Nucleic Acids Research, 1999, Vol. 27, No. 1 79–84

Unified display of *Arabidopsis thaliana* physical maps from AtDB, the *A.thaliana* database

Seung Y. Rhee, Shuai Weng, Deverie K. Bongard-Pierce, Margarita García-Hernández, Alice Malekian, David J. Flanders and J. Michael Cherry*

Department of Genetics, Stanford University, Stanford, CA 94305-5120, USA

Received October 1, 1998; Revised and Accepted October 13, 1998

ABSTRACT

In the past several years, there has been a tremendous effort to construct physical maps and to sequence the genome of *Arabidopsis thaliana*. As a result, four of the five chromosomes are completely covered by overlapping clones except at the centromeric and nucleolus organizer regions (NOR). In addition, over 30% of the genome has been sequenced and completion is anticipated by the end of the year 2000. Despite these accomplishments, the physical maps are provided in many formats on laboratories' Web sites. These data are thus difficult to obtain in a coherent manner for researchers. To alleviate this problem, AtDB (*Arabidopsis thaliana* DataBase, URL: http://genome-www.stanford.edu/Arabidopsis/) has constructed a unified display of the physical maps where all publicly available physical-map data for all chromosomes are presented through the Web in a clickable, 'on-the-fly' graphic, created by CGI programs that directly consult our relational database.

INTRODUCTION

Arabidopsis thaliana is a small annual of the *Cruciferae* family and is well suited for molecular and genetic studies of biological processes (1,2). It has one of the smallest known genome sizes for a plant, ~120 megabases (Mb). This is projected to be sequenced completely by the end of the year 2000, thus becoming the first organism in the plant kingdom to have its genome revealed.

In parallel with this recent burst of information, it is essential to develop coherent and systematic methods of storing and retrieving the data. One example of the high information to organization ratio can be found in the physical mapping of the *Arabidopsis* genome. Completion of physical maps is crucial for understanding genome organization, systematic sequencing and positional cloning of genes. Thus, in the past several years, many groups have put tremendous efforts into constructing physical maps of different regions of *Arabidopsis* (for a review, see ref. 3). Most *Arabidopsis* physical-map data are available to researchers over the Web, from databases such as AtDB or from numerous—usually non-database—individual laboratories' Web sites. This scattering of data, combined with the difference in method, resolution, molecular probes, and clones used to construct contigs, makes it difficult for researchers to obtain, in a tractable manner, all the available physical-map information in a given region of interest.

To help alleviate this problem, AtDB has assembled currently-available physical-map data into a single map for each chromosome. This display does not attempt to resolve conflicts or ambiguities, but presents the user with the most consistent interpretation of the comprehensive clone and probe information. The unified display of physical maps has three main purposes, as follows.

(i) To provide all the available physical-map information for any given region of the genome. This display is designed to aid researchers conducting positional cloning by providing all the relevant probe and clone information in one image for their region of interest.

(ii) To provide links between genetic and physical maps in a comprehensive and unambiguous manner. Probes that have also been used as genetic markers have now been positioned physically. This aspect of the unified display is useful for researchers wishing to position new genetic loci on the physical map.

(iii) To be the first site researchers come to for physical information about the *Arabidopsis* genome. As the physical map gets more resolved and the genome sequenced, the display will show the complete sequence, its annotation, all associated clones and genetic markers.

In addition to the physical-map display, AtDB is in the process of coalescing *Arabidopsis* genomic sequences to build segments of contiguous sequence that will be placed on the physical map. Sequencing is being conducted at a rapid rate by the *Arabidopsis* Genome Initiative (AGI), which comprises six main groups: three US, one Japanese, one French and one European Union (4). AtDB's initial role in AGI was to supply a central site where users could answer questions such as; who is sequencing what, where, and how are they doing. To this end, graphical displays of AGI data were created, which have previously been described in some detail (5). In brief, the information for these data-driven displays came from a combination of direct database entry (on-line transaction processing) by AGI participants using Web forms, and from programmatic scanning of new or changed *Arabidopsis* sequences parsed from GenBank by AtDB. Recently, these

*To whom correspondence should be addressed. Tel: +1 650 723 7541; Fax: +1 650 723 7016; Email: cherry@genome.stanford.edu

displays from AtDB have been augmented by a graphical 'progress meter' of sequence in GenBank. From the sequence map that is under development, all the relevant genetic-marker and sequence-analysis information will be provided with a detailed annotation of each open reading frame. In this paper, we describe in detail the construction of the unified display of the physical maps and briefly work in progress and updates.

CONSTRUCTION OF THE UNIFIED DISPLAY OF CURRENT PHYSICAL-MAP DATA

Status of physical-map data

There has been an international effort of physical map construction and more than 90% of the *Arabidopsis* genome has been covered by overlapping clones into contigs (1). A variety of methods and resources has been used by many groups to build the physical maps (3). Accordingly, the results of the physical mapping are displayed and/or published in many divergent ways ranging from text files to spreadsheets to graphical representations using computer drawing programs (Table 1). The different methods of construction and display of the physical maps make it difficult for researchers to extract easily the status of physical-map information for a given region of the genome. In addition, these individual displays are rarely produced from a database. Thus, updates of the physical-map information have to be done manually, which discourages regular updating of data. In order to help alleviate these problems, AtDB has placed all currently available physical-map data into its database and used this information to construct a unified display of these maps.

Different types of physical-mapping methods were categorized as described in the following. In all cases, clones are defined as pieces of *Arabidopsis* genomic DNA that have been inserted in a variety of vectors. These vectors include Yeast Artificial Chromosome (YAC), Bacterial Artificial Chromosome (BAC), P1, Transformation-competent Artificial Chromosome (TAC), and cosmids. Probes are defined as pieces of DNA (clones, part of a clone, or oligomers) used to identify, in whatever way, clones containing sequence common to the probe. Markers are probes that have been placed onto a genetic map.

(i) Tiling path. This is defined as a hybridization-based physical map where the order of the probes and the hybridizing clones is known, but little or no size information is known about either the clones or the probes. The distance between each probe is unknown and they are displayed equidistantly. Examples of tiling paths include YAC and BAC contigs on chromosome I (Joe Ecker, Thomas Altmann and colleagues, unpublished data).

(ii) Fingerprinted contig. This map is constructed using fragment sizes from restriction digests of clones to calculate likely overlaps. This map rarely has probe data and has to be located onto the chromosome through common clones that have already been placed by other means, such as through additional hybridization data. Most of the fingerprinted contigs use BAC clones and are carried out by the Genome Sequencing Center (GSC) at Washington University (Marco Marra and colleagues, unpublished data).

(iii) Clone-sized tiling path. This is the same as the tiling path except the size of the individual clones in the contig is known and the overall size of the contig can be estimated by dividing the sum of all the clone sizes by the redundancy of the clones hybridizing to each probe. Examples of clone-sized tiling paths include YAC contigs on chromosomes II (6) and III (7).

Table 1. Sources of physical-map data

Chr I:
• K. Dewar, C. Kim, Y. P. Li, P. Dunn, J. Ecker; *Arabidopsis thaliana* Genome Center at University of Pennsylvania, USA
http://genome.bio.upenn.edu/ATGCUP.html
• M. Marra, M. Sekhon, R. Martienssen (Cold Spring Harbor Laboratory), R. Wilson; Genome Sequencing Center at Washington University, USA
http://genome.wustl.edu/gsc/Web_pages/projects.html#*thaliana*
• T. Mozo, S. Meier-Ewert (Berlin), T. Altmann; Max-Planck-Institut für molekulare Genetik, Golm, Germany
http://194.94.225.1/private_workgroups/pg_101/bac.html
• O. Leyser, D. Rouse, P. Mackay; University of York, UK
• W. Lukowitz, G. Jürgens; Genetic Institute, University of Munich, Germany
• C. S. Hardtke, T. Berleth; Yale University, USA
Chr II:
• E. A. Zachgo, M. L. Wang, J. Dewdney, D. Bouchez, C. Camilleri, S. Belmonte, L. Huang, M. Dolan, D. K. Bongard-Pierce, J. W. Morris, H. Goodman; Massachusetts General Hospital/Harvard University, USA
http://weeds.mgh.harvard.edu/goodman/
• R. Finkelstein, T. Lynch; U.C. Santa Barbara, USA
• S. Rounsley, X. Lin, S. Kaul, A. Glodek, L. Zhou; The Institute of Genomic Research, USA
http://www.tigr.org/tdb/at/atgenome/atgenome.html
• R. Buell, S. Somerville; Carnegie Institution of Washington, USA
• A. Sessions; U.C. Berkeley, USA
• K. Hicks, D. R. Meeks-Wagner; Vanderbilt University, USA
Chr III:
• H. Kotani, S. Sato, M. Fukami, T. Hosouchi, N. Nakazaki, S. Okumura, T. Wada, Y.-G. Liu, D. Shibata, Y. Nakamura, S. Tabata; Kazusa DNA Research Institute, Japan
http://www.kazusa.or.jp/arabi/
• C. Camilleri, J. Lafleuriel, C. Macadre, F. Varoquanuz, Y. Parmentier, G. Picard, M. Caboche, D. Bouchez; Laboratoire de Biologie Cellulaire, INRA, France
http://genome-www.stanford.edu/*Arabidopsis*/Chr3-INRA/
• K. Century, B. Staskawicz; University of California at Berkeley, USA
• T. Mozo, S. Meier-Ewert (Berlin), T. Altmann; Max-Planck-Institut für molekulare Genetik, Golm, Germany (URL see Chr I)
• H. Sakai, E. Meyerowitz; California Institute of Technology, USA
• A. Bachmair; University of Vienna, Austria
• Y. Levy, C. Dean; John Innes Centre, UK
Chr IV:
• R. Schmidt; Max-Delbrück Laboratorium in der MPG, Köln, Germany
• C. Dean, C. Lister, M. Bevan, I. Bancroft, M. Stammers; John Innes Centre, UK
• C. Schueller, K. Mayer; Munich Information Centre for Protein Sequences, Germany
http://muntjac.mips.biochem.mpg.de/arabi/index.html
• T. Arioli; Australian National University, Australia
• L. Parnell, R. Martienssen, D. McCombie; Cold Spring Harbor Laboratory, USA
http://nucleus.cshl.org/protarab/
• W. Soppe; Wageningen Agricultural University, The Netherlands
Chr V:
• H. Kotani, S. Sato, M. Fukami, T. Hosouchi, N. Nakazaki, S. Okumura, T. Wada, Y.-G. Liu, D. Shibata, Y. Nakamura, S. Tabata; Kazusa DNA Research Institute, Japan (URL see Chr III)
• L. Parnell, R. Martienssen, D. McCombie; Cold Spring Harbor Laboratory, USA (URL see Chr IV)
• R. Chapple; CSIRO, Canberra, Australia
• R. Schmidt; Max-Delbrück, Köln, Germany
• C. Dean, C. Lister, M. Stammers; John Innes Centre, UK
• M. Anderson, S. Walsh; Nottingham *Arabidopsis* Stock Centre, UK
All Chromosomes:
• The members of the *Arabidopsis* Genome Initiative.

(iv) Physical map. The sizes of the clones, most of the probes, and the distance between probes are known or can be inferred. This category includes most of the smaller-scale chromosomal walks performed for positional cloning of genes. Examples of physical maps include YAC contigs on chromosome IV (8), YAC, P1 and TAC contigs on chromosomes III and V (9,10), and a BAC contig on chromosome II (11).

(v) Sequence contigs. This is the definitive and ultimate physical map, where the sequence is known for a chromosomal region. Examples of sequence contigs include ESSA BAC contigs on chromosome IV (12) and a P1 contig on chromosome V (13).

At the time of writing (September 1998), all chromosomes with the exception of chromosome I are completely covered by overlapping contigs except at the centromeric and nucleolus organizer (NOR) regions. Chromosome I is covered by 31 YAC contigs (Pat Dunn, personal communication). Four BAC contigs (Thomas Altmann and Marco Marra, personal communication) overlap some of the YAC contigs. In total, there are 23 gaps, including the centromeric region, for chromosome I. Chromosome II contains ~3.5 Mb of NOR containing the rDNA repeats (3). The remainder of the chromosome is covered by four YAC contigs (6). Recently two of these contigs were closed by a YAC walk (Ruth Finkelstein, personal communication) and a BAC contig (11). In addition, there are 15 BAC contigs that are being sequenced (Steve Rounsley, unpublished data). Chromosome III is covered by 10 YAC contigs (7). All of the gaps except at the centromeric region have been closed by P1, TAC and BAC clones (7,9). Chromosome IV contains ~3.5 Mb of NOR and is covered by four YAC and cosmid contigs (8). There are eight BAC contigs, two of which close two of the YAC gaps (Mike Bevan and Larry Parnell, personal communication). Chromosome V is covered by 31 YAC contigs (14). All of these contigs, except at the centromere, are closed by two contigs comprising YAC, P1, TAC and P1 clones (10).

Collection of physical-map data

In order to facilitate the retrieval of all the different physical-map information and to link the genetic and physical maps, all available physical-mapping data were collected and a unified display of these data was developed. Sources of the physical-mapping data include published papers and Web sites (Table 1). In many cases, researchers were kind enough to provide unpublished data to aid with these efforts. The strategy was to collect initially the data from large-scale mapping projects and then to incorporate data from many smaller-scale endeavors, e.g., chromosome walks to positionally clone genes. Particular efforts were made to find information that bridged gaps between adjacent contigs. For example, some of the BAC contigs for chromosome I extended or bridged a few YAC tiling path contigs (Thomas Altmann and Joe Ecker, unpublished data). Fingerprinted BAC contigs from GSC (Marco Marra and colleagues, unpublished data) were searched for the presence of any BAC clones present in the BAC contigs that extended the YAC contigs to close gaps. By this approach, the number of contigs on chromosome I was reduced from 31 to 24.

As most of the map data were represented in fixed graphic images or in a form that could not be queried, the coordinates of clones and probes were measured directly from the graphics of physical-map contigs, fingerprinted contigs and tiling paths, and entered into the database. For the clone-sized tiling-path contigs, the contig size was estimated by adding all the clones and dividing the sum by the average number of redundant hits (number of clones/probe) and placing each probe equidistant as described elsewhere (7). For BAC hybridization tiling paths that have been constructed using the webprobeorder program (15), the contig size was estimated by the same calculation used for the clone-sized tiling paths. For BAC clones whose sizes are unknown, 100 kb was used as an average size, which is an average insert size for sequenced BAC clones (data not shown). The positions of the overlapping clones were estimated by converting the webprobeorder display into kilobases using a program written in Perl at AtDB. In the webprobeorder display, each probe and clone are displayed in each column and row, respectively. This display places a number representing the strength of hybridization (e.g., 1, 2, 3) where each probe hybridizes to the clone (See URL: http://194.94.225.1/private_workgroups/pg_101/about_igfbac_con.html). This Perl program converts the column (probe) positions into numbers and records the first and last hybridization positions. These two numbers are then converted into kilobases calculated by the contig-size estimation described above.

Display of the physical-map data

The unified display of the physical maps is clickable as are other AtDB displays (5). By this method, new mapping information can be incorporated into the display simply by updating the database. The picture presented to the user will then be altered as soon as new data are entered into the database.

The unified display of the physical map has three main graphics. The user is first presented with a cartoon of the five chromosomes of *Arabidopsis* (Table 2, Physical Maps). Each chromosome is presented as a horizontal bar, with contigs shown below as narrow bars color-coded according to contig type. For example, chromosome III has two physical-map contigs comprising YAC, TAC, P1 and BAC clones (Fig. 1, light blue bars). There are 10 YAC clone-sized tiling paths (Fig. 1, purple bars), four BAC clone-sized tiling paths (dark blue bars), and two YAC walks (yellow and green bars). A selection of probes used in more than one contig type is shown along the chromosome to provide landmarks. Genetic markers are shown in red (Fig. 1). A Javascript box gives a contig's name, source and chromosomal coordinates when the mouse pointer is placed over the bar representing it (not shown).

Clicking on the chromosomal bar or a contig takes the user to a second display, which shows the relevant region in much more detail (Fig. 2). For each chromosome, a set of contigs that covers the chromosome most extensively using a single method of physical mapping was chosen as the 'framework' set of contigs. Measurements of distances of probes and clones from these contigs were used to provide the chromosomal coordinates. The size of the gaps between the framework contigs was estimated by using the average genetic/physical distance (cM/kb) ratio for the chromosome and determining the estimated physical distance between markers nearest to the proximal ends of adjacent contigs. Any known distance between the end of the contig to the nearest end marker was subtracted from the estimated gap size. If the framework contigs were bridged by other contigs, the gap size was determined by the bridging contigs. Other contigs were then placed on the display relative to the framework by placing these

Table 2. Useful URLs for the *A.thaliana* database

Home Page	http://genome-www.stanford.edu/Arabidopsis/
Maps	http://genome-www.stanford.edu/Arabidopsis/maps.html
Physical Maps	http://genome-www4.stanford.edu:8400/cgi-bin/AtDB/Pchrom
Arabidopsis Genomic View (Java)	http://genome-www3.stanford.edu/Arabidopsis/chromosomes/
AGI Sequencing Totals	http://genome-www3.stanford.edu/cgi-bin/Webdriver?MIval=atdb_agi_total
AGI's Web Sites	http://genome-www.stanford.edu/Arabidopsis/AGI/AGI_links.html
BLAST	http://genome-www2.stanford.edu/cgi-bin/AtDB/nph-blast2atdb
FASTA	http://genome-www2.stanford.edu/cgi-bin/AtDB/nph-fastaatdb
Pattern Matching	http://genome-www2.stanford.edu/cgi-bin/AtDB/PATMATCH/nph-patmatch
Restriction Analysis	http://genome-www2.stanford.edu/cgi-bin/AtDB/PATMATCH/RestrictionMapper
ABRC cDNA Clones	http://genome-www.stanford.edu/Arabidopsis/ABRC/index.html
Arabidopsis Gene Hunter	http://genome-www.stanford.edu/cgi-bin/AtDB/geneform
GenBank	http://www.ncbi.nlm.nih.gov/Web/Search/index.html

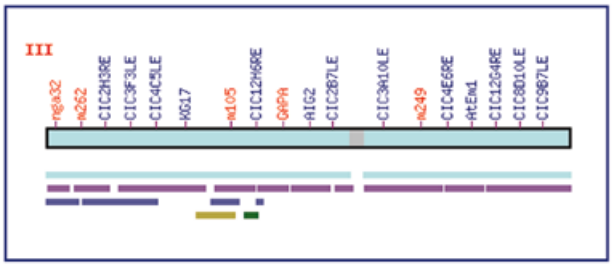

Figure 1. First graphic page of the chromosome III physical-map display. The thick bar represents chromosome III distinguished by regions covered by contigs (light blue) and a gap at the centromeric region (gray). A selection of genetic markers (red) and non-genetic probes (blue) are drawn with tick marks on top of the chromosome. Below the chromosome, contigs are represented as thin bars. For Chr III, there are five types of contigs: physical map from Kazusa (light blue); YAC clone-sized tiling paths from INRA (purple); BAC clone-sized tiling paths from Altmann (IGF) (dark blue); and YAC walks from Century (yellow) and Sakai (green). All of the contigs and the chromosome are clickable and clicking on any region will take the user to the second graphic page.

contigs centrally relative to the leftmost- and rightmost-markers that are common to the framework (Fig. 2). If only one marker was common, the contig was positioned relative to the framework using that probe. Contigs that did not share a probe with the framework, but did share common probes with other contigs that have been positioned relative to the framework, were placed relative to those contigs. There are contigs that do not share probes with any other contigs, but whose probes are present as clones in the framework. These contigs were aligned to the framework by these clones.

The second graphic has several features. First, the relevant chromosomal bar is presented along the top of the display, with a red box highlighting the area displayed (Fig. 2). Clicking anywhere on the bar changes the display to that chromosomal location. A zoom in/out button toggles the region displayed between 2000 (the default) and 500 kb. In the vertical 'layer' below the chromosomal bar, the framework contigs are displayed with the chromosomal coordinates and probe names. In successive layers below the framework layer are the other, non-framework, contigs. Each contig has its own, internal coordinates displayed and is linked to the framework and to other contigs in the same region by vertical, dotted lines that connect each common probe. Each layer has a different background color and the name of the individual/lab who produced it is shown at the left. Hybridizations are indicated by purple dots on the clones. These are positioned vertically below the tick mark of the hybridizing probe. To aid in identifying individual hybridizations, a Javascript box duplicated above and below the graphic presents all hybridizing probes or clones whenever the mouse pointer is placed over a clone or probe.

This second display represents comprehensive physical-map data, and, as such, many regions are inundated with multiple contigs that may be more confusing than useful. For this reason, a feature is provided that allows the user to choose the display of any number of contigs. Below the contigs, the user is presented with a box that can be clicked to either show or hide any contig. The user therefore has the option of displaying the most relevant contigs for his or her particular interest.

All clones and probes in each contig are clickable and spawn a new page to the appropriate 'clone' entry in AtDB's database. This includes details of hybridizing clones and probes and the source and type of the hybridization information, clone library source, probe source and type, and links to AtDB's genetic map displays. If a clone is undergoing AGI sequencing, its status is provided with links to: AtDB's AGI graphical display, the relevant annotation page of the sequencing groups and GenBank. There is also a link back to the appropriate physical-map page from any of the three graphical displays.

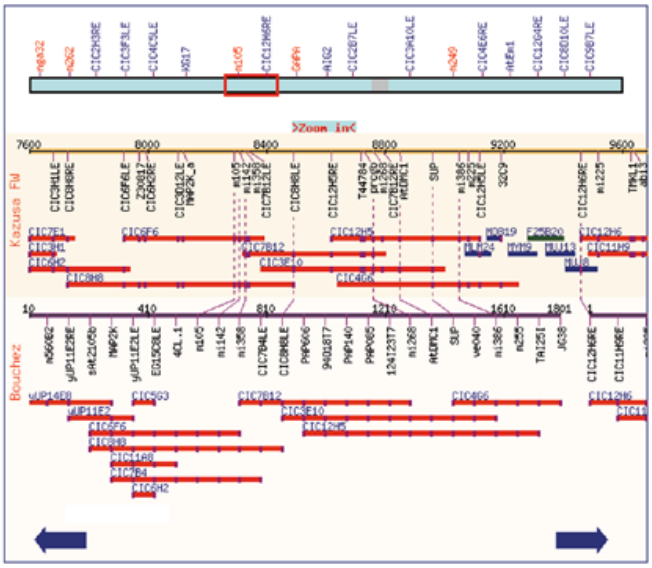

Figure 2. A representative second graphic page of the physical-map display. A region of chromosome III between 7800 and 9800 kb is shown in detail with the same chromosomal picture as the first graphic page shown at the top. A red box on the chromosome shows the area displayed in more detail below. There are five contig types in this region, but only the Kazusa framework and Bouchez (INRA) contigs are shown. Each contig type is presented as a thin bar with kb coordinates. The Kazusa framework contig (yellow background) shows chromosomal coordinates and probes along the tick marks of the contigs. Bouchez (INRA) contigs have internal kb coordinates and probes along the tick marks. Below the contigs are clones: YAC (red), BAC (green) and P1 (blue) clones. Hybridizations are indicated on the clones as purple tick marks at the positions of the hybridizing probes. Bouchez (INRA) contigs are positioned on the physical map using probes that also occur in the framework. Purple dash lines indicate the common probes. Arrows at the bottom of the page take the reader to the adjacent chromosomal coordinates. Clicking on probes or clones will take the user to the next Web page containing the AtDB clone entry information.

CONSTRUCTION OF SEQUENCE CONTIGS DISPLAY

At the time of writing, 37.1 Mb of AGI sequence is in GenBank, ~29% of the expected total of 120 Mb (Table 2, AGI Sequence Totals). This comprises: 5.8 Mb on chromosome I, 8.9 Mb on II, 0.2 Mb on III, 10.1 Mb on IV and 9.6 Mb on V (some sequence is from clones with an as yet unknown location). In addition, there is ~6.2 Mb of 'annotated' *Arabidopsis* sequence that originates from non-systematic sequencing (Table 2, GenBank).

With the sequencing project well underway, issues of presenting and retrieving the sequence information have arisen. At present, there is no one Web site where all the genomic sequences are linked into contigs and placed on the physical map. An additional need for a single site displaying sequence contigs arises from the relatively simple level of annotation permitted in GenBank entries. This is inadequate for many users who wish to know, for example, which gene-prediction or splice-site programs were used. Extensive annotation pages about individual clones are often provided on AGI participants' Web sites (Table 2, AGI's Web sites). Although the user's task has been made simpler by direct linking from the relevant clone entry in AtDB's database, the requirement to search several sites, with different data access and displays, is problematical. Furthermore, as sequencing from different groups on the same chromosome reaches each other, there will arise a need for annotation that passes over the boundary between the groups' efforts.

AtDB has thus embarked on a project to construct sequence contigs, place them on the unified physical map and, eventually, to provide graphical annotation. Construction of sequence contigs and their placement on the physical map are under development by a combination of manual curation and automation. Sequence contigs will be placed onto the physical-map display using markers or clones contained within them.

SEQUENCE ANALYSIS RESOURCES

BLAST and FASTA

AtDB's BLAST and FASTA programs have been previously described (5). The thrice-weekly-updated nucleic acid and protein *Arabidopsis*-data sets used by these programs have been expanded and now include: (i) all public *Arabidopsis* sequences (GenBank and more, including Cold Spring Harbor Laboratory's data set of repetitive sequences); (ii) an additional data set containing all DNA sequences from higher plants contained within GenBank; and (iii) the Mendel database-curated set of sequenced plant proteins (URL: http://jiio6.jic.bbsrc.ac.uk:80/index.html). This latter database, based at the John Innes Centre in Norwich, UK, is mirrored by AtDB in order to provide faster access for North American and other users.

Restriction Analysis and Pattern Matching

Restriction Analysis and Pattern Matching are CGI programs, which search the same data sets used by AtDB's BLAST and FASTA programs and use a Perl script for displaying results (Table 2). Restriction Analysis can be performed on either *Arabidopsis* sequences (using GenBank Locus or Accession Numbers) or on any sequence the user enters into the form. The user can choose to use all known enzymes, or select the enzymes by their cutting character (six-base; once or twice in the sequence) or by the type of end produced (blunt; 3′ or 5′ overhang). Results are presented graphically with the site of each cut for each enzyme shown. Clicking on the enzyme's name gives a list of the fragment sizes. The enzymes that do not cut the sequence are also listed.

Pattern Matching allows searching for short and ambiguous nucleotide or peptide patterns (sometimes called motifs). For DNA searches, both strands, the given strand or the reverse complement may be searched. Simple matching queries may be performed, although users are given many examples to help them on their way to use regular expressions as search terms. For example, X{66,76}C[AVLI][AVLI]X> finds protein subunits whose size is 70–80 amino acids (66+4, 76+4) and C-terminal sequence (>) is cysteine, then two aliphatic amino acids, followed by any amino acid. Pattern Matching also allows the use of up to three mismatches, insertions or deletions in the search string. Results show the position of the matching pattern in target sequences and provide hot links to their GenBank entries.

An additional sequence-related resource from AtDB is provided in collaboration with the *Arabidopsis* Biological Resource Center (ABRC) in Ohio, who provide researchers with *Arabidopsis* EST clones. AtDB's 'ABRC cDNA Clones' (Table 2) allows users to search a table of all available ESTs from ABRC, the results of which include links to: pre-computed BLAST results for each EST against the nucleotide nr (non-redundant) data set

from GenBank; the clone's GenBank, University of Minnesota, and TIGR-EST database entries; and the ABRC Web-form for ordering that clone.

OTHER DEVELOPMENTS AND WORK IN PROGRESS FROM AtDB

In addition to the unified display of the physical map and sequence analysis resources, there have been several updates and changes implemented by AtDB in the past year. Among them is the development of the 'Genomic View', a graphic display of the five *Arabidopsis* chromosomes, with a few clones or markers serving as landmarks (Table 2). From this display, one can get to the Physical Map, Genetic and Physical Map, AGI Map, Meinke Classical Map, Lister & Dean RI Map, and mi RFLP-Marker Map for any selected location within the *Arabidopsis* genome.

Also, AtDB's Web pages were re-organized to allow an easier access to the different kinds of information presented on the Web site. The new layout consists of several main pages organized according to a given category of information (e.g., Search Options, Illustra database, Sequence Analysis Tools, AGI, Maps, Stock Centers, etc.). To allow easier navigation within the Web site, these categories are displayed as a clickable menu on the left side of the main pages. In addition, every page has links to the Illustra database, Search, BLAST, FASTA, Pattern Matching and Restriction Analysis tools.

A recent development in AtDB is the Web-search tool '*Arabidopsis* Gene Hunter' (Table 2). This feature allows users, with a single click, to find information about a given *Arabidopsis* gene from a selectable list of several *Arabidopsis*-related Web sites and databases. The result is a concatenation of clickable Web pages from the sites searched.

Currently in progress is an update of a crucial set of information of any genomic database, the 'Locus' (any mappable entity) data set. We are collecting and curating all available information (sequence ID, mutant phenotype, map location, gene function, publications, expression, etc.) on each known *Arabidopsis* locus. Starting with genetic map loci, subsets of the new data will be made available as improvements are made.

Users of the AtDB are asked to reference this paper when information from the database has provided information for their research.

ACKNOWLEDGEMENTS

The authors thank the numerous researchers who generously provided unpublished data to AtDB, particularly to T. Altmann, T. Arioli, A. Bachmair, M. Bevan, D. Bouchez, K. Century, R. Chapple, C. Dean, M. Dhan, P. Dunn, R. Finkelstein, C. Hardtke, K. Hicks, Y. Levy, O. Leyser, X. Lin, M. Marra, R. Martienssen, T. Mozo, L. Parnell, S. Rounsley, H. Sakai, R. Schmidt, M. Sekhon, W. Soppe and S. Tabata; R. Buell, W. Lukowitz and I. Wilson for helpful discussions; Gavin Sherlock for reading the manuscript; and Gail Juvik, Mark Schroeder and Yan Zhu for software. Funding for AtDB is provided by the National Science Foundation, grant numbers DBI-9503776 and DBI-9714160.

REFERENCES

1 Meinke,D.W., Cherry,J.M., Dean,C., Rounsley,S.D. and Koornneef,M. (1998) *Science*, **282**, 662–682.
2 Meyerowitz,E.M. and Somerville,C.R. (1994) *Arabidopsis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3 Schmidt,R. (1998) In Anderson,M. and Roberts,J.A. (eds), *Arabidopsis Annual Plant Reviews*. CRC Press, Sheffield, Vol. I, pp. 1–23.
4 Stiekema,W.J. and Pereira,A. (1988) In Anderson,M. and Roberts,J.A. (eds), *Arabidopsis Annual Plant Reviews*. CRC Press, Sheffield, Vol. I, pp. 31–57.
5 Flanders,D.J., Weng,S., Petel,F.X. and Cherry,J.M. (1998) *Nucleic Acids Res*., **26**, 80–84.
6 Zachgo,E.A., Wang,M.L., Dewdney,J., Bouchez,D., Camilleri,C., Belmonte,S., Huang,L., Dolan,M. and Goodman,H.M. (1996) *Genome Res*., **6**, 19–25.
7 Camilleri,C., Lafleuriel,J., Macadre,C., Varoquaux,F., Parmentier,Y., Picard,G., Caboche,M. and Bouchez,D. (1998) *Plant J*., **14**, 633–642.
8 Schmidt,R., West,J., Cnops,G., Love,K., Balestrazzi,A. and Dean,C. (1996) *Plant J*., **9**, 755–765.
9 Sato,S., Kotani,H., Hayashi,R., Liu,Y.-G., Shibata,D. and Tabata,S. (1998) *DNA Res*., **5**, 163–168.
10 Kotani,H., Sato,S., Fukami,M., Hosouchi,T., Nakazaki,N., Okumura,S., Wada,T., Liu,Y-G., Shibata,D. and Tabata,S. (1997) *DNA Res*., **4**, 371–378.
11 Wang,M.L., Huang,L., Bongard-Pierce,D.K., Belmonte,S., Zachgo,E.A., Morris,J.W., Dolan,M. and Goodman,H.M. (1997) *Plant J*., **12**, 711–730.
12 Bevan,M., Bancroft,I., Bent,E., Love,K., Goodman,H., Dean,C., Bergkamp,R., Dirkse,W., Vanstaveren,M., Stiekema,W. *et al.* (1998) *Nature*, **391**, 485–488.
13 Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M., Miyajima,N. and Tabata,S. (1998) *DNA Res*., **4**, 215–230.
14 Schmidt,R., Love,K., West,J., Lenehan,Z. and Dean,C. (1997) *Plant J*., **11**, 563–72.
15 Mott,R., Grigoriev,A., Maier,E., Hoheisel,J. and Lehrach,H. (1993) *Nucleic Acids Res*., **21**, 1965–1974.