serial number (e.g. HJW117 or HDT6). These names will be added as the **Systematic Name**. All published names will also be listed as names and synonyms. Nomenclature guidelines for strains, and also for genes and proteins, can be found at: http://dictyBase.org/Dicty_Info/nomenclature_guidelines.html

8.3. The Strain Details Page—Each strain has its own **Strain Details Page** (Fig. 7), which can be reached through direct dictyBase search, or, somewhat faster, through the Stock Center search (see **Subheading 8.1**). Strain pages are also linked from the gene page of the associated gene in the Strains and Phenotypes section (compare Fig. 3I). The strain details page includes all names and identifiers, a summary, the genetic modification, strain characteristics (e.g. uracil auxotroph, null mutant), the parental strain, the reference, and the

genotype. If the strain is available the depositor is also listed and the strain can be ordered through two buttons ‘Add to Cart’ and ‘Check Out’ at the bottom of the page (see **Subheading 8.4**). In the upper right corner of the page a link to the shopping cart can be found, which is convenient when shopping for several strains (Fig. 7A and B).

Because strain annotations are needed for phenotype curation from literature, many strains in the database are not yet available in the Stock Center. If phenotypes are curated for a strain, the Strain details page is instead named the **Phenotype and Strain Details Page**, and phenotypes are listed at the top of the page. The Phenotype and strain page for the strains that are not available contain a clear message “This strain is not available at the Stock Center” in a red bar on top of the strain section (Fig. 7B).

TIP: If a strain you would like to order is **not available**, send an email to the author(s) of the listed reference and to dictyBase, and we will try to make the strain available through the Stock Center.

8.4. Ordering Strains and Plasmids—There are two major gateways to order strains and plasmids: the ‘Add to Cart’ button on every Strain and Plasmid Details page, and the green shopping cart next to the strain on the phenotype page (Fig. 8A). View your selected items by clicking on the shopping cart on the Strain/Plasmid Details page, or click on ‘View Cart’ at the bottom of the Phenotype Page. In addition, every other Stock Center page has a shopping cart icon in the upper right corner to access your order. When your order is complete (Fig. 8B), click on the ‘Check Out’ button and add your shipping information in the form to process your order. You will receive a confirmation E-mail that the order has been placed. The strains and plasmids are free when ordered for research purposes, but shipping will be charged. Carefully check that your shipping information is correct to avoid unnecessary delays. General order information is available from the Stock Center drop-down (<http://dictybase.org/StockCenter/OrderInfo.html>). When strains and plasmids are received, recipients are responsible for storing these materials in their lab for long-term use following procedures available from dictyBase (http://dictybase.org/techniques/media/dicty_storage.html).

TIP: Answers to many questions regarding the Stock Center can be found on the **Stock Center FAQ** page (http://dictybase.org/StockCenter/FAQ_StockCenter.html)

TIP: To reward researchers who voluntarily submit their materials to the Stock Center, please **cite the original depositor** (reference is included in the order confirmation email) and dictyBase when publishing work that includes these materials.

8.5. Depositing Strains and Plasmids—The Stock Center prefers strains frozen on dry ice or as colonies on lawns of bacteria, but also accepts strains as axenic cultures, lyophilized spores, or spores in silica gel. If strains are sent on plates, please identify the medium and the bacterial strain that were used. Plasmids can be deposited as either DNA or as a transformed bacterial culture. When planning to deposit materials to the Stock Center, send an email to dictystocks@northwestern.edu for notification and to receive the Stock Center FedEx account number. You will then be instructed to fill in strain – and/or plasmid submission forms. These forms are also available from the Deposit link in the Stock Center drop-down (<http://dictybase.org/StockCenter/Deposit.html>). For large orders we have special tables, which will be provided on request.

TIP: Carefully filling out the **Submission Forms** helps tremendously to achieve complete and correct strain annotations. For plasmid submission, it is also highly desirable to include a **Plasmid Map** and if possible, **Sequence** (or a GenBank accession number).

9. New Data and Future Directions

In addition to the continuous curation of the database, dictyBase strives to update the database and add new tools to accommodate an increasing flow of data. As dictyBase has now been public for almost a decade, major operating system upgrades and database refiguring are underway. The following list briefly describes our priorities for the expansion of dictyBase.

- *Comprehensive data mining (Intermine).* We will implement Intermine (<http://intermine.org/>), a data warehouse system with a user-friendly web interface. This allows the user to query the data available in creative ways and download the results to their computer, or store them in the browser.
- *Using text-mining to boost GO annotations.* In collaboration with WormBase (<http://www.wormbase.org/>), we are beginning to use Textpresso [29] to semi-automatically annotate cell components from papers. This is also in the trial phase for molecular function annotations.
- *Annotate other genomes.* dictyBase contains a growing number of ‘sister’ databases with genomes of other Dictyostelids. Because these genomes have little or no annotations, we plan to create a pipeline to semi-automatically annotate gene ontology, gene names and products to orthologous genes using *D. discoideum* as the reference genome.
- *Represent more genomes.* More Dictyostelid species are continually being sequenced and we will add several more databases in our multigenome environment.
- *Integrate SNP data in GBrowse.* We will create another instance of the genome browser displaying different *D. discoideum* strains and plan to add SNP data for easy comparison.

Acknowledgments

We thank Kerry Sheppard for her expert technical work at the Dicty Stock Center, Yulia Bushmanova for her contributions to the new gene page layout and the gene model curation tool, and Pascale Gaudet for her many years as a curator of dictyBase. dictyBase and the Dicty Stock Center are funded by National Institutes of Health GM64426, GM087371 and HG0022.

10. References

1. Chisholm R, Gaudet P, Just EM, Pilcher KE, Fey P, Merchant SN, Kibbe WA. dictyBase, the model organism database for Dictyostelium discoideum. Nucleic Acids Res. 2006; 34:D423–7. [PubMed: 16381903]
2. Gaudet P, Fey P, Basu S, Bushmanova YA, Dodson R, Sheppard KA, Just EM, Kibbe WA, Chisholm RL. dictyBase update 2011: web 2.0 functionality and the initial steps towards a genome portal for the Amoebozoa. Nucleic Acids Res. 2011; 39:D620–4. [PubMed: 21087999]
3. Eichinger L, Pachebat JA, Glöckner G, Rajandream MA, Sugang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hemphill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Goodhead I, Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Madan Babu M, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Loulseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabinowitsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G,

- Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A. *Nature*. 2005; 435:43–57. [PubMed: 15875012]
4. Ogawa S, Yoshino R, Angata K, Iwamoto M, Pi M, Kuroe K, Matsuo K, Morio T, Urushihara H, Yanagisawa K, Tanaka Y. The mitochondrial DNA of *Dictyostelium discoideum*: complete sequence, gene content and genome organization. *Mol Gen Genet*. 2000; 263:514–519. [PubMed: 10821186]
 5. Sugang R, Chen G, Liu W, Lindsay R, Lu J, Muzny D, Shaulsky G, Loomis W, Gibbs R, Kuspa A. Sequence and structure of the extrachromosomal palindrome encoding the ribosomal RNA genes in *Dictyostelium*. *Nucleic Acids Res*. 2003; 31:2361–2368. [PubMed: 12711681]
 6. Urushihara H, Morio T, Saito T, Kohara Y, Koriki E, Ochiai H, Maeda M, Williams JG, Takeuchi I, Tanaka Y. Analyses of cDNAs from growth and slug stages of *Dictyostelium discoideum*. *Nucleic Acids Res*. 2004; 32:1647–1653. [PubMed: 15010511]
 7. Morio T, Urushihara H, Saito T, Ugawa Y, Mizuno H, Yoshida M, Yoshino R, Mitra BN, Pi M, Sato T, Takemoto K, Yasukawa H, Williams J, Maeda M, Takeuchi I, Ochiai H, Tanaka Y. The *Dictyostelium* Developmental cDNA Project: Generation and Analysis of Expressed Sequence Tags from the First-Finger Stage of Development. *DNA Research*. 1998; 5:335–340. [PubMed: 10048482]
 8. Sugang R, Kuo A, Tian X, Salerno W, Parikh A, Feasley CL, Dalin E, Tu H, Huang E, Barry K, Lindquist E, Shapiro H, Bruce D, Schmutz J, Salamov A, Fey P, Gaudet P, Anjard C, Babu MM, Basu S, Bushmanova Y, van der Wel H, Katoh-Kurasawa M, Dinh C, Coutinho PM, Saito T, Elias M, Schaap P, Kay RR, Henrissat B, Eichinger L, Rivero F, Putnam NH, West CM, Loomis WF, Chisholm RL, Shaulsky G, Strassmann JE, Queller DC, Kuspa A, Grigoriev IV. Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. *Genome Biol*. 2011; 12:R20. [PubMed: 21356102]
 9. Heidel AJ, Lawal HM, Felder M, Schilde C, Helps NR, Tunggal B, Rivero F, John U, Schleicher M, Eichinger L, Platzer M, Noegel AA, Schaap P, Glöckner G. Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. *Genome Res*. 2011; 21:1882–1891. [PubMed: 21757610]
 10. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res*. 2012; 40:D48–53. [PubMed: 22144687]
 11. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. InterPro: the integrative protein signature database. *Nucleic Acids Res*. 2009; 37:D211–5. [PubMed: 18940856]
 12. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. The Pfam protein families database. *Nucleic Acids Res*. 2012; 40:D290–301. [PubMed: 22127870]
 13. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25:25–29. [PubMed: 10802651]
 14. Drew K, Winters P, Butterfoss GL, Berstis V, Uplinger K, Armstrong J, Riffle M, Schweighofer E, Bovermann B, Goodlett DR, Davis TN, Shasha D, Malmström L, Bonneau R. The Proteome Folding Project: proteome-scale prediction of structure and function. *Genome Res*. 2011; 21:1981–1994. [PubMed: 21824995]
 15. Plessis, du L, Skunca N, Dessimoz C. The what, where, how and why of gene ontology--a primer for bioinformaticians. *Brief. Bioinformatics*. 2011; 12:723–735. [PubMed: 21330331]
 16. Masseroli M, Pinciroli F. Using Gene Ontology and genomic controlled vocabularies to analyze high-throughput gene lists: three tool comparison. *Comput Biol Med*. 2006; 36:731–747. [PubMed: 16165121]
 17. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, AmiGO Hub; the Web Presence Working Group. AmiGO: online access to ontology and annotation data. *Bioinformatics*. 2009; 25:288–289. [PubMed: 19033274]

18. Ostlund G, Schmitt T, Forslund K, Köstler T, Messina DN, Roopra S, Frings O, Sonnhammer ELL. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* 2010; 38:D196–203. [PubMed: 19892828]
19. Chen F. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.* 2006; 34:D363–D368. [PubMed: 16381887]
20. Kersey PJ, Lawson D, Birney E, Derwent PS, Haimel M, Herrero J, Keenan S, Kerhornou A, Koscielny G, Kahari A, Kinsella RJ, Kulesha E, Maheswari U, Megy K, Nuhn M, Proctor G, Staines D, Valentin F, Vilella AJ, Yates A. Ensembl Genomes: Extending Ensembl across the taxonomic space. *Nucleic Acids Res.* 2009; 38:D563–D569. [PubMed: 19884133]
21. Gaudet P, Williams JG, Fey P, Chisholm RL. An anatomy ontology to represent biological knowledge in *Dictyostelium discoideum*. *BMC Genomics.* 2008; 9:130. [PubMed: 18366659]
22. Fey P, Gaudet P, Curk T, Zupan B, Just EM, Basu S, Merchant SN, Bushmanova YA, Shaulsky G, Kibbe WA, Chisholm RL. dictyBase—a *Dictyostelium* bioinformatics resource update. *Nucleic Acids Res.* 2009; 37:D515–9. [PubMed: 18974179]
23. Rot G, Parikh A, Curk T, Kuspa A, Shaulsky G, Zupan B. dictyExpress: a *Dictyostelium* *discoideum* gene expression database with an explorative data analysis web-based interface. *BMC Bioinformatics.* 2009; 10:265. [PubMed: 19706156]
24. Gaudet P, Lane L, Fey P, Bridge A, Poux S, Auchincloss A, Axelsen K, Braconi Quintaje S, Boutet E, Brown P, Coudert E, Datta RS, de Lima WC, de Oliveira Lima T, Duvaud S, Farriol-Mathis N, Ferro Rojas S, Feuermann M, Gateau A, Hinz U, Hulo C, James J, Jimenez S, Jungo F, Keller G, Lemercier P, Lieberherr D, Moinat M, Nikolskaya A, Pedruzzi I, Rivoire C, Roechert B, Schneider M, Stanley E, Tognolli M, Sjölander K, Bougueleret L, Chisholm RL, Bairoch A. Collaborative annotation of genes and proteins between UniProtKB/Swiss-Prot and dictyBase. *Database (Oxford).* 2009; 2009:bap016. [PubMed: 20157489]
25. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford).* 2011; bar009. 2011. [PubMed: 21447597]
26. Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, Cheng Y, Cleland I, Faruque N, Goodgame N, Gibson R, Hoad G, Jang M, Pakseresht N, Plaister S, Radhakrishnan R, Reddy K, Sobhany S, Ten Hoopen P, Vaughan R, Zalunin V, Cochrane G. The European Nucleotide Archive. *Nucleic Acids Res.* 2011; 39:D28–31. [PubMed: 20972220]
27. Parikh A, Miranda ER, Katoh-Kurasawa M, Fuller D, Rot G, Zagar L, Curk T, Sugang R, Chen R, Zupan B, Loomis WF, Kuspa A, Shaulsky G. Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biol.* 2010; 11:R35. [PubMed: 20236529]
28. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–3402. [PubMed: 9254694]
29. Müller H-M, Kenny EE, Sternberg PW. Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol.* 2004; 2:e309. [PubMed: 15383839]
30. Sawai S, Guan X-J, Kuspa A, Cox EC. High-throughput analysis of spatio-temporal dynamics in *Dictyostelium*. *Genome Biol.* 2007; 8:R144. [PubMed: 17659086]
31. Maeda M, Sakamoto H, Iranfar N, Fuller D, Maruo T, Ogihara S, Morio T, Urushihara H, Tanaka Y, Loomis WF. Changing patterns of gene expression in *dictyostelium* prestalk cell subtypes recognized by in situ hybridization with genes from microarray analyses. *Eukaryotic Cell.* 2003; 2:627–637. [PubMed: 12796308]
32. Maruo T, Sakamoto H, Iranfar N, Fuller D, Morio T, Urushihara H, Tanaka Y, Maeda M, Loomis WF. Control of cell type proportioning in *Dictyostelium discoideum* by differentiation-inducing factor as determined by in situ hybridization. *Eukaryotic Cell.* 2004; 3:1241–1248. [PubMed: 15470253]

Fig. 1.

Graphical gene display and curator statements on the Gene Page. The image (**A**) also serves as a link to the Genome Browser and depicts a gene in the 'Watson' direction (arrow pointing to the right) with the chromosomal coordinates on top. The thinner gene track on top is labeled with the gene ID (DDB_G...), whereas the curated model track is labeled with the sequence ID (DDB...). (**B**) Below the image, optional curator notes are displayed, in this example alerting the user that ESTs indicate an intron in the 5' UTR of the gene. In (**C**), the DDB ID of the curated model links to a separate tab for the sequence page. Next to the ID is a green check mark indicating that a curator has reviewed the gene. Below, notes describe where the curated model is derived from and what evidence supports the curated model.

Gene Name	<i>pich</i>
Name Description	<i>pich</i> = Polo-Interacting Checkpoint Helicase <i>ercc6l</i> = Excision Repair Cross-Complementing repair deficiency, complementation group 6-Like
Alternative Gene Names	<i>ercc6l</i>
Gene ID	DDG_G0288873
Gene Product	polo-interacting checkpoint helicase
Alternative Protein Names	Ercc6-like protein
Description	<p>ortholog of human DNA excision repair protein ERCC-6-like (ERCC6L)</p> <p>A Community annotation: DDB_G0288873, but not <i>ercc6l</i>, is strongly overexpressed (8x) in a Dicty strain in which the retinoblastoma-like gene <i>rbA</i> has been disrupted. Most cells involved in S-phase or mitosis are overexpressed in this strain. <i>Ercc6l</i>-like is an essential component of the spindle checkpoint, which localizes to the kinetochore; in prometaphase it is found on thin threads that stretch between sister kinetochores. It is thought to function as part of the tension sensor, the functional core of the spindle checkpoint [see Baumann et al. Cell 128, 101-114 (2007)]. Other spindle checkpoint components are strongly overexpressed in the <i>rbA</i> null. This suggests that DDB_G0288873 may actually code for the spindle tension sensor in Dicty. <i>Harry MacWilliams, May 2010</i></p> <p>There is more information on the wiki page</p>
Community Annotations	<p>B View annotation for <i>pich</i> Add an annotation for <i>pich</i> Community Annotations Help</p>

Fig. 2.

Community annotation display on the Gene Page. When a community annotation has been added to the wiki page for the gene, curators add a summary of that annotation in the description field, followed by the name of the submitter, date of annotation, and a link noting that there is more information on the wiki page (**A**). Curators also add a bold red link 'View annotation for [gene name]' (**B**). To the right of this link are more links to the wiki that you can find on each gene page: 'Add an annotation for [gene name]' and 'Community Annotations Help'.

Gene Information for p2xA

A Gene Summary | Protein Information | Gene Ontology | Orthologs | Phenotypes | References | BLAST

C **General Information**

Gene Name: p2xA
 Name Description: p2xA = P2X receptor A
 Gene ID: DCB_02272004
 Gene Product: ATP-gated ion channel P2XA
 Alternative Protein Names: DpP2X
 Description: membrane ion channel activated by ATP; may be involved in intracellular calcium signaling
 Community Annotations: [View annotation for p2xA](#) Add an annotation for p2xA Community Annotations Help

D **Genomic Information**

Location: Chromosome 2 coordinates 1165671 to 1167128, Crick strand
 Genomic Map: [Click on the map to browse the genome from this location]

E **Gene Product Information**

Protein Coding Gene: DCB02203446 (Curator reviewed)
 Derived from gene prediction. Supported by mRNA, ESTs, sequence similarity, unpublished transcript sequence.
 Protein Length: 378 aa
 Protein Molecular Weight: 42,185.6 Da
 More Protein Data: Protein sequence, domains and much more...
 Sequence:

F **Associated Sequences**

GenBank mRNA: EU047052
 ESTs: DCB0203948

H **Gene Ontology Annotations** [View evidence and references](#)

Molecular Function: receptor activity (REA), ATP binding (IDA), ligand-gated ion channel activity (IDA), ligand-gated ion channel activity (IDA)
 Biological Process: ion transport (IDA), hypotonic response (IMP), hypotonic response (GPI), cellular water homeostasis (IMP), regulation of calcium-mediated signaling (GPI)
 Cellular Component: contractile vacuolar membrane (IDA)

I **Strains and Phenotypes** [View Phenotype Information](#)

p2xA-: aberrant response to hypo-osmotic stress
 p2xA-: wild type
 p2xA- (p2xB- p2xC- p2xD- p2xE-): decreased growth rate | delayed response to hypo-osmotic stress
 p2xA- (p2E112) p2xA-GFP: wild type

J **Links**

Expression: dictyExpress (RNA-Seq)
 dictyBase Collogues: p2xA Researchers
 External Resources: Interpro | GenBank Protein | UniProtKB: Q8CJN7 | Ensembl | ENA |

K **Summary**

Curator Notes: The p2xA gene is one of 6 Dictyostelium genes encoding putative P2X receptors. p2xA has been heterologously expressed in human HEK 293 cells and was shown to have ATP-gated cation channel activity. Mutation of conserved lysine (K27A, K289A) or a conserved aspartate (D330A) to alanine rendered the molecule non-functional. This P2X receptor was also found to be sensitive to Cu²⁺ ions. Deletion of the p2xA gene in AX4 wild type cells led to extensive swelling of the cells when placed in water, indicating a

L **Curation Status**

Gene has been comprehensively annotated, 11-MAR-2010 PF

M **Latest References** [View complete list of references \(11 papers\)](#)

Journef, Klein, Brugiere, Vandenbroeck, Chapel, Kieffer, Bruley, Masselon & Aubry (2011) Investigating the macrophagocytic proteome of Dictyostelium amoebae by high-resolution mass spectrometry. Proteomics
 Valente, Watterson, Parker, Ford & Young (2011) Expression, purification, electron microscopy, N-glycosylation mutagenesis and molecular modeling of human P2XA and Dictyostelium discoideum P2XA. Biochimica et Biophysica Acta

Fig. 3. The dictyBase Gene Page. **(A)** Tabbed browsing. The tabs on top link to Detail Pages as depicted in Fig. 4. **(B)** Page Navigation tools and link to Help. **(C)** General Information including gene names and product. **(D)** Genomic Information with coordinates and a map that links to Genome Browser. **(E)** Gene Product and gene model Information. **(F)** Links to sequences and BLAST server. **(G)** Associated Sequences. **(H)** Gene Ontology annotations. **(I)** Strain and Phenotype information. **(J)** External Links. **(K)** Gene Summary and **(L)** curation status note. **(M)** Latest References with links to paper details and PubMed.

Fig. 4. dictyBase Details Pages accessible from tabs on the Gene Page. **(A)** Protein Information including protein domains and sequence. **(B)** Complete Gene Ontology Info. **(C)** List of Orthologs. **(D)** Complete gene-related Strain and Phenotype info. **(E).** Complete list of References.

Fig. 5.

The GBrowse Display. A view of 20 kb sequence on *D. purpureum* scaffold 12, coordinates 98,644 – 118,643. Note that the overview for this organism contains the genes present on the scaffold (E; see also **Subheading 5.2**). Tracks selected are: **H**, Genes; **I**, Gene Predictions from JGI; **J**, EST Alignments; and **K**, *D. discoideum* protein alignments. The view depicts 6 genes, one (DPU_G0052772) on the Watson strand (arrow head to the right) and the other genes on the Crick strand (arrow heads to the left). Three genes have EST alignments, and all genes have at least one *D. discoideum* protein aligned. Note the small ruler on top of the Details region (G); it can be expanded and moved to any desired location on the genomic stretch to check alignments. **L** shows an expansion of the available track options. From left to right, the functions are as follows: **a** add track to favorites; **b**, show or hide this track; **c**, turn off the track (may be activated again in track selection panel); **d**, share or export this track to another Gbrowse instance; **e**, configure the appearance of this track; **f**, download this track (Fasta, GFF3, GenBank); **g**, about this track.

Fig. 6.

A. RNAseq tracks of different developmental time points displayed in Gbrowse. On top are the gene track and gene prediction (1) followed by the RNAseq developmental time points with a 4 hour interval, displayed chronologically from top down: 2, 0 hrs; 3, 4 hrs; 4, 8 hours; 5, 12 hrs; 6, 16 hrs; 7, 20 hrs; 8, 24 hrs. This is followed by two tracks that are “RNAseq 48 hrs prestalk” (9) and “RNAseq 48 hrs prespore” (10), where slugs were allowed to migrate for 48 hours before separating the cell types. The final track (11) contains all time points compiled into a single track. This gene, DPU_G0057736, the ortholog of *D. discoideum calA*, show increasing expression early in development (3) and a drop after 16 hrs (7); expression is higher in prestalk (9) than in prespore (10) cells. **B.** The GBrowse decorated FastA file. On the left (1) the decoration of this file is as follows: Yellow highlight: curated model; pink letters: Overlapping Sequencing Center and Geneid gene predictions; underline: ESTs. To the right (2) is an overview with the gene track and the curated gene model on top. Both gene predictions (3) are below, followed by ESTs (collapsed into one line), and RNAseq (all time points) aligned at the bottom. Note that the first exon of both gene predictions is much further upstream than the correct very short first exon (bases AT of the start codon) supported by ESTs and RNAseq. Start sites of gene predictions, curated model and supporting sequences are indicated by **arrowheads**.

A Strain Details for DH1			
Strain Details			
Strain Descriptor	DH1	Strain ID	DBS0235701
Strain Names	DH1-DR, dh1(dr), dh1-dr, dh1 d.r., DH1 DR, DH1DR	Systematic Name	DH1
Strain Summary	Doug Robinson lab strain; axenic; uracil auxotroph	Genotype	axeA1, axeB1, axeC1, pyr5-6-[pRG24], ura-
Genetic Modification	endogenous deletion	Mutagenesis Method	Homologous Recombination
Strain Characteristics	uracil auxotroph, axenic	Depositor	Doug Robinson
Parental Strain	AX3 (DBS0235542)	Species	Dictyostelium discoideum
Plasmid	pRG24 + FOA selection	Vector	
Reference(s)	8288619, d4829		
Associated Gene(s)	pyr56		
		Add to Cart Check Out	

B Phenotype and Strain Details for fszB-			
Phenotype	Notes	Reference	
decreased number of mitochondria	Null mutant contains about 30% fewer mitochondria than parent	Gilson, Yu, Hereld, Barth, Savage, Kiefel, Lay, Fisher, Margolin & Beech (2003) Two Dictyostelium orthologs of the prokaryotic cell division protein FtsZ local Eukaryot Cell 2:1315-26
Strain Details			
This strain is not available at the Dicty Stock Center			
Strain Descriptor	fszB-	Strain ID	DBS0235980
Strain Names	ΔfszB	Systematic Name	DBS0235980
Strain Summary	fszB null mutant in AX3	Genotype	axeA1, axeB1, axeC1, bsR, fszB-
Genetic Modification	endogenous deletion	Mutagenesis Method	Homologous Recombination
Strain Characteristics	blestidin resistant, axenic, null mutant	Depositor	
Parental Strain	AX3	Species	Dictyostelium discoideum
Plasmid	pWM1325	Vector	
Reference(s)	14665465		
Associated Gene(s)	fszB		

Fig. 7. Strain Details Pages. As for all Stock Center pages, all strain records also have a shopping cart icon in the upper right corner making browsing of strains easy while the shopping cart can be accessed from everywhere. **(A)** A strain page of a wild type strain that is available at the Stock Center. Note the ‘Add to Cart’ and ‘Check Out’ buttons at the bottom. **(B)** Details of a mutant strain with one phenotype associated, but not yet available with a note in red above the strain record stating “This strain is not available at the Dicty Stock Center”.

A

abpC-/[ecmA]: abpC:GFP (DBS0237948)	null mutant, rescued	wild type
abpC- (HG1270)	mutant	aberrant phototaxis
View Cart			

B Shopping Cart Contents

The items listed below are currently in your shopping cart:

ID	Strain / Plasmid name	Remove From Cart
Strain-HG1270	abpC-	Remove
Strain-DBS0235456	abpA- / abpC-	Remove

[Check Out](#)

Fig. 8.

Strain Selection from Phenotype List on Gene Page and Shopping Cart Content. **(A)** A strain can be added to the shopping cart directly from the list of phenotypes on the Gene Page Tab (see also Fig. 4D). At the left of each strain/phenotype is either a green shopping cart when available (bottom strain), or a crossed-out gray cart when unavailable (top strain). **(B)** The Shopping Cart lists all items added to the cart with an option to remove it. When finished selecting click the 'Check Out' button at the bottom.

Table 1

Linked Pages from the dictyBase Blue Top Bar. Links are available from all dictyBase pages, categorized into topics (bold) and organized as dropdown menus. In case of Download the top item directly links to the central download page (see Table 4).

Genomes	Explore	Research	Tools	Stock Center	Download	Community
<i>Dictyostelium</i> discoideum	Learn About Dicty	Techniques	Genome Browser *	DSC Home		Submit abstract dictyNews
<i>Dictyostelium</i> purpureum	Teaching Protocols	Dicty Anatomy Ontology	BLAST	About Stock Center		Read dictyNews
<i>Dictyostelium</i> fasciculatum	<i>Dictyostelium</i> Genome Resources	Mutant Phenotypes	ID Converter	Search Stock Center		ListServ Archive
<i>Polysphondylium</i> pallidum	<i>D. discoideum</i> Genome Statistics	HTP Phenotyping Princeton	dictyMart	Order		Add/Update Colleague Profile
	Pictures/ Videos	Transcriptome Browser Baylor	Textpresso	Deposit		Dicty Annual Conference
	dictyArt	Codon Bias Table	Biochemical Pathways	Strain Catalog Plasmid Catalog		Job opportunity
	Useful Links	Nomenclature Guidelines	Third Party Tools	Bacterial Strain Catalog		History
	Virtual Library <i>Dictyostelium</i>	Axenic Strain History		Additional Materials		Dicty Labs
		Franke Reference Library		Stock Center FAQ		Citing dictyBase
				Nomenclature Guidelines		
Other Stock Centers						

* The Genome Browser opens another dropdown on mouse-over that links to the different

Table 2

Searchable Items in General Search. The fields and pages searched when doing a general search via the search box found in the upper right corner of each dictyBase page.

<ul style="list-style-type: none">• Gene Names or Synonyms• Gene IDs• Gene Product Names• dictyBase IDs• Gene Descriptions• Protein Domains• EST identifiers• Gene Ontology Terms or Synonyms• Strains, Synonyms, Systematic Names or dictyBase Strain IDs• Phenotype Terms or Synonyms• Plasmid Names• Colleagues by Last Name• Authors by Last Name• GenBank Nucleotide Accession Numbers• Web pages (including dictyNews)• UniProt IDs
--

Table 3*D. discoideum* Contents at dictyBase as of March 2011.

DATA	ANNOTATIONS
<ul style="list-style-type: none"> • 13,541 Automated Gene Predictions [3] • 1,770 GenBank Records • 163,182 Expressed Sequence Tags [6, 7] • 7,158 PubMed References • External Data: <ul style="list-style-type: none"> ○ 4,123 dictyExpress Microarray expression profiles [23] ○ 12,087 dictyExpress RNAseq expression profiles [23, 27] ○ 2,257 High Through-Put Phenotypes/Princeton [30] ○ <i>In situ</i> hybridization: 150 images (Tsukuba Atlas; [31, 32]) ○ Insertional Mutants: 817 Links (BCM) • 8,620 Genes with Orthologs • 9,955 Proteins with InterPro domains • 1,891 Colleagues • Dicty Stock Center: <ul style="list-style-type: none"> ○ 1,853 strains ○ 716 plasmids 	<ul style="list-style-type: none"> • 11,989 Curated Models[#] • 34 Alternative transcripts • 651 Pseudogenes[#] • 520 Transposable elements (not curated gene models[#]) • Gene products for 9,076 genes • Brief Descriptions for 7,338 genes • Name descriptions for 4,305 genes • Mutant Phenotypes for 822 genes • Gene Ontology annotations for 7,559 genes • Summary paragraphs for 647 genes • 5,735 Genes with basic annotations • 2,822 Genes comprehensively annotated • 275 Genes with Community Annotations[*]

[#] explained in Subheading 3.2,^{*} discussed in Subheading 3.5.

Table 4

Downloads for *D. discoideum* and other Dictyostelid species. A. Downloads for *D. discoideum*. All downloadable files, which contain information that changes or expands regularly, are updated on a weekly or monthly basis.

A. Downloads for *D. discoideum*

- Gene Information
 - dictyBase ID, gene names, synonyms, and gene products
 - DDB-DDB_G-UniProt mapping
 - DDB_G curation status
 - Curated Model History
 - Ortholog information
 - Alternative transcripts
- *Dictyostelium* Sequences and Annotations
 - Sequences in FASTA format [download area]
 - *D. discoideum* protein sequences
 - *D. discoideum* coding sequences
 - *D. discoideum* noncoding sequences
 - *D. discoideum* genomic sequences
 - *D. discoideum* EST sequences
 - *D. discoideum* Chromosomal DNA: 1,2,3,4,5,6,M, and floating contigs
 - Sequences and annotations in GFF3 format
 - Promoter sequences (5' flanking sequence up to next gene)
- Genes on AX4/AX3 Chromosome 2 Duplication1
- Mutant Phenotypes
 - All curated mutants
 - Null mutants
 - Overexpression mutants
 - Multiple mutants
 - Mutants with developmental defects
 - Other mutants
 - Insertional mutants at BCM
- *Dictyostelium* Anatomy Ontology [21]
- *Dictyostelium* Phenotype Ontology²
- Protein Domains
 - InterPro domains
 - N-terminal myristoylation sites
- GO Association File
 - Gene Ontology annotations
 - GO annotations mapped to GO-slim.obo
- References
 - Franke *Dictyostelium* Reference Library [EndNote]
 - List of PubMed IDs and associated genes
 - List of High throughput papers and associated genes

○ List of Reviews and associated genes

○ List of Papers excluding Reviews and High throughput papers and associated genes

- Tools

○ Third Party Tools

<i>B. Downloads for other species</i>			
Downloadable Items	Dictyostelium purpureum	Dictyostelium fasciculatum	Polysphondylium pallidum
Nuclear Chromosomal	X	X	X
Nuclear Coding Sequence	X	X	X
Nuclear Protein Sequences	X	X	X
Nuclear Genome Annotations-GFF3	X	X	X
Mitochondrial Chromosomal	X	X	X
Mitochondrial Genome Annotations-GFF3	X	X	X
EST Sequences	X		
DPU_G -JGI ID mapping	X		
Ortholog information	X		

¹In strains AX3 and AX4 (the sequenced strain represented in dictyBase), chromosome 2 contains a duplication of approximately 750 kb, in which the region from base 2263132 to 3015703 is repeated between bases 3016083 and 3768654.

²The phenotype ontology contains terms used to annotate Dictyostelium phenotypes, updated continuously during literature annotation. B. Downloads for other Dictyostelid species available at <http://genomes.dictybase.org/>. These downloads are standardized across the three species, with species-special downloads added individually.

Table. 5

The dictyBase BLAST databases. Every sequence can be blasted against a single species or all 4 available species simultaneously. The Blast server can be accessed at <http://dictybase.org/tools/bblast>.

BLAST database	All	D. discoideum	D. fasciculatum	D. purpureum	P. pallidum
Protein sequences	X	X	X	X	X
Coding sequences	X	X	X	X	X
Genomic sequences *		X			
EST sequences		X		X	
Non-coding sequences		X			
Chromosomal DNA #		X	X	X	X

* Genomic sequences for *D. discoideum* are defined as coding sequences plus 1,000 bp flanking sequence on each side;

Chromosomal DNA includes all 6 chromosomes, the mitochondrial genome and floating contigs for *D. discoideum*; for *D. purpureum* it is all sequence scaffolds, and for *D. fasciculatum* and *P. pallidum* it contains chromosomal super-contigs plus the mitochondrial genome.

Table 6

GBrowse Tracks and Track Options. **A.** There are many optional tracks that can be turned on in the genome browser. Select tracks and return to browser to view. Tracks can be closed and customized in the browser window. By default, the primary gene model is activated for each genome. This is the curated model in *D. discoideum* (except the approximately 1,000 genes that do not have a curated model where it is the Sequencing Center gene prediction), the JGI gene prediction in *D. purpureum*, and the GenBank submitted genes for *D. fasciculatum* and *P. pallidum*. Note not all tracks available for *D. discoideum* (listed) are available for other species. **B.** Each track has buttons to its left that allow formatting, downloading and sharing, among other functions.

A	
•	Dicty Genes
	○ Genes
	○ Gene Models
	○ Curated Models
	○ tRNAs
	○ ncRNAs
	○ Gene Predictions from Sequencing Center
	○ Geneid Gene Predictions
•	Dicty Alignments
	○ EST Alignments
	○ GenBank mRNA (cDNA) Alignments
	○ CDS from GenBank Genomic Fragments Alignments
•	Genome Assembly/Sequence
	○ 3-frame translation (forward)
	○ 3-frame translation (reverse)
	○ DNA/GC Content
•	Quantitative Data
	○ RNA Sequence Profile Baylor
	○ RNA Sequence Profile HMW
•	Overview
	○ Contigs Overview
•	Region
	○ Genes Overview
•	Analysis
	○ Putative TF binding sites
	○ Restriction Sites

Table 7

Stock Center Searchable fields. The Stock Center items Strains and Plasmids have a separate search interface with specified options to make it easier for Stock Center users to identify the strains they are interested in.

STRAINS	PLASMIDS
<ul style="list-style-type: none"> • All • Depositor • Genotype • Keyword • Mutagenesis method* • Parental Strain • Stock Center Phenotype • Plasmid • Species • Strain ID • Strain descriptor /Synonyms /Systematic names • Strain characteristics * 	<ul style="list-style-type: none"> • Depositor • GenBank accession number • ID • Keywords • Name

* A pop-up window appears when this field is selected, from which search options can be specified; links below the search box on the Dicty Stock Center search page achieve the same result.