BIOINFORMATICS

Vol. 00 no. 00 2011
Pages 1–3

An individual-based model simulator for transposable elements dynamics and evolution

Felipe Figueiredo [1,2]*, Claudio Struchiner [2]

[1]Programa de Ps-graduao em Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, BCS/IOC/Fiocruz
[2]Programa de Computao Cientfica, PROCC/Fiocruz

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation: Transposable Elements (TEs) or Mobile Elements are small genomic elements present in almost all genomes sequenced so far, that appear repeatedly in the individuals genome. Much debate is ongoing about their origin, but some mechanisms for the invasion of a host individual and the later spread within a host population are known from both *in silico* and wet laboratory experiments.

In order to take into account the forces that promote variation in TEs three independent levels of biological organization must be evaluated: (a) the population demographics should be represented by a proper ecological model, (b) the population genetics of the TEs should be given by a proper transposition model and (c) the molecular evolution of actual TE sequences must be represented by a evolutionary model. Previous efforts at modelling TE dynamics focus mostly on the net quantity of TEs, either per individual or the total ammount in the population. Models that deal with the molecular evolution of the TEs are scarce. Most mathematical or computational models deal with either one or two of the levels of organization above.

Results: We present TRepid, an individual-based model that can be used to simulate Population Genetics of a host population with a focus in Transposable Elements dynamics within a host population. This system can be used for simulating TE invasions, fixation and competition between TE families.

TEs dynamics features presently available in our simulator include: transposition by a selectable method (copy and paste or cut and paste), selectable transposition models, nucleotide substitution according to a selectable molecular evolutionary model, active and inactive TEs, inactivation of TE over generations, selective pressure on host individuals by accumulation of TEs and deleterious transposition events, and more. Population dynamics features include random mating, selectable population models (including exponential and logistic) and recombination by crossing over of gametes.

Availability: The software is available upon request to the author, either as a packaged Perl module, or as a Debian/Ubuntu .DEB package.

Contact: ffigueiredo@ioc.fiocruz.br

1 INTRODUCTION

Transposable elements are DNA segments that appear in several genomic sites in a host. They occur in several sizes and structures, ranging from small segments flanked by inverted repeats that get copied whenever there's an apropriate enzyme, to autonomous virus-like elements that contain the necessary genes to produce its own transposase or retrotransposase.

They are so common, in fact, that they have been observed in such quantities corresponding to 40-50% of mammals genomes, and almost 90% in some plants (Kazazian, 2004), and seem to be present inside every genome they have been looked for so far (Batzer and Deininger, 2002). This is why these elements previously thought to be *junk DNA* have now been promoted to agents of evolutionary force (Biemont *et al.*, 2003; Le Rouzic and Capy, 2005; Gotea and Makalowski, 2006). A throughout classification work of TEs is available in Wicker (2007).

Several approaches have been proposed to understand and predict how the amount of TEs vary in hosts genomes (Le Rouzic and Deceliere, 2005). Models try to assess the invasion capabilities, accumulation and fixation of new copies and are usually based on mathematical and computational techniques.

FIXME: citar (Hedges *et al.*, 2005).

Most of the previous attempts in the modelling community focus in reproducing the dynamics through differential or difference equations that expressed how the total amount of TE copies vary in terms of acquisition of new copies and excision of old copies (Quesneville and Anxolabehere, 1997, 1998; Deceliere *et al.*, 2005, 2006), although some explore less popular techniques such as genetic algorithms (Quesneville and Anxolabehere, 2001), (FIXME)(Egorova *et al.*, 2003) or even combination of more than one technique (Katzourakis *et al.*, 2005). These types of models can typically be characterized in a continuum between two extreme paradigms: the *master gene* type of model, in which only one TE copy is active and generates inactive copies, and the *transposon* type of model in which every TE actively bears copies, leading to an exponential growth of TE copy number in the absence of some regulation mechanism.

*to whom correspondence should be addressed

2 METHODS

Throughout the computational model description, we will refer to transposable elements as *TEs*, and the individuals that carry them *Hosts*. This is also the nomenclature of the objects involved in the software.

The simulator takes into consideration distinct and independent modeling techniques for each level of biological organization: a population model, a transposition model and an evolutionary model of the DNA sequences. Each of these sections are based on either differential or difference equations, or probabilistic models.

The population model produces a trend for the population dynamics. The basic constant population model and the logistic equation are available, as well as the Hassell equation(Hassell, 1975). In each generation, the necessary ammount of individuals is sampled and coupled generate the required ammount of offspring to follow the ecological model as closely as possible, notwithstanding fitness effects from TEs. The modular framework provide the means to implement any ecological model.

The transposition model does the same thing for the ammount of TEs in each new individual, based on the ammount of TEs in the parental gametes.

The evolutionary model determines how the TE sequences change over time, after sucessive transposition events (Yang, 2006).

2.1 The generation algorithm

The general algorithm that happens at each generation models the basic life cycle of a diploid sexual species subject to transposition events in the gametogenesis.

1. Host couples are chosen randomly from available mature hosts at the beginning of each reproductive season. Males are chosen with replacement, females are chosen without replacement.
2. Each adult bears new gametes after transposition and recombination.
3. Transposition draws a recruitment amount of new TE copies and the deletion amount of excised copies from the transposition model. This changes the content of the gametes in terms of availability of TEs.
4. Mutations are sampled from the evolutionary model for newly created TE copies.
5. Recombination provides additional shuffling of gamete contents.
6. Each couple gives birth to a number of offspring defined by the user as a parameter.
7. Each newborn individual is composed of one chromosome from each of its parents, and new individuals are introduced to the population pool.
8. The fitness cost from TEs in newborn individuals is calculated and any that exceeds a given threshold is killed before birth and removed from the population.
9. The age of every surviving individuals is incremented ate the end of the generation.

3 CONCLUSION

In this paper we describe a computational model composed of a forward-time individual-based model for the population genetics of transposable elements.

Similar software exist in the for each of the techniques we combine. Population genetics software exist for both forward simulations (Carvajal-Rodriguez, 2008; Guillaume and Rougemont, 2006; Peng and Kimmel, 2005; Padhukasahasram *et al.*, 2008; Hernandez, 2008) and backward-time simulations (Hudson, 2002; Teshima and Innan, 2009), although most of them simply count the distribution and availability of a set of alelles that populate a given *locus* or *loci*. Similarly, there are simulators for transposition phenomena in host populations (Deceliere *et al.*, 2006) but assumes TEs don't change over time. Evolutionary model aware software (Hwang and Green, 2004) are more scarce, and we are aware no one system that deals with all the effect of three levels of biological organization.

We also consider the effect of sampling a host at the end of the simulation. This provides an additional layer in the model, that takes into account a sampling distribution. (blah, FIXME)

ACKNOWLEDGEMENT

Funding: This work was partially supported by a Bill & Melinda Gates Foundation grant (FIXME), and a PhD scholarship by CAPES (FIXME).

REFERENCES

Batzer, M. A. and Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nat Rev Genet*, **3**(5), 370–9.

Biemont, C., Nardon, C., Deceliere, G., Lepetit, D., Loevenbruck, C., and Vieira, C. (2003). Worldwide distribution of transposable element copy number in natural populations of Drosophila simulans. *Evolution*, **57**(1), 159–67.

Carvajal-Rodriguez, A. (2008). GENOMEPOP: a program to simulate genomes in populations. *BMC Bioinformatics*, **9**, 223.

Deceliere, G., Charles, S., and Biemont, C. (2005). The dynamics of transposable elements in structured populations. *Genetics*, **169**(1), 467–74.

Deceliere, G., Letrillard, Y., Charles, S., and Biemont, C. (2006). TESD: a transposable element dynamics simulation environment. *Bioinformatics*, **22**(21), 2702–3.

Egorova, A. V., Ratner, V. A., and Iudanin AIa (2003). [Negative selection and computer models of the joint evolution of the patterns of polygenes, transposable elements, and origin identity labels]. *Genetika*, **39**(4), 540–9.

Gotea, V. and Makalowski, W. (2006). Do transposable elements really contribute to proteomes? *Trends Genet*, **22**(5), 260–7.

Guillaume, F. and Rougemont, J. (2006). Nemo: an evolutionary and population genetics programming framework. *Bioinformatics*, **22**(20), 2556–7.

Hassell, M. P. (1975). Density-dependence in single-species populations. *Journal of Animal Ecology*, **44**(1), pp. 283–295.

Hedges, D. J., Cordaux, R., Xing, J., Witherspoon, D. J., Rogers, A. R., Jorde, L. B., and Batzer, M. A. (2005). Modeling the amplification dynamics of human Alu retrotransposons. *PLoS Comput Biol*, **1**(4), e44.

Hernandez, R. D. (2008). A flexible forward simulator for populations subject to selection and demography. *Bioinformatics*, **24**(23), 2786–7.

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**(2), 337–8.

Hwang, D. G. and Green, P. (2004). Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc Natl Acad Sci U S A*, **101**(39), 13994–4001.

Katzourakis, A., Rambaut, A., and Pybus, O. G. (2005). The evolutionary dynamics of endogenous retroviruses. *Trends Microbiol*, **13**(10), 463–8.

Kazazian, Jr, H. H. (2004). Mobile elements: drivers of genome evolution. *Science*, **303**(5664), 1626–32.

Le Rouzic, A. and Capy, P. (2005). The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. *Genetics*, **169**(2), 1033–43.

Le Rouzic, A. and Deceliere, G. (2005). Models of the population genetics of transposable elements. *Genet Res*, **85**(3), 171–81.

Padhukasahasram, B., Marjoram, P., Wall, J. D., Bustamante, C. D., and Nordborg, M. (2008). Exploring population genetic models with recombination using efficient forward-time simulations. *Genetics*, **178**(4), 2417–27.

Peng, B. and Kimmel, M. (2005). simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, **21**(18), 3686–7.

Quesneville, H. and Anxolabehere, D. (1997). A simulation of P element horizontal transfer in Drosophila. *Genetica*, **100**(1-3), 295–307.

Quesneville, H. and Anxolabehere, D. (1998). Dynamics of transposable elements in metapopulations: a model of P element invasion in Drosophila. *Theor Popul Biol*, **54**(2), 175–93.

Quesneville, H. and Anxolabehere, D. (2001). Genetic algorithm-based model of evolutionary dynamics of class II transposable elements. *J Theor Biol*, **213**(1), 21–30.

Teshima, K. M. and Innan, H. (2009). mbs: modifying Hudson's ms software to generate samples of DNA sequences with a biallelic site under selection. *BMC Bioinformatics*, **10**, 166.

Wicker, T. (2007). A unified classification system for eukaryotic transposable elements. *Nature*, **8**.

Yang, Z. (2006). *Computational Molecular Evolution (Oxford Series in Ecology and Evolution)*. Oxford University Press, USA.