APPLICATION NOTES

Genomics & Informatics Vol. 5(1) 30-31, March 2007

ChroView: A Trace Viewer for Browsing and Editing Chromatogram files

Hongseok Tae[1,2], Eun-Bae Kong[2] and Kiejung Park[1]*

[1]Information Technology Institute, SmallSoft Co. Ltd., Daejeon 305-343, Korea, [2]Deptartment of Computer Engineering, Chungnam National University, Daejeon 305-764, Korea

Abstract

Many visualization tools have been designed to aid information processing during whole genome projects. We have developed a trace viewer program, ChroView, which can read a chromatogram file and display the chromatogram traces of the four bases. The program can be used to examine sequencing quality and base-calling errors. It can also help researchers to edit and save base-calling results while browsing the traces. Additionally, this program has a base-calling feature which can produce supplementary data for validation of the results from other base-calling programs.

Availability: The trial version of ChroView is available from http://gate.smallsoft.co.kr:8006/~hstae/docs/ChroView.html

Keywords: ChroView, base-calling, chromatogram, trace view, contig assembly

Introduction

The Sanger method for DNA sequencing was invented in the late 1970s (Sanger *et al*., 1977). Since then, whole-genome shotgun (WGS) sequencing projects have been producing enormous amounts genomic information. Base-calling processes identify sequences of DNA fragments by analyzing chromatogram files from sequencing machines such as the ABI DNA Analyzer. To support the high throughput of sequencing projects, several base-calling algorithms (Giddings *et al*., 1998; Ewing *et al*., 1998) have been developed. Although Phred (Ewing *et al*., 1998) is the most commonly used base-calling program, its error rate is reported to be 5 - 6%. The reasons for this include contamination, faint signal and signal interference during the shotgun process. The chromatogram traces can often be ambiguous and base-calling programs can predict incorrect bases. As this causes the miss-assembly of contigs, manual intervention is still required to correct the bases. To aid in the validation of base-called sequences, the visualization of chromatograms is essential. Although a few chromatogram viewer programs such as ABIView (Klatte, 1996) and Chromas (McCarthy, 1996) have been developed to make the validation step easier, the features these tools provide are insufficient.

We have developed a practical chromatogram viewer, ChroView, to aid in the validation step. The program runs on PCs using the Windows™ operating system, and has several useful functions including a base-calling feature based on simple fourier transform.

Features and Results

ChroView reads chromatogram files in ABI or SCF format, and displays the chromatogram traces. SCF is a standard format for the representation of chromatogram data for DNA sequencing. ABI format is a representation generated by ABI DNA Analyzer. Chromatogram files contain the DNA sequence and spectrum values of the four bases at discrete time points. ChroView uses a Bezier function to generate smooth curves for the spectrum values, and shows the DNA bases at assigned positions.

The reverse strands of DNA can also participate in contig assembly. ChroView shows chromatogram traces of the reverse strand of displayed DNA fragments. To allow the detailed investigation of chromatograms, ChroView can draw a single trace for a selected base (Fig. 1A). With this feature, users can compare the trace patterns of four bases, and thereby check the base-calling result. Slide bars at left and top sides of the window frame are used to control the vertical and horizontal scales of the traces respectively.

When bases need to be corrected, users can edit them by deletion, insertion or substitution. ChroView also provides an undo operation to cancel editing actions. This feature aids users to revise their mistake without reading again the original data file. Users can compare the edited sequences with the original sequence (Fig. 1B). The edited result can be stored as an SCF or ABI file. Printing the chromatogram is important for the observation of the overall pattern of traces. Most researchers studying contig assembly examine the pattern rather than local peaks. ChroView users are able to print the traces in multiple rows. The program also provides a print preview option which is useful when adjusting the scale of a chromatogram before printing (Fig. 1C).

*Corresponding author: E-mail kjpark@smallsoft.co.kr, Tel +82-42-864-2524, Fax +82-42-385-9240
Accepted 8 Jan 2007

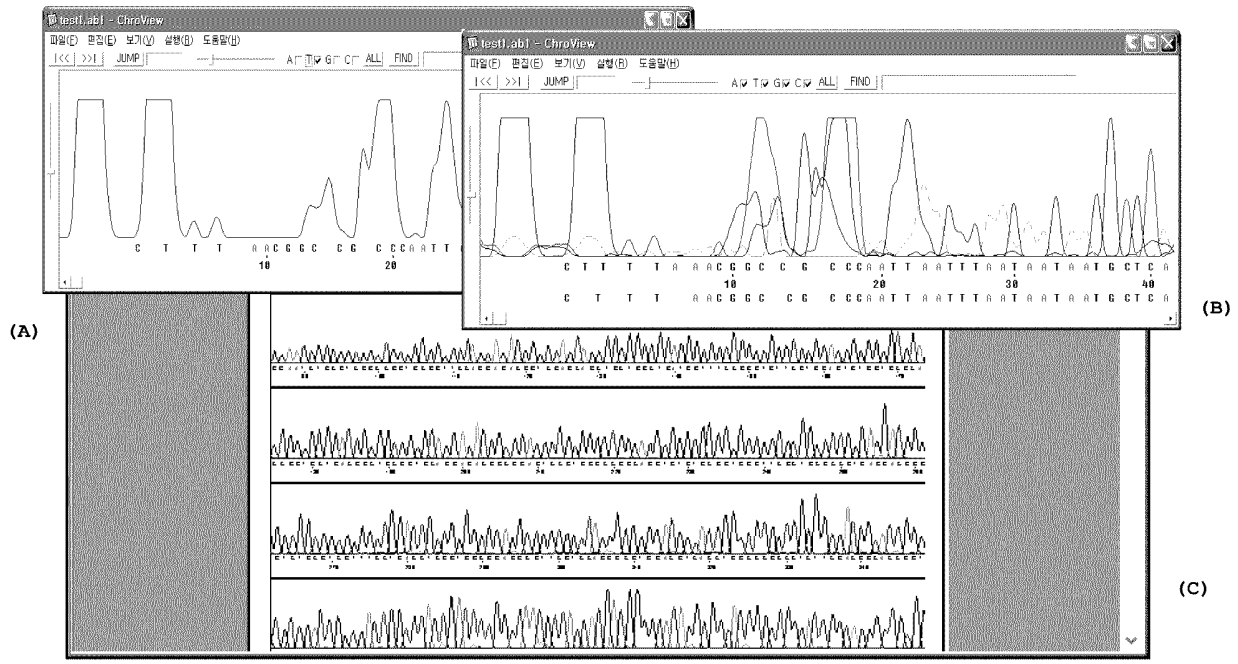

Fig. 1. The interface of ChroView. (A) Users can observe a single trace to closely study the pattern of the trace. (B) After editing the erroneous bases, the edited bases can be compared with original bases. (C) ChroView provides a print preview option and can print traces in multiple rows.

ChroView can carry out base-calling for traces and replace the original sequence with its base-called sequence. Comparing the base-calling result of ChroView with the original sequence is an efficient method for validation. The first step of the base-calling algorithm is the examination of four traces to detect preliminary peaks. The peaks are scanned to find out regions showing the most uniformly spaced peaks in a window of 200 trace points. The second step is to determine solid peaks by the application of simple Fourier transform. A period of peaks is estimated from the region, and a cosine curve with the corresponding period is drawn on the traces. The center of each peak in the cosine curve can shift from -0.3 to +0.3 along x-axis. The peaks of traces which are well matched with peaks of the cosine curve are marked as solid peaks. The third step is to refine ambiguous peaks. If a peak is considerably broader than the other peaks, it is considered to be a merged peak of two or three peaks and should be split. After all of the peaks are scanned in the window, the window shifts by 100 trace points and the algorithm is applied iteratively. The final step is to store the predicted peaks and DNA bases to a chromatogram file.

Discussion

Base-calling errors are the cause of several problems in contig assembly. ChroView has been developed as a visualization program to aid researchers in correcting these errors with editing, print preview and several features. The base-calling feature is not included in other trace viewers. The base-calling results of ChroView analyses, which remain to be rigorously tested, could be used as supplementary data to validate the data generated by other base-calling programs. Further research on base calling algorithms and related programs is an important and challenging task which will have a great impact on genome research.

References

Ewing, B., Hillier, L., Wendl, M., and Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* 8, 175-185.

Giddings, M.C., Severin, J., Westphall, M., Wu, J., and Smith, L.M. (1998). A software system for data analysis in automated DNA sequencing. *Genome Res.* 8, 644-665.

Klatte, D.H. (1996). ABIView. *http://bioinformatics.weizmann.ac.il/software/abiview/abiview.html.*

McCarthy, C. (1996). Chromas. *http://www.mb.mahidol.ac.th/pub/chromas/chromas.htm.*

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). Chain Sequencing with chain terminating inhibitors. *Proc. Nat. Acad. Sci. USA* 74, 5463-5467.