FPC Web Tools for Rice, Maize, and Distribution[1]

Vishal Pampanwar, Friedrich Engler, James Hatfield, Steve Blundy, Gaurav Gupta, and Carol Soderlund*

Arizona Genomic Computational Laboratory, BIO5 Institute, University of Arizona, Tucson, Arizona 85721 (V.P., F.E., J.H., G.G., C.S.); and Clemson University, Clemson, South Carolina 29634 (S.B.)

Many clone-based physical maps have been built with the FingerPrinted Contig (FPC) software, which is written in C and runs locally for fast and flexible analysis. If the maps were viewable only from FPC, they would not be as useful to the whole community since FPC must be installed on the user machine and the database downloaded. Hence, we have created a set of Web tools so users can easily view the FPC data and perform salient queries with standard browsers. This set includes the following four programs: WebFPC, a view of the contigs; WebChrom, the location of the contigs and genetic markers along the chromosome; WebBSS, locating user-supplied sequence on the map; and WebFCmp, comparing fingerprints. For additional FPC support, we have developed an FPC module for BioPerl and an FPC browser using the Generic Model Organism Project (GMOD) genome browser (GBrowse), where the FPC BioPerl module generates the data files for input into GBrowse. This provides an alternative to the WebChrom/WebFPC view. These tools are available to download along with documentation. The tools have been implemented for both the rice (*Oryza sativa*) and maize (*Zea mays*) FPC maps, which both contain the locations of clones, markers, genetic markers, and sequenced clone (along with links to sites that contain additional information).

FingerPrinted Contigs (FPC) is a program that orders clones into contigs based on restriction fragment fingerprints and marker data and orders contigs based on genetic markers. FPC provides the ability to assemble and manually edit contigs (Soderlund et al., 1997), integrate markers and frameworks (Soderlund et al., 2000), add electronic markers, and automatically select a minimal tiling path (Engler et al., 2003). In addition to many other whole genome physical maps, FPC has been used to build the physical maps of rice (*Oryza sativa*; Chen et al., 2001) and maize (*Zea mays*; Coe et al., 2002). These maps integrate clones, anchored markers, unanchored markers, genetic markers, and sequenced clones. For these maps to be used by the community, they have been made available for viewing using the FPC Web tools at www.genome.arizona.edu (the pages referred to in this paper at this URL were created jointly by Arizona Genomics Computational Laboratory [AGCoL] and Arizona Genomics Institute [AGI]).

The WebAGCoL package is a set of four tools: WebFPC displays contigs in a view very similar to the FPC display. WebChrom shows contigs and genetic markers aligned to the chromosome. It also allows the user to view the distribution of markers based on name or remark. WebFCmp allows fingerprint comparisons of a user-selected clone set against the entire FPC database. WebBSS locates a user supplied sequence on an FPC map based on its similarity to sequences associated with other clones in the map. All of these tools work with a standard browser.

The WebAGCoL package has been made available for distribution. Such a package can be difficult to install since it has Java, CGI, and HTML files that all belong in different directories. To simplify the setup, we have written a script that automatically installs the different files based on a configuration file. The set of tools and setup scripts were released in August 2004.

This manuscript also discusses two other FPC support efforts: an FPC module for BioPerl (www.bioperl.org) that provides a simple interface to query an FPC database and an FPC browser using the Generic Genome Browser software (Stein et al., 2002). The BioPerl FPC module and Genome Browser files were released in August 2003.

A previous version of WebFPC and WebChrom was released in August 2003 and is used at multiple sites. Some sites have developed their own FPC Web browsers, for example, ICE (Internet Contig Explorer; Fjell et al., 2003). The advantages of having one distributable set of FPC Web tools are: (1) everyone is not reinventing the same functionality, (2) the community can view FPC results at different sites and have the same look and feel, and (3) as changes are made to FPC, they can be easily included in the Web tools.

It should be noted that these tools were not designed to replace FPC. If a user will be executing more than a few occasional queries, then downloading the FPC database and FPC executable will save time. There are FPC executables for Sun, Linux, and Mac OSX, but not for Windows. For Windows users, we recommend using a VNC (http://www.realvnc.com) session on a Unix machine, which allows the user to run FPC on Unix from a Windows machine. A tutorial on building maps is given in Engler and Soderlund (2002), and a set

[1] This work was supported in part by the U.S. Department of Agriculture Initiative for Future Agriculture and Food Systems (grant no. 11180) and by the National Science Foundation (grant no. 0213764).

* Corresponding author; e-mail cari@agcol.arizona.edu; fax 5206262632.

www.plantphysiol.org/cgi/doi/10.1104/pp.104.056291.

of tutorials are available on specialized features. The FPC site (www.agcol.arizona.edu/software/fpc) provides the following: (1) the release of FPC; (2) access to the tutorials; (3) a list of sites that use WebFPC or have FPC Web browsers; and (4) the software discussed in this manuscript.

RESULTS

Rice Map

The rice FPC map has 72,703 clones, 8,870 markers, 180 contigs, and 2,918 anchors. FPC calls any marker that has a location on a chromosome or linkage group an anchor. There are two types of anchors: (1) frameworks are well ordered and (2) placements are binned between frameworks. For the rice map, the 1,378 frameworks are the Japanese Genetic markers (Harushima et al., 1998). The 1,040 placement markers have the prefix OJ and originate from Monsanto-International Rice Genome Sequencing Project Map Integration (Chen et al., 2001). Chen et al. (2001) reported 458 contigs, which have since been reduced to 180 (H.R. Kim, personal communication). Of the 180 contigs, 93 are anchored to chromosomes by the genetic markers. Of the remaining contigs, 49 have only two clones and 26 have less than 10 clones; these contigs probably contain clones with bad fingerprints and are generally ignored. By these criteria, only 12 good contigs (i.e. with $\geq$10 clones) remain unanchored.

The initial WebFPC display for the rice physical map is shown in Figure 1. It provides options to search by clone, marker, or contig. If a marker is contained in more than one contig, all contigs containing that marker will be listed. Substrings can be used for marker and clone names. For example, to view all the sequenced clones, one would enter the string ''sd1'' (the ''sd'' stands for simulated digest, as explained below).

WebFPC is implemented in Java, which allows fast navigation around an entire contig, avoiding the slow redisplay common in Web displays using the paging method. Each marker is centered over the largest stack of clones to which it is attached. WebFPC features a filtering window that gives user options to show or hide information. For example, there may be many markers in a small region, causing the marker track to become very deep. Marker filtering allows the user to limit the depth by showing only the markers of interest, as shown in Figure 2a. In Figure 2b, markers prefixed by SOG or OJ have been removed from the display. Anchors are shown at the bottom of the display (Fig. 2c). By default, only the frameworks are shown, but the placements can be made visible via the filtering window. While only a region of the contig may be visible, all frameworks are shown, providing an overview of the whole contig. Selecting an anchor centers the contig display on the corresponding region. A pull-down at the top of the display lets the user filter the clones by ''No Buried,'' ''All,'' or ''Seq Only.'' A buried clone is one whose fingerprint pattern matches that of another clone either exactly or approximately. Selecting No Buried hides the buried clones to limit redundant information. The Seq Only option shows only the simulated digest (SD) clones, which are generated from sequenced clones.

The SD clones are generated by a nightly cronjob (a script that is scheduled to run automatically at a given time). The cronjob executes the following steps: (1) download updated rice sequence from GenBank; (2) run a simulated digest on each sequence (Engler et al., 2003) to create an in silico fingerprint; (3) if a sequence digests into more than 55 bands, breaks it up into

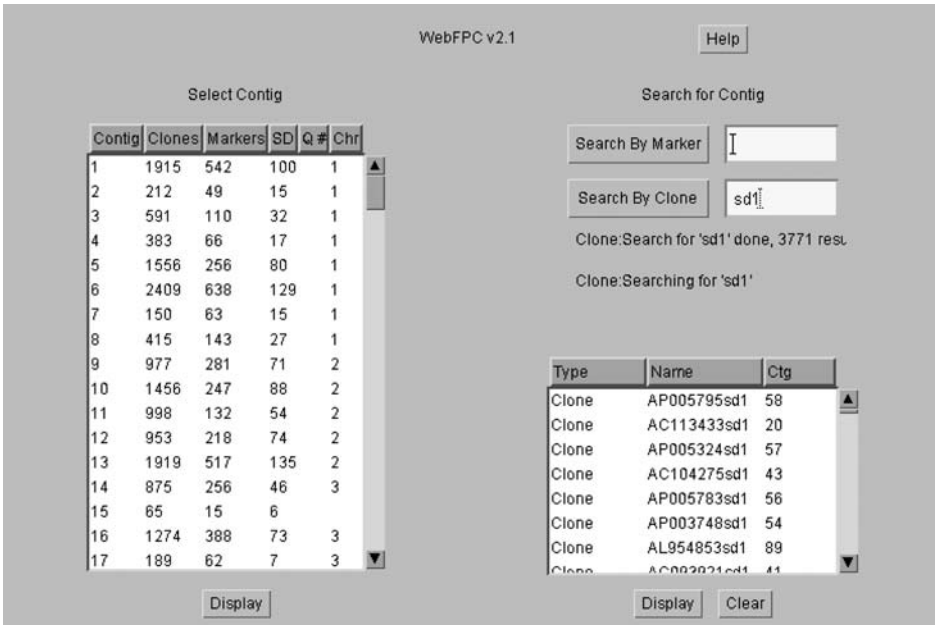

Figure 1. WebFPC initial display. The contig to display can be selected either from a table or by searching for a marker or clone. In this example, the substring sd1 was entered for clones. All clone names containing the substring sd1 are listed in the bottom right scrollable window. Any clone can be clicked from this window and the contig will be displayed centered on the clone, which will be highlighted.

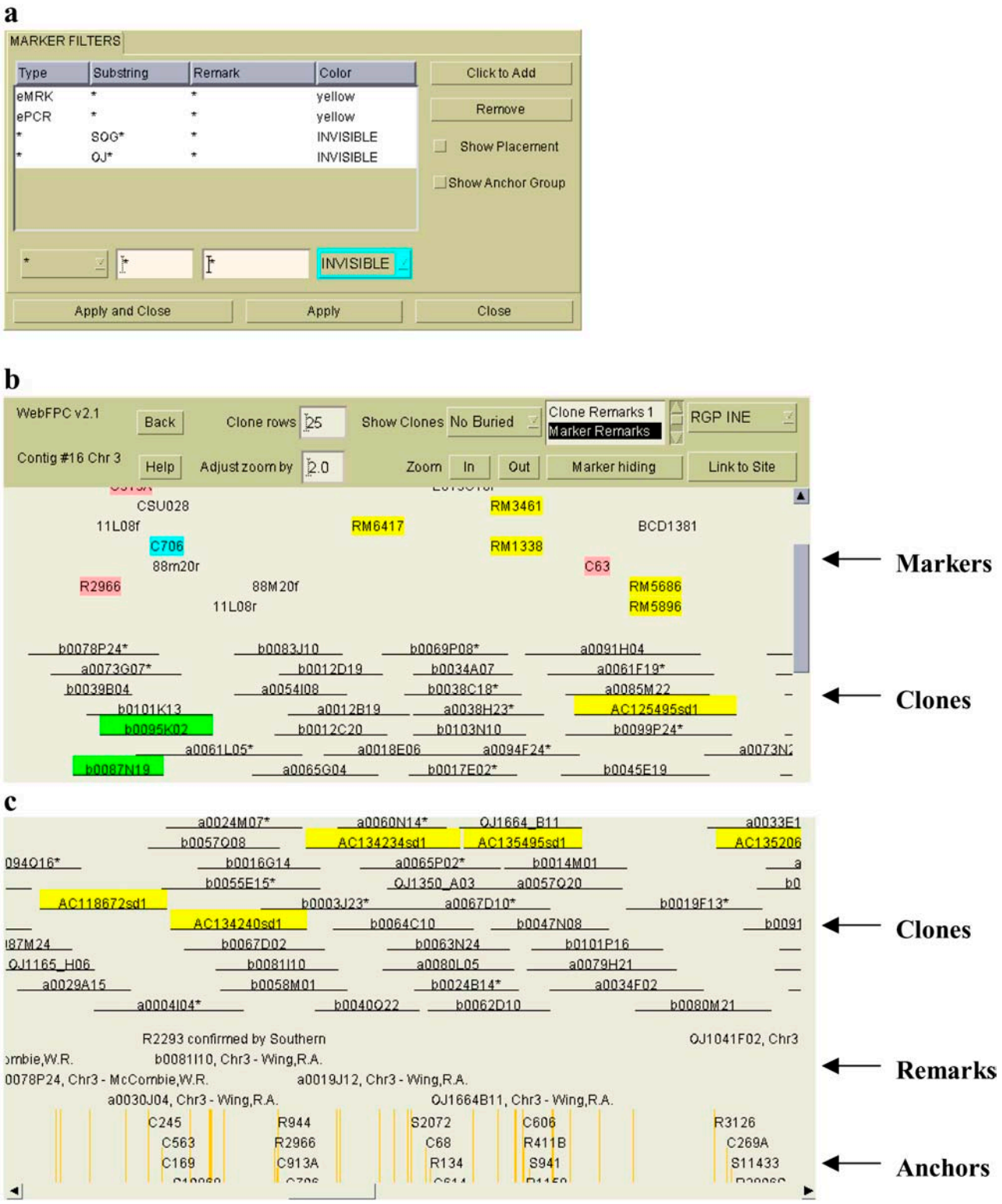

Figure 2. WebFPC contig. a, For the rice FPC, there are so many markers that the marker track takes up half the window, so filtering is helpful using this window. This example shows that ePCR and eMRK (electronic markers) are colored yellow, and markers with the prefix of SOG or OJ are made invisible in the display show in b. b, The WebFPC contig display closely resembles that of FPC, showing an ordering of clones along with markers centered over their attached clones. By selecting the RGP INE database (top right corner), all entities that have links to that database our highlighted in pink. c, The bottom half of the WebFPC display shows the remarks and the anchors.

multiple overlapping clones, where they are consecutively suffixed by sd1, sd2, etc. (this is done because a clone with many more bands than the typical clone will have a low probability of overlap with the other clones); (4) add the fingerprints to the FPC database; (5) compute the overlap probability between each SD clone and all other clones in the database; (6) automatically place each clone in the same position as its best match; (7) generate a file of links between FPC SD clones and GenBank sequences; and (8) update the WebFPC site. The number of SD clones are shown on the initial display (Fig. 1), and they are also colored yellow in the contig display (Fig. 2b) and remarked (Fig. 2c). The remark is the original clone name, the author, and the chromosome assignment. Since this information is automatically parsed from the GenBank record, it may not be correct if it was not entered into GenBank in the canonical format. But the automatic entry means that WebFPC is always up to date with the latest sequenced clones. The rice sequencing status page is also updated nightly and shows the coincidence score between the original fingerprinted clone and the SD clone; this provides both clone and sequence assembly confirmation (see www.genome.arizona.edu/shotgun/rice/status).

In the top right hand corner of the contig display, there is a pull-down that lists other on-line databases to which the user can link. For the rice map, there are links to GenBank (Benson et al., 2004), Gramene (Ware et al., 2002), and RGP INE (Harushima et al., 1998). Since RGP INE is selected in Figure 2, all markers and clones with links to records in RGP INE are highlighted. Selecting a marker and then selecting ''Link to Site'' will bring up the corresponding record in the given on-line database.

Figure 3 shows the rice WebChrom display, which provides a view of the ordering of the contigs and anchors along the chromosome. Selecting a contig will bring up the WebFPC display for the contig. The markers link to WebFPC, Gramene, and RGP INE. It is not unusual for the genetic location of anchors to disagree with their location in a contig. For example, contig 14 in Figure 3 shows all of its markers in close succession (bounded box with tick marks representing genetic markers), but the long unbounded yellow box indicates that one marker is on the lower part of chromosome 3. The contig's chromosome and the position on the chromosome are calculated by FPC (Engler et al., 2003). A contig is assigned to a chromosome based on a majority rule of all the anchors in the contig. The contig's position is calculated from all anchors located on the assigned chromosome, excluding anchors that are too far from the average location.

If the user wants to search for a particular marker on the chromosome map or see the distribution of a set of markers based on a substring of the marker name or marker remark, he or she can use the WebChrom Search tool. Many markers are from expressed sequence tags, making it advantageous to remark them with their annotation. As a test case, we blasted (Altschul et al., 1997) all the markers against SwissProt (Boeckmann et al., 2003) and added the highest scoring hit as a remark for each marker. Since the remark contains the word SwissProt, the user can search on it to view the distribution of annotated entries (see Fig. 4).

Though WebFPC shows overlapping clones, the amount of overlap is not exact due to the error in the data (Soderlund et al., 2000). Hence, if a user wants to know which clones strongly overlap with a given clone, he or she can do so by using the WebFCmp tool shown in Figure 5. The user inputs one or more clones and a cutoff. An input clone is compared with each clone in the FPC database by first counting the number of bands N that have the same value within a tolerance, then computing the probability score that the N bands are shared by coincidence (Sulston et al., 1988; Soderlund et al., 1997). If two clones have a score below the cutoff they are said to overlap. WebFCmp outputs all overlapping clones with links to the corresponding contigs in WebFPC.

Suppose one wants to determine whether a sequence from a related organism is found in rice and, if so, what other markers are surrounding the given marker. Such information can be gained by querying all sequences associated with FPC clones, namely the sequences for all SD clones and/or the bacterial artificial chromosome end sequence (BES) for all fingerprinted clones. FPC has a feature called Blast Some Sequence (BSS) that blasts a file of sequences against a directory of BESs or genomic sequences and creates a report of all hits and their location on the FPC contigs. For the WebBSS, the user inputs a sequence, the FPC BSS routine is called, and the output is parsed and displayed as shown in Figure 6. Selecting a contig will bring it up in WebFPC, where the user can view the surrounding markers and clones. As of February 7, 2005, there are 4,058 genomic sequences and 98,286 BESs.

For an alternative view to WebFPC and WebChrom, we have developed an FPC configuration file and GFF file description that are used to create a Generic Model Organism Project (GMOD) genome browser (Stein et al., 2002). This allows the user to view all the contigs and frameworks for a chromosome in a bird-eye's view. Selecting a region of genetic marker displays the corresponding region of the contig in the bottom section. Browsing can be done by chromosome or by contig. In chromosome browsing, contigs are aligned to a chromosome based on the position calculated by FPC. The length of a contig is the number of CBunits (i.e. approximately the number of consensus bands in that contig). The approximate length in base pairs is the number of CBunits times the average band length (typically around 4,096 bp for agarose). The length of the chromosome represents the total length of all its contigs with 100 bp between contigs. Features like markers, frameworks, clones, and sequenced SD clones are shown in different tracks (see Fig. 7).

A full description of the origin of the clones and markers in the rice FPC is presented at the Web site (www.genome.arizona.edu/fpc/rice). The FPC file is

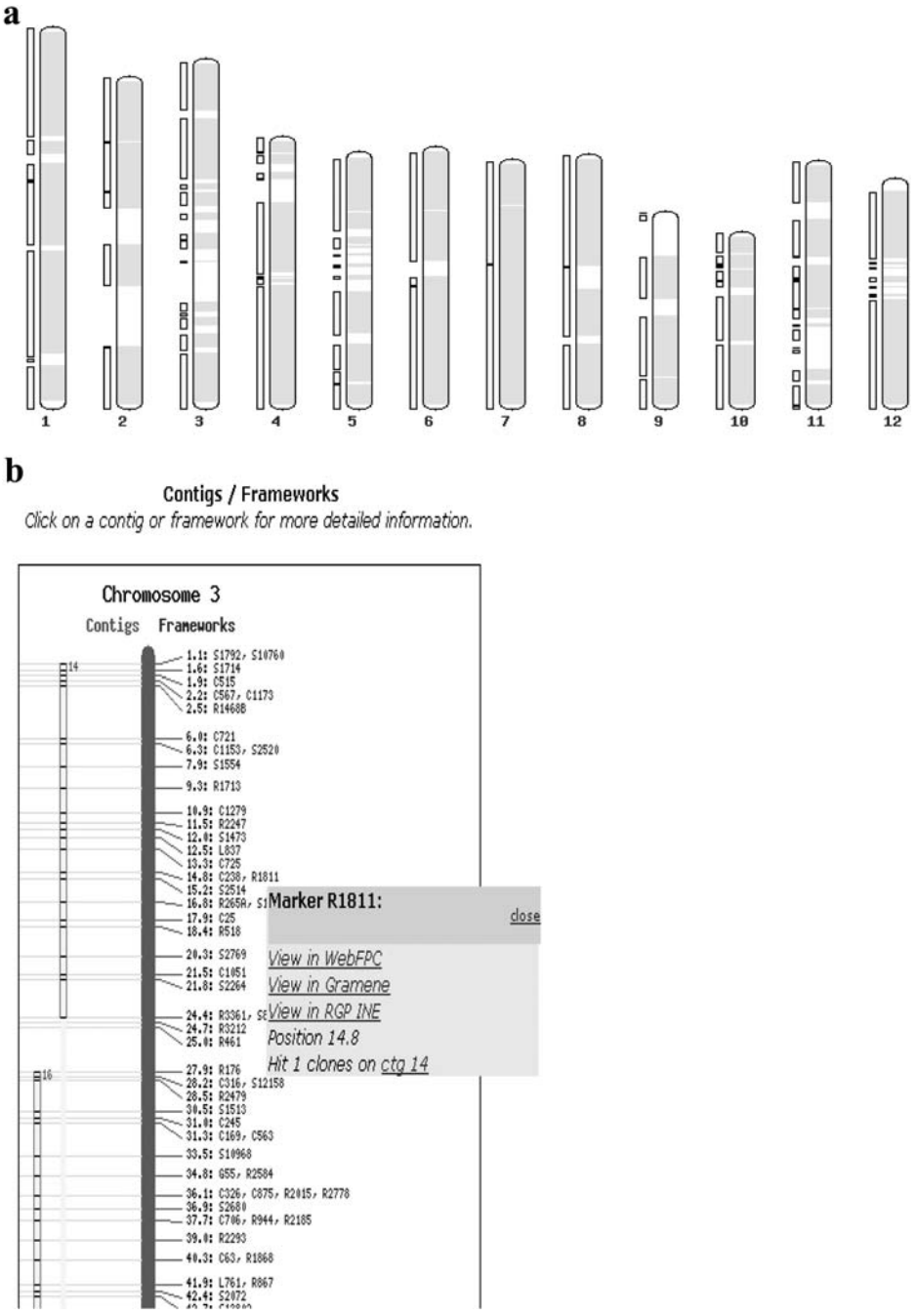

Figure 3. WebChrom chromosome displays. a, An overview of all chromosomes indicates the approximate lengths of each chromosome, the location of the centromeres, and which areas of the chromosome have been sequenced (shaded area). b, By clicking on a chromosome in the overview, the arrangement of FPC contigs along the chromosome together with the positions of anchored markers can be viewed. Links to WebFPC and other sites are shown in a popup when the user clicks an item.

available from our FTP site (ftp.genome.arizona.edu/pub/fpc/rice/).

Maize Map

The maize genome (approximately 2,500 Mb) will be the next cereal genome to be sequenced and will thus require a robust physical map and software tools to support the sequencing project as well as a variety of positional cloning projects. The maize FPC map currently contains 292,168 clones, 17,523 markers, and 2,998 anchors (Coe et al., 2002; Cone et al., 2002). A total of 13,810 of the markers are overgos derived from expressed sequence tags (Gardiner et al., 2004). The anchors are from the IBM neighbors map, of which the 1,931 frameworks are the IBM genetic markers and the 1,067 placements are from other genetic maps that have been binned in the IBM map (Fang et al., 2003). There are currently 766 contigs, of which 413 are assigned to chromosomes (release October 27, 2004). The maize FPC map is still under construction. The set of markers will continue to change as new ones are added and faulty ones are removed. The contigs will continue to change as they are being manually edited

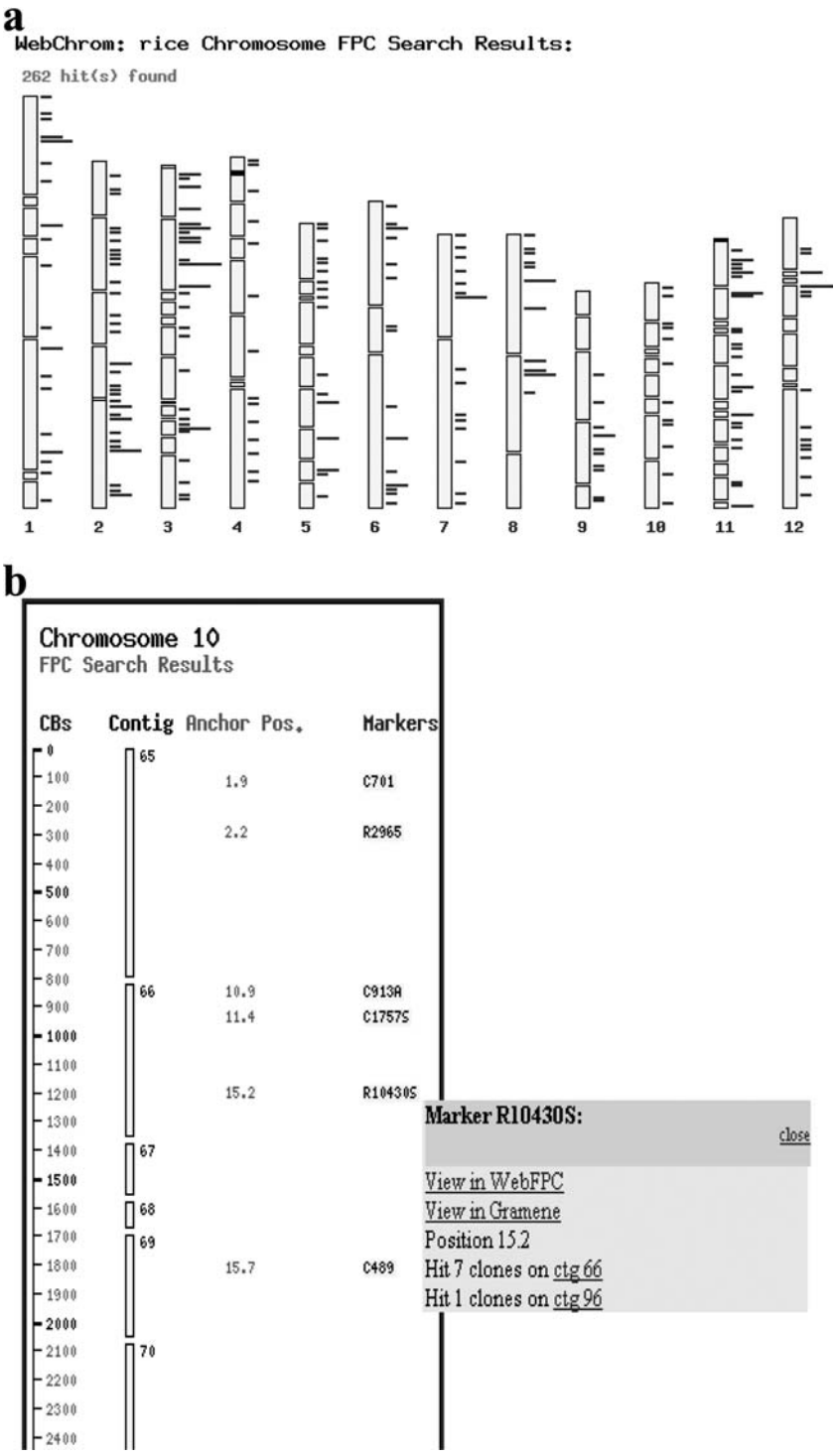

Figure 4. WebChrom Search. a, The location of all markers with SwissProt remarks. b, Selecting a chromosome shows the names and locations of the markers.

to merge contigs and split chimeric contigs (F. Wei, personal communication). The contigs are located on chromosomes using the FPC routine, but due to false positives, false negatives, and ancient polyploidization (Paterson et al., 2004), there is a fair amount of ambiguity. The locations are often manually edited so they are positioned correctly on the chromosome. For major releases, the contigs are renumbered to reflect the ordering along the chromosome, which gives the contig number a meaning in relation to the other contigs.

The maize FPC Web site has the same set of tools as the rice Web site. WebFPC links to maizeGDB (Lawrence et al., 2004), iMap (Fang et al., 2003), and GenBank (Benson et al., 2004). Both maizeGDB and iMap link to WebFPC based on a marker or clone

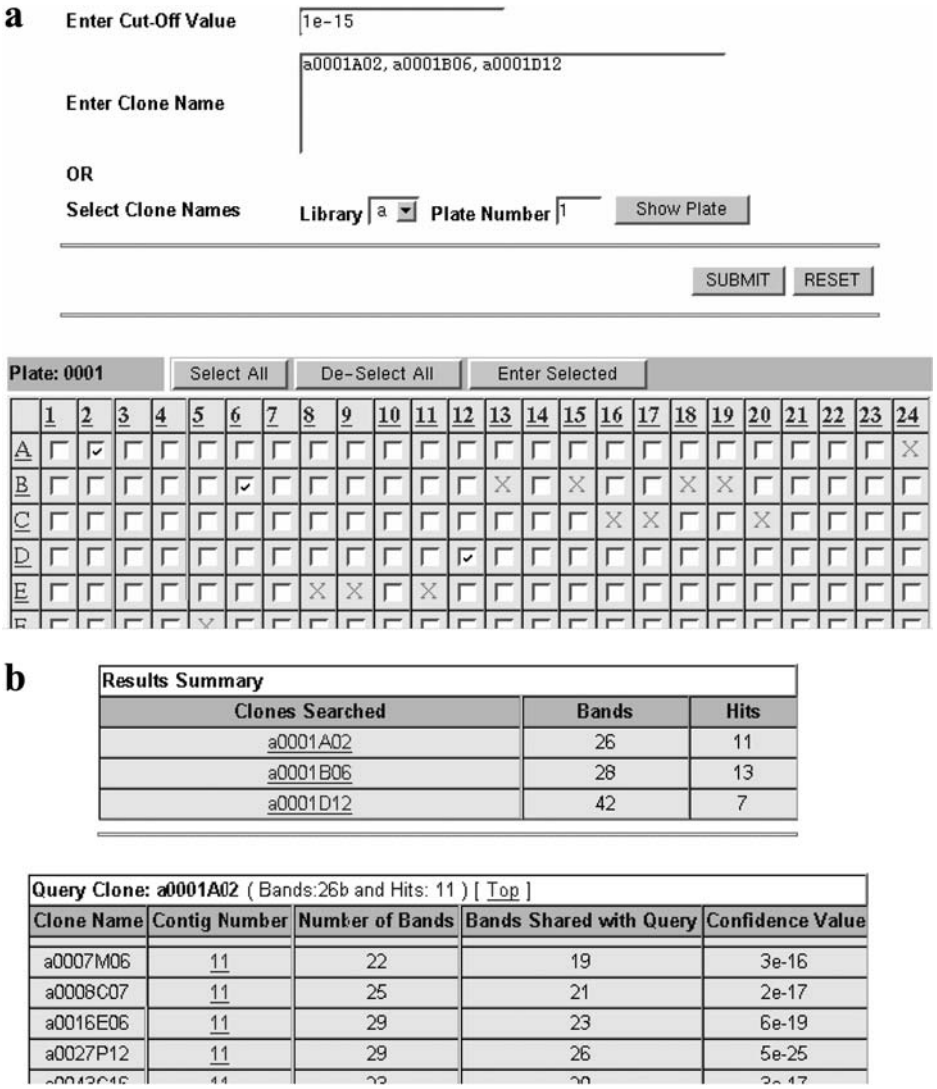

Figure 5. WebFCmp clone selection. a, The user can directly enter a clone name or select a plate to view and check which of the clones to compare against the FPC fingerprint database. Cells with an X indicate that the clone in that well is not available in FPC, most likely due to a failed fingerprint. b, The output of the comparator provides links into WebFPC.

name. WebChrom links to GenBank, maizeGDB, and WebFPC. The maize contigs tend to have many long yellow unbounded boxes indicating the ambiguity of marker locations on the contigs.

For the WebBSS, there are currently (as of February 7, 2005) 490 genomic sequences and 682,116 BESs to search against. As with rice, new and updated sequences are downloaded nightly so all available sequences are in the database. The maize FPC map currently contains 503 SD clones, of which 13 have the suffix of sd2 or greater, indicating they come from GenBank sequences that result in more than 55 bands. The number of SD clones will continue to grow since clones are currently being sequenced. The maize sequencing status page shows what clones have been sequenced and where they are located, along with their similarity to the original clone (see www.genome.arizona.edu/shotgun/maize/status).

A full description of the origin of the clones and markers in the maize FPC is presented at the Web site www.genome.arizona.edu/fpc/maize. Additionally, a High Information Contig Fingerprinting (HICF; Ding et al., 1999; Meyers et al., 2004; Nelson and Soderlund, 2005) map has been assembled by FPC and can be viewed from this URL.

Testing the Large Numbers of Anchors in the Human Map

Rice has a finished map and an almost finished sequence. Maize has almost 300,000 clones, and the map and sequencing are still in progress. These two datasets provide good test cases for maps with a large number of sequences and clones, respectively. To test the tools on a dataset with a large number of markers, we downloaded the human map from www.bcgsc.bc.ca/perl/humanbac (The International Human Genome Mapping Consortium, 2001), which has 69,507 markers and 26,164 anchors. WebFPC and WebChrom were installed for the human map (www.agcol.arizona.edu/fpc/human). The only

Markers	Len	Hits	BestCtg
Seq	463	5	1

Contig	CST	CloneHits
1	2297	5

BES Matches

First | Previous 20 Page 1 of 1 Next 20 | Last

BES	RC	Clone	Contig	Markers	Start	End	HitLen	Score	E-Value	Identity
OSJNBa0067C11r	n	a0067C11	1	Seq	187	463	486	549	e-156	100%
OSJNBa0051I19f	n	a0051I19	1	Seq	183	463	450	525	e-148	98%
OSJNBa0032A01f	y	a0032A01	1	Seq	1	174	647	337	7e-92	99%
OSJNBa0030P03f	n	a0030P03	1	Seq	183	463	594	549	e-156	99%
OSJNBa0001M01f	y	a0001M01	1	Seq	1	174	683	337	7e-92	99%

Figure 6. WebBSS results. The first table contains an entry for each FASTA format sequence in the uploaded query file. The second table gives a summary of the hits on a per-contig level. The third table lists all hits, and each hit has a link to that clone in WebFPC.

noticeable increase in wait time is during the display of results for the WebChrom Search software.

Setup and Distribution

The FPC tools with the prefix of Web are distributed as a package. A difficulty in distributing code for the Web is that it is a mix of Java, HTML, and CGI, where the three types of files go in different directories. An additional complexity is that the WebAGCoL package is composed of four tools with different requirements. A manual could be written to explain how to set up the files, but that would be tedious and error prone. Hence, we have written a setup script that reads

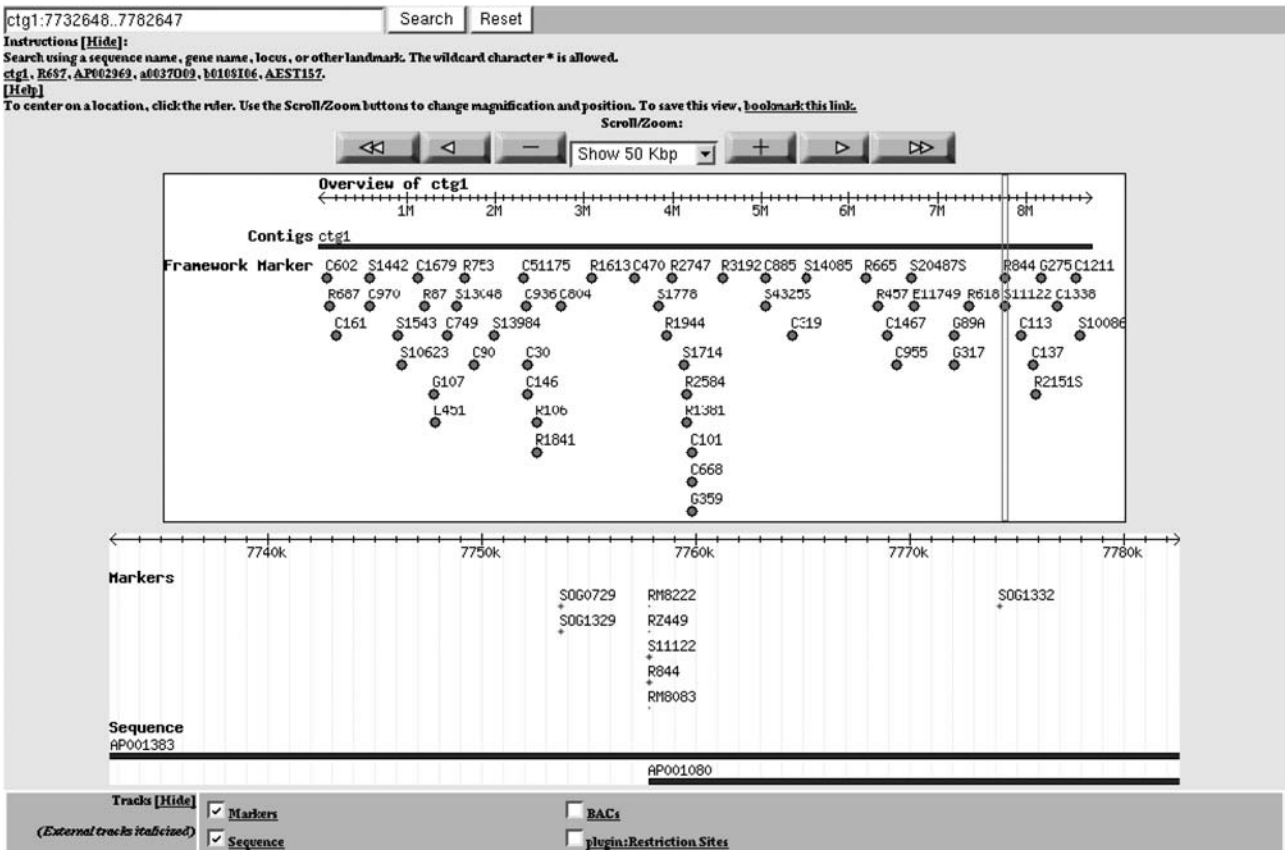

Figure 7. Rice contig 1 in GBrowse. The contig overview is shown at the top, and a 50-kb segment is displayed at the bottom. Marker and Sequence tracks are shown.

a configuration file and automatically runs the correct scripts and puts files in their correct directories. It also creates a script that can be run to update the Web sites based on an updated FPC database.

DISCUSSION

The philosophy of making data available to the public as soon as it is generated helps investigators stay up-to-date on the progress of their genomes of interest. To address this need, we developed a set of tools for viewing and analyzing FPC maps via the Internet. A significant advantage of Web based displays is that the host institution can automatically update the display when there is an updated FPC database instead of the user having to manually download new FPC releases. As an example, for the Maize Mapping Project (www.maizemap.org), the WebAGCoL tools were regularly updated as new clones were fingerprinted. Therefore, the community could easily follow the progress being made by simply visiting the maize FPC Web site. When the contig numbers have changed, the user can simply search in WebFPC by clone or marker for the new contig number.

Regular updates on the host site can be greatly simplified through the use of a cronjob. At AGCoL, a cronjob automatically updates the WebAGCoL tools when a given FPC database has changed. New and updated sequences are downloaded nightly from GenBank, a simulated digest is performed on them using Fingerprinted Simulated Digest (FSD; Engler et al., 2003), and they are positioned on the map when possible. Therefore, a sequence submitted to GenBank today will be present in the WebAGCoL tools tomorrow. The FTP site is updated with the modified FPC database, which triggers the nightly update of the WebAGCoL tools. Remote sites can have a cronjob that checks whether the FPC file has changed on the FTP site and, if so, downloads the new copy and updates the Web pages as necessary. Such automated coordination greatly enhances the relevancy of on-line tools since investigators can be confident they are looking at the most recent data. Cross links between sites, e.g. MaizeGDB and WebFPC, allow the user to see all the relevant data without requiring any one site to maintain everything. In the future, more elegant schemes for sharing data, such as Moby (Wilkinson and Links, 2002), will be used. Until then, this simple link scheme is easy to implement and keep updated.

The FPC Generic Genome Browser (GBrowse) displays the same data as WebFPC and WebChrom, but the layout and functionality are quite different. WebFPC closely models the way FPC displays data, but GBrowse resembles other well-known sequence-based Web browsers. In GBrowse, each scroll, zoom, or position shift requires the entire page to be redrawn, which can make extended browsing a bit tedious. Since WebFPC is designed as a Java applet, scrolling through a contig is immediate once it is loaded. GBrowse is flexible in what tracks are shown, whereas WebFPC is flexible in that entities can be colored or made invisible. GBrowse will show adjacent contigs while WebFPC does not. A final difference is that when a marker is clicked on in WebFPC, the clones it is associated to are highlighted; similarly, a clone can be clicked and its markers and remarks are highlighted; this feature is not in GBrowse.

While viewing FPC data via Web pages is quite useful, the ability to perform computations on the data on-line is desired as well since labs often do not have the high-performance computing resources nor the technical knowledge required for setting up the process locally. WebBSS addresses this demand by providing searches against a database of sequences associated with clones in FPC. To do this locally, the researcher would need to install BLAST (Altschul et al., 1997), download all BES and genomic sequences, and install FPC to run the BSS. Alternatively, they could run BLAST themselves, look through pages of BLAST output, and find each location in WebFPC by searching for each clone that was hit by the input sequence(s). Needless to say, running the WebBSS brings this functionality to the user without the overhead of setting up the process or searching through pages of BLAST output.

Whereas WebBSS shows the location(s) of a sequence on the map, the WebChrom Search tool shows the location of markers based on name or remark. We have currently demonstrated the ability to search on annotations based on SwissProt hits, which will be extended to use Gene Ontology annotations (The Gene Ontology Consortium, 2000).

We want to stress that WebFPC and its associated tools are not a replacement for FPC. If the user will be doing any serious querying of the data, FPC has much better search tools, is much faster, and FPC V7 (Engler et al., 2003) has a very sophisticated contig display. Additionally, the WebBSS only allows files of a maximum of 5,000 bases to be used as input, whereas the FPC BSS allows unlimited size; it also has much better querying support, and hits can be added as electronic markers. Several tutorials are provided to make it easy for the user to learn FPC. For example, Engler and Soderlund (2002) provide a tutorial on the basic usage of FPC, which should only take a few hours of the user's time. The tutorials are available from the FPC Web site.

MATERIALS AND METHODS

WebAGCoL Package

The WebAGCoL package contains WebFPC, WebChrom, WebBSS, and WebFCmp, along with a setup script and on-line help. Each tool has a preprocessor script that creates the files necessary for fast display. They all read a shared configuration file. All preprocessor scripts expect to read an FPC file written with FPC V7 or a later version. All preprocessors are written in Perl, all graphics except WebFPC are generated using the GD library (www.boutell.com/gd/), and Perl CGI is used for run time execution. At AGCoL, all processing and updates are performed on a Sun 280R with 2GB RAM.

WebFPC is written in the Java programming language and is implemented as an applet. The preprocessor perl script splits the FPC database at a contig level, writing information for each contig as XML. This allows a Java SAX (www.saxproject.org) parser to retrieve and parse information simultaneously, reducing the time spent waiting for a contig to be displayed. The preprocessor reads a site specific file to determine how to color clones and markers. For example, for the rice (*Oryza sativa*) and maize (*Zea mays*) FPCs, the sequenced clones and electronic markers are highlighted yellow. It reads the directory of reference files to set up links to external sites. The preprocessor also creates the initial HTML that accesses the WebFPC Java Jar file. A CGI script is used to execute an external lookup in order to display a contig without going through the initial page.

The WebChrom preprocessor splits the FPC file into one HTML file per chromosome and writes perl GD code to display the graphics. The WebChrom Search tool stores the markers in a ''Storable,'' which is a hash table and is used by the CGI script for fast searching.

For WebBSS, the preprocessor creates a modified FPC file that contains only the necessary information; this speeds up the reading of the file during the Web-based execution. The configuration file contains the paths to the BES directory and genomic sequence directory. FPC is run in batch mode to execute the BSS. The BSS saves the output in a file, which is read by a CGI script and displayed.

For WebFCmp, the preprocessor creates a modified FPC file that only contains the index into the file of bands in order to increase speed. FPC is run in batch mode to compare the fingerprints. A CGI script is run to read these files, compute the overlap score, and display the results.

The setup script reads a configuration file to determine where to put the CGI, HTML, and Jar file. It also reads the location of the reference file, FPC file, and target directories and runs the preprocessors. It creates the initial HTML file and writes a file called update.sh that can be used to update all the Web tools when a new version of the organism's FPC file is updated.

Processing at AGCoL

For the rice and maize Web sites, a cronjob is run nightly to download new and updated sequences from GenBank (as mentioned previously). Each GenBank file is parsed into a FASTA file of sequences and saved into a directory whose location is known by WebBSS; hence, this feature is always run on the latest sequences. A program called FSD2 (FPC Simulated Digest, Version 2) reads the GenBank file, cuts the sequence into overlapping clones, and creates the file of restriction fragment sizes. It also extracts the first author, clone name, and chromosome and writes the information into a file to be loaded as a remark for the clone. The size2band program uses the file of marker lanes (used by Image) to convert the sizes to bands. FPC is then run in batch mode to enter the new fingerprints and remarks into the database and position each clone in the same location as its best match. Once a clone has been positioned in FPC, it is not automatically repositioned when an updated record is entered. Therefore, we periodically remove all the SD clones from their contigs by making a keyset of them and executing ''Move to Ctg0'' from the pull down menu. They are then repositioned by executing ''Keyset->FPC'' on the keyset of SD clones. FSD2 is available from the FPC Web site.

BioPerl FPC Module

BioPerl (www.bioperl.org) is an initiative that seeks to simplify bioinformatics development by providing perl objects that perform mundane tasks such as parsing a file and retrieving information from it. To further this undertaking, we have developed a BioPerl module that reads an FPC file and allows the user to extract information from it. For example, one may retrieve all markers in a particular contig or find all clones attached to a marker. This module also converts FPC data into GFF format suitable for input into a Generic Genome Browser database, discussed in the next section.

GBrowse for FPC Map

The GMOD is ''a joint effort …to develop reusable components suitable for creating new community databases of biology'' (Stein et al., 2002; www.gmod.org). This community has developed GBrowse, which is a viewer designed to present linear genomic data on the Web. A GFF file is created to populate a mySQL database, and a configuration file is created that describes the tracks to be displayed in the GBrowse. The GBrowse software reads the configuration file and extracts data from the database in order to display the data. We have created the configuration file for displaying the FPC data. The BioPerl FPC module creates the GFF file to be loaded into the database.

Web Resources

www.agcol.arizona.edu/software/fpc: (1) FPC software, (2) tutorials, (3) links to other Web based FPC displays, and (4) the WebAGCoL, FPC GBrowse, and FPC BioPerl.
www.genome.arizona.edu/fpc/rice: The Rice FPC.
www.genome.arizona.edu/fpc/maize: The Maize FPC.
www.genome.arizona.edu/fpc_hicf/maize: The Maize HICF FPC.
www.agcol.arizona.edu/fpc/human: The Human FPC.
www.genome.arizona.edu/shotgun/maize/status: Maize sequencing status.
www.genome.arizona.edu/shogtun/rice/status: Rice sequencing status.
www.bcgsc.bc.ca/perl/humanbac: The human FPC download.
www.maizegdb.org: A central repository for public maize information.
www.gramene.org: A curated comparative genome analysis in the grasses.
www.maizemap.org: Maize mapping project.
www.maizemap.org/iMapDB/iMap.html: iMap display of maize genetic markers.
rgp.dna.affrc.go.jp/publicdata/geneticmap98/geneticmap98.html: Rice genetic markers.
www.bioperl.org: Open source Perl tools for bioinformatics, genomics and life science research.
www.gmod.org: The model organism system databases.
www.ebi.ac.uk/swissprot: Protein database of annotated protein sequence.
www.realvnc.com: VNC a tool to view and fully interact with one computer from any other computer.
www.boutell.com/gd/: C library for the dynamic creation of images.
stein.cshl.org/WWW/software/GD/: Perl Interface to GD Graphics Library
www.saxproject.org: Java API for XML.

ACKNOWLEDGMENTS

Scott Pearson and Luke Delorna wrote the WebChrom. Jayesh Sharma wrote the original WebBSS. Kiran Rao developed the rice and maize sequencing status pages and wrote the FSD2 script from the original FSD/ESD scripts. We thank Rod Wing and William Nelson for their valuable feedback on this manuscript. The Maize Mapping Project is a collaboration with the University of Missouri (PI Ed Coe, Karen Cone, Georgia Davis, Jack Gardiner, Michael McMullen, Mary Polacco, and Hector Sanchez Villeda), the University of Georgia (Andrew Paterson), and the University of Arizona (Rod Wing and Cari Soderlund).

Received November 18, 2004; returned for revision February 15, 2005; accepted February 20, 2005.

LITERATURE CITED

Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389–3402

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2004) GenBank update. Nucleic Acids Res **32:** D23–D26

Boeckmann B, Bairoch A, Apweiler R, Blatter M, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, et al (2003) The Swiss-Prot protein knowledgebase and its supplement TrEMBL. Nucleic Acids Res **31:** 365–370

Chen M, Presting G, Barbazuk W, Goicoechea J, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Higingbottom S, et al (2001) An integrated physical and genetic map of the rice genome. Plant Cell **14:** 537–545

Coe E, Cone K, McMullen M, Chen S-S, Davis G, Gardiner J, Liscum E, Polacco M, Paterson A, Sanchez-Villeda H, et al (2002) Access to the maize genome: an integrated physical and genetic map. Plant Physiol **128:** 9–12

Cone K, McMullen M, Bi IV, Davis G, Yim Y, Gardiner J, Polacco M, Sanchez-Villeda H, Fang Z, Schroeder S, et al (2002) Genetic, physical, and informatics resources for maize. On the road to an integrated map. Plant Physiol **130:** 1598–1605

Ding Y, Johnson MD, Colayco R, Chen YJ, Melnyk J, Schmitt H, Shizuya H (1999) Contig assembly of bacterial artificial chromosome clones through multiplexed fluorescence-labeled fingerprinting. Genomics **56:** 237–246

Engler F, Hatfield J, Nelson W, Soderlund C (2003) Locating sequence on FPC maps and selecting a minimal tiling path. Genome Research **13:** 2152, 2163

Engler F, Soderlund C (2002) Software for physical maps. *In* Ian Dunham, ed, Genomic Mapping and Sequencing, Genome Technology Series. Horizon Press, Norfolk, UK, pp 200–236

Fang Z, Cone K, Sanchez-Villeda H, Polacco M, McMullen M, Schroeder S, Gardiner J, Davis G, Havermann S, Yim Y, et al (2003) iMap: a database-driven utility to integrate and access the genetic and physical maps of maize. Bioinformatics **19:** 2105–2111

Fjell C, Bosdet I, Schein J, Jones S, Marra M (2003) Internet Contig Explorer (iCE): a tool for visualizing clone fingerprint maps. Genome Res **13:** 1244–1249

Gardiner J, Schroeder S, Polacco ML, Sanchez-Villeda H, Fang Z, Morgante M, Landewe T, Fengler K, Useche F, Hanafey M, et al (2004) Anchoring 9,3971 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiol **134:** 1317–1326

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, et al (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics **148:** 479–494

Lawrence C, Dong Q, Polacco M, Seigfried T, Brendel V (2004) MaizeGDB, the community database for maize genetics and genomics. Nucleic Acids Res **32:** D393–D397

Meyers BC, Scalabrin S, Morgante M (2004) Mapping and sequencing complex genomes: let's get physical! Nat Rev Genet **5:** 578–588

Nelson W, Soderlund C (2005) Software for restriction fragment physical maps. *In* K Meksem, G Kahl, eds, The Handbook of Genome Mapping: Genetic and Physical Mapping. Wiley-VCH Verlag GmbH, Weinheim, Germany, pp 285–305

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA **101:** 9903–9908

Soderlund C, Humphrey S, Dunhum A, French L (2000) Contigs built with fingerprints, markers and FPC V4.7. Genome Res **10:** 1772–1787

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. CABIOS **13:** 523–535

Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, et al (2002) The generic genome browser: a building block for a model organism system database. Genome Res **12:** 1599–1610

Sulston J, Mallet F, Staden R, Durbin R, Horsnell T, Coulson A (1988) Software for genome mapping by fingerprinting techniques. CABIOS **4:** 125–132

The Gene Ontology Consortium (2000) Gene ontology: tool for the unification of biology Nat Genet **25:** 25–29

The International Human Genome Mapping Consortium (2001) A physical map of the human genome. Nature **409:** 934–941

Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, et al (2002) Gramene: a resource for comparative grass genomics. Nucleic Acids Res **30:** 103–105

Wilkinson MD, Links M (2002) BioMOBY: an open-source biological web services proposal. Brief Bioinform **3:** 331–341