Genome analysis

Seevolution: visualizing chromosome evolution

Andrés Esteban-Marcos, Aaron E. Darling∗ and Mark A. Ragan

Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD 4072, Australia

Received on August 5, 2008; revised on January 30, 2009; accepted on February 16, 2009
Advance Access publication February 20, 2009
Associate Editor: John Quackenbush

ABSTRACT
Summary: Genome evolution underpins all of biology, yet its principles can be difficult to communicate to the non-specialist. To facilitate broader understanding of genome evolution, we have designed an interactive 3D environment that enables visualization of diverse genome evolution processes. The system can intuitively and interactively animate mutation histories involving genome rearrangement, point mutation, recombination, insertion and deletion. Multiple organisms related by a phylogeny can be visualized simultaneously. As methods to infer evolutionary histories of genomes become increasingly complex, visualization of the evolutionary process will not only be useful for communication, but will also serve as an exploratory tool for discovering new patterns of genome evolution.
Availability: The software is licensed under the GNU GPL and available for download from http://seevolution.org.
Contact: aarondarling@ucdavis.edu

1 INTRODUCTION
As genomes evolve, they undergo mutational processes that can alter not only individual nucleotides but also the large-scale structure of chromosomes. Although genome sequencing has helped to characterize the rates and patterns of such chromosomal evolution, communicating findings to a broad audience can be challenging.

Modern inference methods can reconstruct likely evolutionary histories using genome sequence data alone. Methods such as GRAPPA (Tang and Moret, 2003) can infer inversion histories and phylogenies on single chromosomes, while GRIMM/MGR (Tesler, 2002) can be applied to multi-chromosomal genomes. Yet another method, BADGER (Larget et al., 2005), implements a Bayesian framework to sample likely phylogenetic inversion histories and can report the uncertainty in individual reconstructions (Darling et al., 2008). Bayesian methods have also been developed to sample histories of inversions and transposition (Miklos, 2003), duplications (Zhang, 2008), gene gain and loss and gene conversion and nucleotide substitution (Didelot and Falush, 2007). Similar methods have been applied to mammals (Blanchette et al., 2008). All of these inference methods are typically predicated on a multiple-genome alignment (Darling et al., 2004). Although no method currently infers mutation histories under a single model incorporating all these mutation types, inferences from each method can potentially be integrated into a single reconstruction.

Fig. 1. (a) Snapshots of inversion animation on a circular chromosome. (b) Heatmap view of chromosomes. Seevolution can load user-defined heatmaps that can represent, for example, GC-skew, local repeat abundance or gene expression levels. Heatmaps are simply a list of real-valued numbers in [0,1].

Output from the programs listed above usually consists of a text file containing the complete history for the mutation type and genomes of interest. The textual representation is not always easy to interpret, especially for complex histories. An appropriate visualization has the potential to substantially aid interpretation.

To this end we introduce Seevolution, a novel system for visualizing complex histories of chromosomal evolution. Seevolution displays single- and multi-species phylogenies, animating the inferred series of events that occurred in the history of the organisms. Mutations such as inversion, transposition, deletion, insertion, nucleotide substitution and gene conversion can all be visualized.

Other programs offer some visualization of rearrangements, for example, the GRIMM web server (Tesler, 2002) and PEGR (Fremez et al., 2007). However, to our knowledge no other software visualizes the range of events implemented in Seevolution.

2 METHODS
Seevolution utilizes Java3D to render animations of evolutionary events in chromosomes. Java3D implements a scene graph-based rendering paradigm. In Seevolution, a chromosome is represented visually as either a cylinder (linear chromosomes) or torus (circular chromosomes). In the scene graph, each chromosome is composed of an arbitrary number of segments, each of which is also a cylinder. Thus, a circular chromosome is a composite of many small cylindrical segments each rotated around a center point in x,y,z space. The segments are themselves composed of numerous triangular surfaces. Each segment can be assigned a color and a texture, and Seevolution uses these colors to communicate information about breakpoints of rearrangement, spatial organization of the chromosome such as distance from the origin of replication and other features (Fig. 1).

∗To whom correspondence should be addressed.

© 2009 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
We designed the XML format to represent chromosome evolution. The described viewer has been implemented as a Java Web Start application available from http://seevolution.org. The application takes as input an XML file listing the history of evolutionary events. We designed the XML format to represent chromosome evolution in the simplest possible way (Fig. 3). The following mutation types are currently supported: substitutions, inversions, gene conversions, insertions and deletions. The XML can also represent genealogical trees relating the organisms of interest. Example XML and a tutorial are available online. Future work visualizing mutation histories might include means to summarize the uncertainty inherent in most reconstructions. Substantial ambiguity exists in mutation event ordering, as orderings are often informed only by branching events in the tree.

**Funding:** This work was supported by National Science Foundation (DBI-0630765 to A.E.D.); Australian Research Council (CE0348221 to M.A.R.).

**Conflict of Interest:** None declared.