PlasMapper: a web server for drawing and auto-annotating plasmid maps

Xiaoli Dong, Paul Stothard, Ian J. Forsythe and David S. Wishart*

Departments of Biological Sciences and Computing Science, University of Alberta, Edmonton, AB, T6G 2E8, Canada

ABSTRACT

PlasMapper is a comprehensive web server that automatically generates and annotates high-quality circular plasmid maps. Taking only the plasmid/vector DNA sequence as input, PlasMapper uses sequence pattern matching and BLAST alignment to automatically identify and label common promoters, terminators, cloning sites, restriction sites, reporter genes, affinity tags, selectable marker genes, replication origins and open reading frames. PlasMapper then presents the identified features in textual form and as high-resolution, multicolored graphical output. The appearance and contents of the output can be customized in numerous ways using several supplied options. Further, PlasMapper images can be rendered in both rasterized (PNG and JPG) and vector graphics (SVG) formats to accommodate a variety of user needs or preferences. The images and textual output are of sufficient quality that they may be used directly in publications or presentations. The PlasMapper web server is freely accessible at http://wishart.biology.ualberta.ca/PlasMapper.

INTRODUCTION

Plasmid map generation is one of the oldest and most frequently performed operations in bioinformatics. Indeed, probably almost every practicing molecular biologist has worked with or generated a plasmid map to guide them through the cloning or plasmid manipulation process. Because of the size and complexity of plasmid molecules, computer-generated maps are absolutely essential to identify, locate and analyze key regions in a vector sequence. As early as the 1980s standalone computer programs were being described that supported the presentation and manipulation of plasmid maps on specific platforms and computer operating systems (1–7). Many of these early freeware packages have since been replaced by more sophisticated and far more user-friendly commercial

packages such as SimVector (Premier BioSoft), GeneTool (BioTools), VectorNTI (Informax, Invitrogen), MacVector (Accelrys), DNA Strider and LaserGene (DNAStar). Currently there are remarkably few freeware plasmid mapping programs still available, although pDRAW32 (AcaClone) is one example of an installable standalone package that supports plasmid mapping.

With the growing trend toward using freeware and freely available web tools in bioinformatics, it seems that the continuing dependency on expensive commercial packages to perform just a single operation (plasmid mapping) is somewhat questionable. Furthermore, with the increasing diversity of operating systems seen in many laboratories and the expanding level of inter-laboratory and international collaboration, we believe that a platform-independent solution to plasmid mapping is needed. One obvious solution is a plasmid mapping web server.

Here we describe a web server, called PlasMapper, which is able to accept FASTA-formatted DNA sequences and generate a fully labeled/annotated plasmid map (both graphical and textual output) with essentially no further user input. A central innovation in PlasMapper is its capacity to automatically identify and label the plasmid control sequences found in both eukaryotic and prokaryotic vectors using a large database of common plasmid sequences and common plasmid subsequences (replication origins, promoters, terminators, marker genes, etc.). PlasMapper supports a wide range of textual and visual display options that allow users to easily customize the image or textual output. It is able to generate plasmid maps of sufficient quality and resolution that they may be readily used in publications or presentations. PlasMapper is specifically designed to make plasmid annotation trivially simple and to facilitate the sharing and dissemination of plasmid images and plasmid data across all computer platforms.

PROGRAM DESCRIPTION

PlasMapper is composed of three parts: a front-end web interface (generated using Java), a back-end for rendering and sequence matching (written in Java and C) and a Feature Site Database (FSD) consisting of 336 DNA sequence motifs

---

*To whom correspondence should be addressed. Tel: +1 780 492 0383; Fax: +1 780 492 5305; Email: david.wishart@ualberta.ca

Table 1. Features in each category of the Feature Site Database consists of 10 different feature categories and consists of 336 total features

Category	No. of features
Promoters	43
Regulatory sequences	63
Terminators	29
Selectable markers	25
Replication origins	11
Reporter genes	21
Affinity tags	26
Localization sequences	17
Two-hybrid genes	10
Miscellaneous features	91
Total	336

(promoters, terminators, selectable markers, etc.; Table 1) and 457 restriction enzymes from the Restriction Enzyme Database (8). The FSD was compiled from an extensive survey of commercially and publicly available plasmids. PlasMapper accepts FASTA DNA sequences up to 20 000 bases in length as input to its sequence window and performs checks on both the length and validity of the DNA sequence prior to conducting any further analysis. To facilitate the generation of maps for commonly used plasmids, the PlasMapper website also maintains a growing repository of 288 vector sequences available from various vendors and suppliers. These sequences may be selected and automatically uploaded into the sequence window using the 'Plasmid Library' button. To facilitate tracking, editing and 'virtual cloning' the sequence text box has a '(Re)Format' button that allows the raw FASTA sequence file to be block formatted and numbered. Insertions (cloned genes), deletions, mutations, edits and corrections can all be made readily in this specially formatted view. As soon as any edits are completed, the user can press the '(Re)Format' button to reformat and renumber the modified sequence. After a sequence has been pasted, edited, selected or reformatted, the 'Submit' button can be used to begin the map generation process. As seen in Figure 1, several options are available for specifying which features are displayed and how the map image is rendered. More detailed descriptions of how annotation, labeling and rendering can be controlled are provided in the on-line Help page.

The feature identification and image rendering portion of PlasMapper consists of four separate programs: BLAST, FIND-SITE, FORMAT and CGView. BLAST (9) is used to identify portions of the supplied sequence that match promoters, terminators, selectable markers, reporter genes and replication origins stored in the FSD. The BLAST program parameters have been optimized for PlasMapper by testing more than 50 randomly selected commercial plasmids to ensure that the resulting annotations completely matched those reported by the vendors. The FIND-SITE program, which uses several components of BioJava (http://www.biojava.org/), is used to identify open reading frames (ORFs) and type II restriction enzyme cutting sites. The third program (FORMAT), also written in Java, generates formatted text output which displays the plasmid sequence (60 bases per line, numbered, courier font) with the requested annotations displayed in stacked, non-overlapping positions

above each sequence line. The final program (CGView) uses the Java2D API to convert the results obtained from BLAST and FIND-SITE into a graphical map. Specifically, CGView accepts sequence feature information (feature name, feature type, position and strand) and generates a collection of two-dimensional objects. Each object's shape (an arrow or an arc), color, opacity and position are adjusted according to the attributes of the feature represented. After the objects are drawn, the feature names are placed on the map using an iterative collision detection and shifting process that results in a visually pleasing label arrangement and no label overlap. Java classes included as part of the Java API are used to convert the map into JPG and PNG images. SVG output is generated using the Batik SVG Toolkit (http://xml.apache.org/batik/). Because SVG is a vector format, SVG images can be scaled without any noticeable degradation. Most web browsers can display SVG images using the freely available Adobe SVG Viewer plug-in (http://www.adobe.com/).

PROGRAM OUTPUT

PlasMapper generates both text and graphics output (JPG, PNG or SVG format). The default view is the graphic image, with a button to create the text view in another window. Figure 2 provides an example of the range of text and graphics outputs that can be produced. PlasMapper is designed to be interactive, allowing users to readily go back and forth between the PlasMapper home page and the output, so that new images or text can be generated using a variety of display options. Using radio buttons, text boxes and pull-down menus, users can control a large number of display variables from the 'Options' portion of the PlasMapper home page. Figure legends, color schemes, shading, directional arrows, restriction enzyme labels and plasmid feature labels (promoters, terminators, replication origins, tags, etc.) may be easily turned on or off using radio buttons. Similarly, custom feature titles, minimum ORF lengths, and image titles may be added or edited using text boxes. Image formats, image sizes, plasmid circle widths and gene arc widths may be altered using pull-down list boxes. The images and textual output generated from PlasMapper can be saved or copied directly to the user's hard disk or placed into standard word-processing, presentation or image manipulation programs. Images and text generated by PlasMapper are kept in a /temp file and are stored for not more than 24 h on the PlasMapper server. Similarly, PlasMapper sessions that are inactive for more than 20 min are terminated.

DISCUSSION

The PlasMapper server provides a convenient and easily accessible solution to plasmid annotation and drawing for users who normally depend on freeware or free web servers such as EMBOSS (10) or The Sequence Manipulation Suite (11). The simplicity and accessibility of PlasMapper should also make it a useful tool for teaching or training high school and university students in introductory molecular biology or genetics courses. Furthermore, we believe that the use of web-server technology should enable or encourage the sharing of plasmid data and images among geographically distant labs or
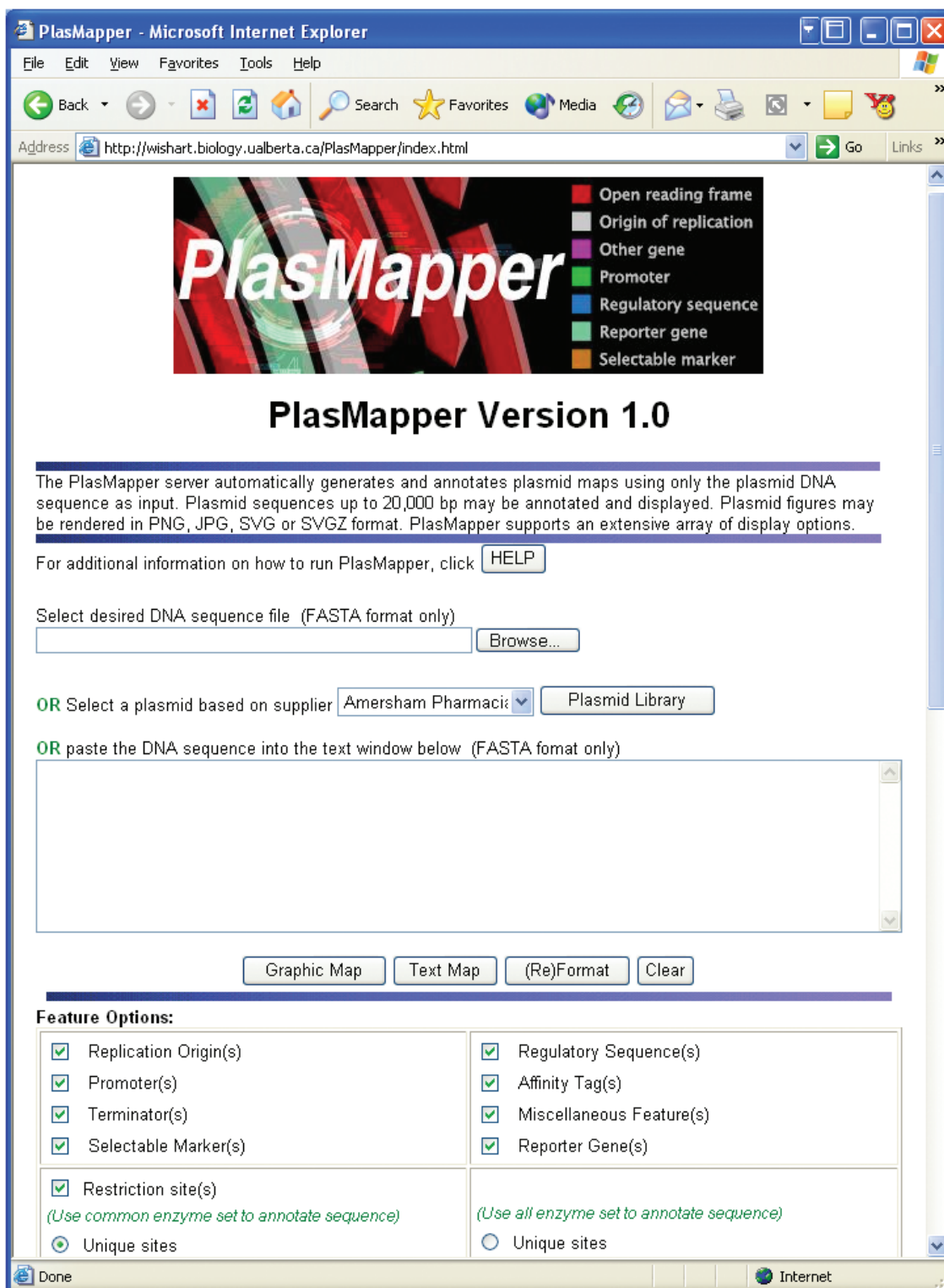
Figure 1. The PlasMapper entry page provides a graphical user interface that allows users to upload a sequence from their local computer or paste a FASTA-formatted sequence into an input box for plasmid map generation. This page also includes many options that enable the user to customize the output.

between labs that normally use incompatible software and/or computer platforms.

In trying to make the PlasMapper interface as simple as possible, some sacrifices in flexibility had to be made. No doubt some users may not want a specific label displayed or will dislike the default color scheme. Others may find certain rare restriction enzymes or gene features missing from the FSD. Likewise, the limited choice of plasmid circle
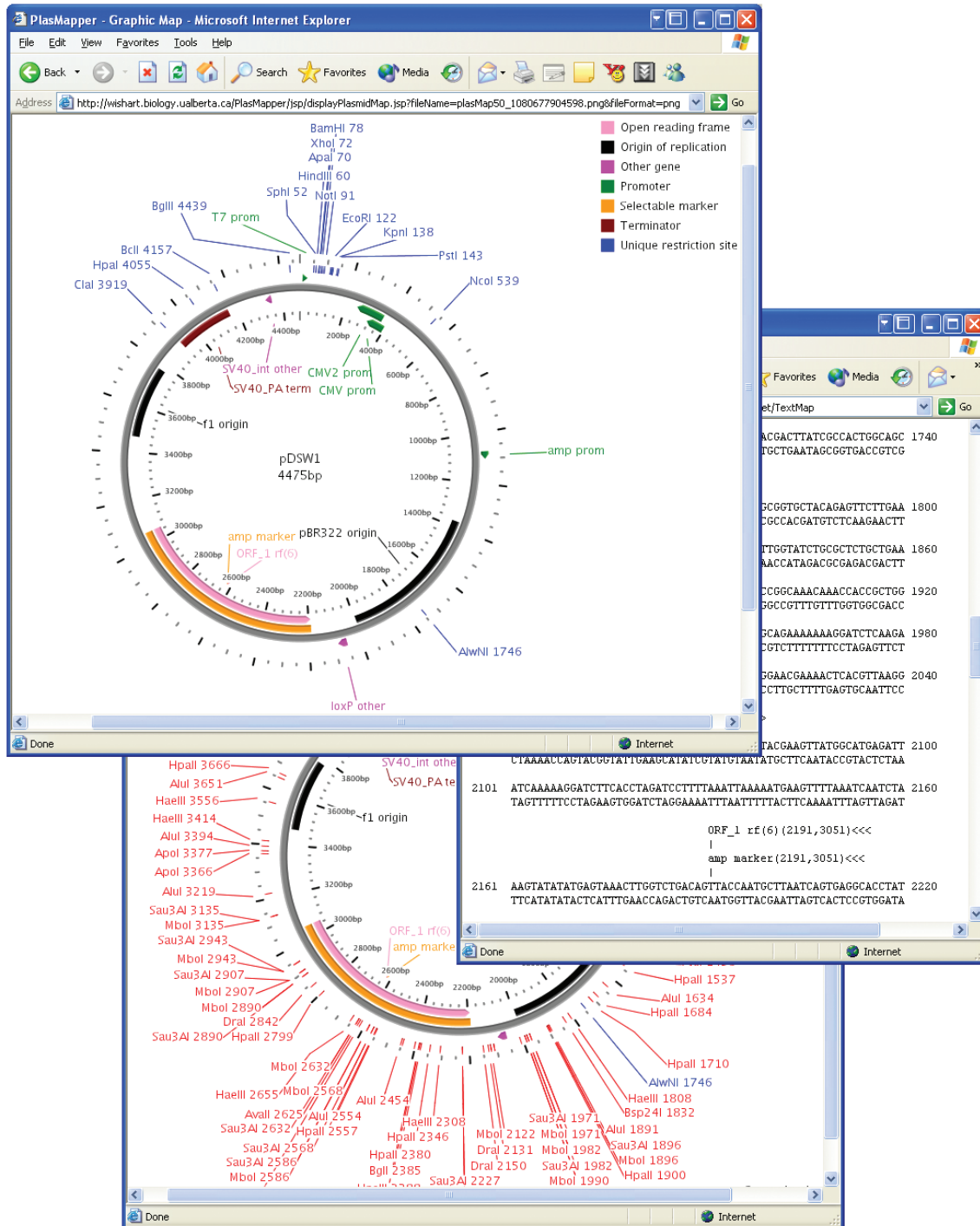
Figure 2. A screenshot montage of PlasMapper output showing an example of two graphical views and a text view. The graphical views show labeled features including replication origins, restriction enzyme sites, selectable markers and reporter genes. The text view shows the individual DNA base pairs as well as the feature names.

or gene/marker widths may seem too restrictive. In many cases, these problems can be addressed by simply pasting the PlasMapper image into an image manipulation package to change the offending feature. The SVG format is helpful in this regard, since individual features and labels can be repositioned or modified using vector-graphics-capable software. For more specialized maps and annotations,

commercial sequence analysis packages may be a more appropriate choice.

In summary, PlasMapper is a web server that permits the automated annotation and rendering of circular plasmids for both eukaryotic and prokaryotic vectors. It combines database-searching and pattern-matching techniques with a unique collection of plasmid-feature sequences to automatically

generate publication-quality text and images. PlasMapper supports a wide range of display and formatting options and should make plasmid analysis and manipulation much simpler and far more accessible. The PlasMapper web server is freely accessible at http://wishart.biology.ualberta.ca/PlasMapper.

ACKNOWLEDGEMENTS

REFERENCES

1. Abremski,K. and Ward,D.F. (1986) Plasmid map: a microcomputer program for display and storage of plasmid data. *Gene*, **46**, 127–130.
2. Filippone,E. and Lurquin,P.F. (1988) PROPLASM: an Apple Macintosh computer program for proportional plasmid map drawing. *Biotechniques*, **6**, 574–575.
3. Liu,J.D. and Parkinson,J.S. (1989) A Macintosh program for drawing circular plasmid maps. *Comput. Appl. Biosci.*, **5**, 237–238.
4. Peterson,E.A. and Ward,D.F. (1990) CLONE 3: plasmid drawing and clone management software program for microcomputers. *Biotechniques*, **8**, 690–693.
5. Dolz,R. (1994) GCG: drawing circular restriction maps. *Methods Mol. Biol.*, **24**, 35–46.
6. Reda,D. and Reda,A.C. (2000) Redasoft Plasmid 1.1: software for easy, efficient cloning and map drawing. *Curr. Issues Mol. Biol.*, **2**, 37–39.
7. Tsudzuki,T. (2000) A graphic tool for circular genome maps. *Nucleic Acids Symp. Ser.*, **44**, 189–190.
8. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2003) REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Res.*, **31**, 418–420.
9. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
10. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
11. Stothard,P. (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, **28**, 1102–1104.