

## ChromaPipe: a pipeline for analysis, quality control and management for a DNA sequencing facility

T.D. Otto<sup>1,2</sup>, E.A. Vasconcellos<sup>1,2</sup>, L.H.F. Gomes<sup>1,3</sup>, A.S. Moreira<sup>1</sup>,  
W.M. Degraive<sup>1</sup>, L. Mendonça-Lima<sup>1</sup> and M. Alves-Ferreira<sup>1</sup>

<sup>1</sup>Laboratório de Genômica Funcional e Bioinformática,  
Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, RJ, Brasil

<sup>2</sup>Fundação Ataulpho de Paiva, Rio de Janeiro, RJ, Brasil

<sup>3</sup>Faculdade de Medicina, Universidade Federal do Rio de Janeiro,  
Rio de Janeiro, RJ, Brasil

Corresponding author: T.D. Otto  
E-mail: otto@ioc.fiocruz.br

Genet. Mol. Res. 7 (3): 861-871 (2008)

Received June 2, 2008

Accepted August 11, 2008

Published September 23, 2008

**ABSTRACT.** Optimizing and monitoring the data flow in high-throughput sequencing facilities is important for data input and output, for tracking the status of results for the users of the facility, and to guarantee a good, high-quality service. In a multi-user system environment with different throughputs, each user wants to access his/her data easily, track his/her sequencing history, analyze sequences and their quality, and apply some basic post-sequencing analysis, without the necessity of installing further software. Recently, Fiocruz established such a core facility as a “technological platform”. Infrastructure includes a 48-capillary 3730 DNA Sequence Analyzer (Applied Biosystems) and supporting equipment. The service includes running samples for large-scale users, performing DNA sequencing reactions and runs for medium and small users, and participation in partial or full genome projects. We implemented a workflow that fulfills these requirements for small and high throughput users. Our implementation also includes the monitoring of data for continuous quality improvement (reports by plate, month

and user) by the sequencing staff. For the user, different analyses of the chromatograms, such as visualization of good quality regions, as well as processing, such as comparisons or assemblies, are available. So far, 180 users have made use of the service, generating 155,000 sequences, 35% of which were produced for the BCG Moreau-RJ genome project. The pipeline (named ChromaPipe for Chromatogram Pipeline) is available for download by the scientific community at the url <http://bioinfo.pdtis.fiocruz.br/ChromaPipe/>. The support for assembly is also configured as a web service: <http://bioinfo.pdtis.fiocruz.br/Assembly/>.

**Key words:** Sequencing pipeline; Chromatogram processing; DNA sequencing