Genetics and population analysis

TESD: a transposable element dynamics simulation environment
Grégoire Decelieire, Yann Letrillard, Sandrine Charles and Christian Biémont*
Laboratoire de Biométrie et Biologie Evolutive (UMR 5558), CNRS, Université Claude Bernard Lyon 1, 69622, Villeurbanne Cedex, France

Received on April 5, 2006; revised on August 16, 2006; accepted on August 18, 2006
Advance Access publication August 23, 2006

ABSTRACT
Various mathematical models have been used to explore the dynamics of transposable elements (TEs) within their host genomes. However, numerous factors can influence their dynamics, and we know only little about the dynamics of TEs when they first began to invade populations. In addition, the influence of population structuring has only recently been investigated. Transposable Element Simulator Dynamics, a population genomics simulation environment, has therefore been developed to provide a simple tool for analyzing the dynamics of TEs in a community based on (i) various TE parameters, such as the transposition and excision rates, the recombination rate and the coefficient of selection against TE insertions; and (ii) population parameters, such as population size and migration rates. The simulations can be used to illustrate the dynamic fate of TEs in structured populations, can be extended by using more specific molecular or demographic models, and can be useful for teaching population genetics and genomics.

Availability: TESD is distributed under GPL from the Pôle Bioinformatique Lyonnais (PBIL) web server at http://pbil.univ-lyon1.fr/software/TESD
Contact: biemont@biomserv.univ-lyon1.fr

1 INTRODUCTION
Transposable elements (TEs) are moderately repeated DNA sequences, which make up a high proportion of most genomes, that are able to move along chromosomes and have a high mutator potential (Biémont and Vieira, 2005). They have played a significant role during evolutionary processes and, by promoting new genetic variability, could still have an important influence in population adaptation. Their dynamics within a genome and within a population are the subject of passionate debates and of various theoretical analyses (for a review see Le Rouzic and Decelieire, 2005). Most of the theoretical models of the dynamics of TEs involve a single population, and although we have obtained evidence on the recent species invasion of the genome of the fruit fly Drosophila melanogaster in the past century by some elements (Biémont and Vieira, 2005), there are few models describing the steps that lead to the invasion of an entire species, apart from the \textit{P} element of \textit{D.melanogaster} (Queasnville and Anxolabéhére, 1998). Recent mathematical developments have made it possible to track the impact of horizontal transfers of TEs in a single population (Le Rouzic and Capy, 2005) and, thereby include the migration between populations of the host species (Decelieire \textit{et al}., 2005). Transposable Element Simulator Dynamics (TESD) is a user-friendly application of the individual-based model of TEs to this inter-population process.

2 METHODS
The simulation developed by Decelieire \textit{et al}.
(2005) is based on the classical model of selection against TE insertions for diploid monocious individuals proposed by Charlesworth and Charlesworth (1983), in which migration between subpopulations of a spatially structured population has been added. The simulations used the stepping stone model of a single species structured into populations of finite size (Kimura \textit{et al}., 1964). TESD, thus, takes into account both molecular processes (TE transposition and excision rates, homologous recombination between TE insertion sites, selection against TE insertions that can be deleterious or involved in ectopic recombination) and demographic processes (population size and rates of migration between populations). TESD is implemented by using the Java 2 Standard Edition specifications (version 5.0) and runs on all OS that have an interpreter with J2SE specifications. The main structure is based on aggregated objects, population, individual, chromosome, species and TE insertions, thus making it possible to add modules by inheritance in order to explore new molecular/demographic models.

TESD can be used in either shell mode or graphic user interface (GUI) mode. Shell mode is recommended for intensive computations. This enables the software to carry out several simulations simultaneously by using the same initial populations. In GUI mode, which can also be started from the shell, TESD has a graphics package to plot the data using curves, histograms and maps. These tools were developed using the JFreeChart libraries (version 1.0.0; http://www.jfree.org/jfreechart) based on the Java2D package, under GNU Lesser GPL and javaHelp under SUN license, version 2 (http://java.sun.com/products/javahelp).

3 RESULTS
TESD is a simulation environment and it can be used to investigate the dynamics of the spread of TEs in a host population that is itself structured into connected subpopulations. In the case of a single population, TESD illustrates the dynamics within a genome of a newly invading TE as in the classical models of Charlesworth and Charlesworth (1983), Tsitrone \textit{et al}.
(1999) and Le Rouzic and Capy (2005). TESD differs from analytical approaches, in that in addition to equilibrium values, the first generations of the TE invasion can be tracked step by step, and the extinction of the TE during this spread can be observed. This means that TESD can be used to estimate the biological impacts of parameters controlling TE dynamics and the influence of the connections (the migration rates) between the subpopulations. TESD cannot be used, however, to estimate parameters from the data of TE dynamics. TESD, which can be applied to any sexual, outbred species, is therefore also useful for illustrating the influence of population structuring, population
et al.

numbers in a population are usually not in equilibrium (Tsitrone et al., 1996) and drift for teaching population genetics and genomics. TESD can be extended to more specific molecular or demographic models which could include presence of active, autonomous and nonautonomous elements, transposition repressors, host factors (Badge and Brookfield, 1997), burst of transposition in some populations (Biémont et al., 1996, 1997), suggesting that fluctuations in TE numbers in the populations after around 800 generations. In Figure 1B, in which the population size was 500 individuals and the migration rate was equal to 0.005, the TE copy number in the population with the highest initial copy number decreases, whereas the copy number in the population with an initially low copy number increases.

A plateau is then reached at around the 100th generation. Although the plateau could be interpreted as an equilibrium value, the copy number subsequently begins to increase at the 200th generation. This is due to mixing of the populations, and the homogenization of their TE copy numbers owing to migration. The copy number, thus, behaves in the same way as in a single genome, and reaches an equilibrium value after numerous generations, as observed classically (Charlesworth and Charlesworth, 1983; Tsitrone et al., 1999). Changing the parameter values in the simulation can lead to various distributions of TEs that have been reported classically (Charlesworth and Charlesworth, 1983; Tsitrone et al., 1996) and laboratory stocks (Nuzhdin et al., 1998, 1997), suggesting that fluctuations in TE numbers in the first generations may be quite considerable and that the TE copy numbers in a population are usually not in equilibrium (Tsitrone et al., 1999).

ACKNOWLEDGEMENTS

This work was funded by the Centre National de la Recherche Scientifique (UMR 5558) and the Bioinformatics programme of the French Minister of Research.

Conflict of Interest: none declared.

REFERENCES


