ContigMaker:
Software Tool for Contig Map Construction

Akira Suyama¹  Masami Hagiya²  Takashi Ito³
suyama@tansei.cc.u-tokyo.ac.jp  hagiya@is.s.u-tokyo.ac.jp

Asao Fujiyama⁴  Akira Ohyama⁵  Toshihisa Takagi³
afujiyam@ddbj.nig.ac.jp  akr@mitsui-knowledge.co.jp  takagi@ims.u-tokyo.ac.jp

¹Institute of Physics, College of Arts and Sciences, The University of Tokyo
3-8-1 Komaba, Meguro-ku, Tokyo 153, Japan
²Department of Information Science, Graduate School of Science, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
³Human Genome Center, Institute of Medical Science, The University of Tokyo
4-6-1 Shiroganedai, Minato-ku, Tokyo 108, Japan
⁴Division of Human Genetics, National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411, Japan
⁵Mitsui Knowledge Industry Co. Ltd.
3-7-4 Kojimachi, Chiyoda-ku, Tokyo 102, Japan

Abstract

ContigMaker is a software tool to aid contig map construction. It is a Motif application running on UNIX workstations with the X Window System. ContigMaker is composed of five
major components: map data manager, map analyzer, map viewer, map aid, and project manager. Contig-mapping data obtained by experiments are stored in a database of the map data manager. The stored data are then subjected to analysis by the map analyzer to generate contigs. ContigMaker supports the two strategies for contig construction: the STS (sequence-tagged sites) strategy and the MOF (mapping by oligonucleotide fingerprinting) strategy. The generated contigs are assembled into a contig map according to positions of landmarks falling on the contigs. ContigMaker allows a user to extract landmark information from a public genome database such as the GDB. The contig maps constructed are graphically drawn by the map viewer. The map aid provides miscellaneous small useful tools to finish a contig-mapping task. A repeated task ContigMaker performs can be automated by a macro created by the project manager. The macro will save time and effort for contig map construction.

1 Introduction

A contig map is a physical map of genome, consisting of overlapping recombinant DNA clones. Construction of a contig map of the whole human genome is a major mid-term goal of the human genome project because the contig map offers a primary guide for the whole genome sequencing [1, 2]. The map also offers immediate access to a particular genetic segment of the genome, thus facilitating experimental analysis of biological functions the segment holds.

Even a contig map extending over the smallest human chromosome consists of hundreds of 1 MB YAC clones. A detailed map, consisting of smaller P1 or cosmid clones, contains thousands of clones. Therefore, construction of a detailed contig map of the whole human genome needs millions of overlapping clones ordered. It means that no contig maps spanning human chromosomes cannot be constructed without a computer aid.

ContigMaker is a software tool to offer a computer aid for constructing contig maps from experimental mapping data. It supports the two strategies for contig construction: the STS (sequence-tagged sites) strategy [3] and the MOF (mapping by oligonucleotide fingerprinting) strategy [4]. ContigMaker assembles contigs generated from experimental data to draw contig maps by consulting positional information on landmarks extracted from a public genome database, the GDB, or a published literature. ContigMaker provides other miscellaneous tools aiding laborious daily mapping experiment.

2 System Architecture
ContigMaker consists of five major components: map data manager, map analyzer, map viewer, map aid, and project manager. The map data manager provides a database of various data used in ContigMaker. Data of markers, landmarks, clones, and contigs, for example, can be managed and be edited by the map data manager. The map analyzer offers analysis tools for generating contigs from experimental data and for constructing contig maps by assembling contigs. The map viewer is a graphical viewer for contigs and contig maps. The map aid contains miscellaneous useful tools especially for contig-mapping experiments. The project manager is introduced to make easy-to-use environment for routine contig mapping work. It is a macro manager to automate a mapping task.

ContigMaker is a Motif application running on UNIX workstations with the X Window System. The present version of ContigMaker requires no commercial DBMS. A simple database manager to manipulate a database of markers, clones, and contigs is included in ContigMaker. The future version is planned to run on a commercially available relational or object-oriented DBMS.

![Figure 1. A record of a clone database](image)

This is reference for clone001, this clone is for contig4001.

This is remark for clone002.

This is remark for clone003.
3 Map Data Manager

A large collection of recombinant DNA clones, organized in a complicated hierarchy created by successive subcloning, is used in experiments for constructing contig maps. A clone database provided by the map data manager keeps information on both the hierarchy and individual clones to maintain and to use clone libraries for contig mapping.

![Figure 2. A record of a marker database](image)

![Figure 3. A button-switch panel for hit-relation input](image)
Figure 1 shows an editor window to create and update a record in a clone database. Fields of "MOTHER CLONE" and "DAUGHTER CLONE" in a record allow a user to trace complicated hierarchy structure of a clone library to seek an objective clone. A "FREEZER" field lets a user know where the clone is actually stored. Information on a clone base sequence is stored in a "SEQUENCE" field. Terminal sequences in this field offer data to make STS markers sitting near each end of a clone.

The present version of ContigMaker supports the two contig mapping strategies, the STS strategy and the MOF strategy. Both strategies generate contigs from hit-relation data obtained by contig mapping experiments. The hit-relation shows a list of clones each STS or each random oligonucleotide marker hits for the STS or the MOF strategy, respectively. A marker is defined as hitting a clone when the clone contains the marker. Thus, a marker database keeping hit relations is the most fundamental data for contig construction based on the STS or the MOF strategy.

Figure 2 shows an example of a record in a STS marker database. A "SEQUENCE" field indicates a base sequence for a PCR primer pair consisting of a STS marker. The size of PCR product attributed to the marker is stored in a "PCR PRODUCT SIZE" field. A "HIT CLONES" filed keeps the hit-relation, that is, names of clones the marker hits. The button-switch panel
shown in Fig. 3 is furnished to facilitate the hit-relation input. STS markers are in the leftmost column; clones in the top row. Clones hit by a marker can be easily selected without mistakes by just clicking the corresponding buttons in the panel.

Contig mapping projects are usually performed library by library containing different chromosome or different part of the same chromosome. These various mapping projects are often going on in parallel. ContigMaker thus keeps all of those mapping data at once. In this case, data of clones and markers, for example, involved in different mapping projects should be divided into different groups to eliminate unnecessary confusion. The grouping is also necessary even for a single mapping project because a group of markers or clones often varies at the stage of the project.

The map data manager has a feature to divide clone and marker data into groups. Figure 4 shows a window listing markers in a maker database. Markers you want to put together into the same group can be selected in this window by clicking marker IDs or by search commands. A created group can be specified by its name. The operation range of instructions in ContigMaker can be restricted to data falling into the specified group.

Figure 5. A contig map view
4 Map Analyzer

The hit-relation stored in a marker database is analyzed by the contig constructor in the map analyzer to generate contigs compatible with the hit-relation. Generated contigs are then assembled into a contig map by the map constructor in the map analyzer. To assemble contigs, positions of landmarks sitting on clones composing the contigs are used. ContigMaker can refer to a public genome database, the GDB, to get information on the landmark positions. Landmark positions reported in a published literature or obtained by original experiments can also be used to get contigs together into a contig map.

The present version of ContigMaker supports the two strategies for creating contigs, the STS strategy and the MOF strategy. For contig construction based on the STS strategy, ContigMaker uses the new algorithm based on the deductive method, which copes with incompleteness and incorrectness unavoidable for actual hit-relation data [5]. For the MOF strategy, the algorithm based on the multi-dimensional scaling with the ability of coping with unavoidable experimental errors is used to create contigs from the hit-relation data [4].
5 Map Viewer

The map viewer is a graphical viewer to look at contigs and contig maps constructed by the map analyzer. Fig. 5 shows a graphical view of a contig the map viewer draws. Clones are designated as horizontal bars. Positions of STS markers sitting near each end of a clone are indicated by small vertical bars on the ruler appeared at the top of the window. A landmark of which position along the genome is determined can also be shown on the ruler as a vertical bar in a different color. Isolated clones are shown in a small window appeared on the right side part of the viewer.

The map viewer will offer an easy-to-use graphical user interface to perform analysis operations on contigs and contig maps. The future version of ContigMaker will support this feature intensively.

6 Map Aid

The map aid is a miscellany of small useful software tools aiding a variety of tasks encountered in daily contig mapping work. One of software tools often used in the map aid is a program to design a base sequence of a PCR primer pair composing a STS marker from each terminal base sequence of a clone. This tool is indispensable for contig construction by the STS strategy. The future version of ContigMaker is planned to be equipped substantially with small useful tools in the map aid.

7 Project Manager

The project manager is a macro utility to automate repeated tasks in ContigMaker. A set of instructions recorded in a project macro can be executed repeatedly just by running the macro. It thus makes it easy to perform daily routine work in ContigMaker. Especially for a newcomer to a computer, a project macro offers a powerful aid because it allows him to use software tools in ContigMaker with little knowledge on a computer.

How to use a project macro is very simple. To record a macro, you just turn on the macro recorder and record every step you perform in ContigMaker to finish your task. The recorded macro can then be run just by double-clicking its name in the list of project macros in Fig. 6. If it is a macro to edit information on a marker, for example, you can immediately reach the window shown in Fig. 2 and start editing. Edited markers can be restricted automatically to the
group of markers involved in your present project if the macro contains instructions to create a marker group to which an editor command is applied. This automatic restriction will help you avoid making unnecessary mistakes.

8 Conclusion

ContigMaker is a software tool designed to support analysis and experiment for constructing contig maps. The present version is now being tested at laboratories involved in contig mapping projects. Features of ContigMaker will be updated substantially to meet demands made at the laboratories during the test.

Acknowledgments

The authors wish to thank Dr. Y. Sakaki and Dr. H. Tanahashi, HGC, Institute of Medical Science, The University of Tokyo, for their valuable discussions. The authors also thank Dr. S. Mitaku, Tokyo University of Agriculture and Technology, for his helpful suggestions. This work was supported by a Grand-in-Aid for Scientific Research on Priority Areas, "Genome Informatics", from the Ministry of Education, Science and Culture of Japan.

References