BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 9 2011, pages 1335–1336
doi:10.1093/bioinformatics/btr139

Genetics and population analysis

Advance Access publication March 16, 2011

ogaraK: a population genetics simulator for malaria

Tiago Antao* and Ian M. Hastings

Department of Molecular and Biochemical Parasitology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

Associate Editor: Jeffrey Barrett

ABSTRACT

Motivation: The evolution of resistance in *Plasmodium falciparum* malaria against most available treatments is a major global health threat. Population genetics approaches are commonly used to model the spread of drug resistance. Due to uncommon features in malaria biology, existing forward-time population genetics simulators cannot suitably model *Plasmodium falciparum* malaria.

Results: Here we present ogaraK, a population genetics simulator for modelling the spread of drug-resistant malaria. OgaraK is designed to make malaria simulation computationally tractable as it models infections, not individual parasites. OgaraK is also able to model the life cycle of the parasite which includes both haploid and diploid phases and sexual and asexual reproduction. We also allow for the simulation of different inbreeding levels, an important difference between high and low transmission areas and a fundamental factor influencing the outcome of strategies to control or eliminate malaria.

Availability: OgaraK is available as free software (GPL) from the address http://popgen.eu/soft/ogaraK.

Contact: tra@popgen.eu

Supplementary information: Supplementary data is available at *Bioinformatics* online.

Received on December 19, 2010; revised on February 18, 2011; accepted on March 9, 2011

1 INTRODUCTION

Malaria is a major public health concern, as one-third of the human population is estimated to be exposed to the threat of the most virulent species, *Plasmodium falciparum*. Antimalarial drug resistance has emerged as one of the major challenges facing malaria control. Drug resistance became widespread to most first-line therapies and treatment failures are now being observed for their replacements, Artemisinin Combination Therapies (ACTs) (Dondorp *el al*., 2009). Mathematical and computational models of the spread of drug resistance are an important tool to understand the emergence and spread of drug resistance.

Most mathematical and computational modelling of malaria have been based on epidemiology (e.g. Koella and Antia, 2003) or population genetics (e.g. Hastings, 1997), though complex simulation models have also been developed (Smith *et al*., 2008). While many forward-time population genetics simulators do exist (e.g. Peng and Kimmel, 2005), they are not suitable to model malaria, therefore all existing computational studies using a population genetics approach are based on applications and scripts developed for each study and not directly subjected to peer review or publicly available.

Standard individual-based forward-time population genetics simulators are not suitable to model malaria biology for two main reasons: (i) population size in malaria can rise up to 10^{12} parasites per human host, making it computationally infeasible to simulate all individuals and (ii) the malaria life cycle includes both haploid and diploid phases and most existing simulators do not allow for the modelling of different genotypic structures over time. In order to address these issues, we developed ogaraK, a population genetics simulator designed to study the spread of drug resistance in *P.falciparum* malaria.

2 APPROACH

OgaraK features and limitations are based on malaria population biology in the presence of treatment pressure. OgaraK was designed to study the spread of existing drug resistance and it can be used to understand how recognized important factors in malaria epidemiology (e.g. different transmission intensities) influence the spread of resistance and also to compare different drug deployment policies.

Drug treatments can be modelled as a form of selection pressure over the parasite population. OgaraK is able to simulate a wide variety of selection pressures modelling both temporal and spacial selection heterogeneity. Temporal heterogeneity is a proxy for a policy of drug rotation while spacial heterogeneity approximates the use of multiple first-line drugs.

A wide range of epistasis modes of loci involved in drug resistance are also supported. While most existing theoretical research assumes that parasites require all mutations (full epistasis) related to a drug in order to resist treatment, ogaraK is able to model other epistasis modes, like duplicate gene function or asymmetry. These modes reflect existing empirical evidence for some drugs (e.g. Chloroquine or SP) where genes vary in their importance to confer resistance (Olliaro, 2005). Multiple epistasis modes are also useful to model poor drug compliance or host immunity: a weaker epistasis mode among drug resistance loci is enough to confer resistance in humans with no acquired immunity or who take an incomplete treatment; in humans who take a full course or have acquired partial immunity against malaria, a stronger epistasis mode is required to resist treatment.

Multiplicity of infection (MOI) has been recognized as one of the factors that differentiate between high and low transmission areas of malaria. MOI affects the spread of resistance, recombination and population inbreeding levels as the mating alternatives on the obligatory sexual phase in the mosquito are limited by the number

*To whom correspondence should be addressed.

of genetically different parasites ingested in a blood meal. OgaraK allows the simulation of different inbreeding levels, allowing for a varying MOI across simulations and also, within each simulation, simulating individuals with different MOI.

Individual-based simulation is replaced by exhaustive enumeration of all possible combinations of infection types (each infection having a specific genotype). This method was first used in Hastings (2006) and comparative analysis between results derived from epidemiological simulations and this approach show consistent results (Boni *et al*., 2008).

The main purpose of the simulator is to study the spread of resistance but we also support mutation therefore allowing the study of *de novo* emergence. In the case of most antimalarials [e.g. Chloroquine or SP (Wellems and Plowe, 2001)], mutation is a rare event. Furthermore, resistance already exists to most drugs, even Artemisinin-based therapies (Dondorp *el al*., 2009), hence our focus is on spread of existing mutations.

In order to study simultaneous usage of multiple ACTs which might share one resistance gene (as they all have an Arteminisin derivative as principal component), we also provide models of multi-drug resistance where part of the resistance mechanism is shared among all drugs. A 'standard' model where all drugs involved have unrelated loci is also available.

OgaraK has an easy to use interface which can be run from the web as a Java Webstart application. The code is available and can be also linked as a library or used in batch mode.

Results are exported in a text format (reporting the frequency of all genotypes over time) and also made available in the widely used Genepop (Rousset, 2008) format. Simple scripts for data analysis are provided using Biopython (Cock *et al*., 2009), but these mainly serve as examples as it is expected that most analysis will be done using standard population genetics packages owing to the ability to export data in Genepop format.

Supplementary material is included where the model is detailed and where example applications and a user manual are also supplied.

3 DISCUSSION

OgaraK is able to easily simulate most existing population genetics models studying the spread of drug resistance in malaria. It is made available as a public framework which can be used to evaluate new models or re-use old ones for new analysis. For instance, it was already used and tested to research the impact of epistasis on linkage disequilibrium between loci involved in drug resistance (Antao and Hastings, 2011) and can also simulate, from a population genetics perspective, promising drug deployment strategies (Boni *et al*., 2008).

We focused on modelling how resistance spreads and not *de novo* emergence, given that emergence is a rare event and that it is already widespread to most drugs, therefore making the management of existing resistance a major concern. Nonetheless ogaraK supports mutation, therefore permitting the study of *de novo* emergence.

While ogaraK was developed with malaria modelling in mind, epistasis and spacial and temporal selection patterns have clear parallels with some known models in sex theory (Otto, 2009), as temporal selection heterogeneity is the fundamental concept behind the Red-Queen Hypothesis, and spacial selection heterogeneity has also been proposed as one explanation for sex and recombination. OgaraK can, therefore, be used to easily simulate and test some models relevant to sex theory.

OgaraK can serve as a framework to more easily evaluate drug deployment policies and help enhance the understanding of fundamental variables underlying the spread of drug-resistant malaria.

Funding: Research grants SFRH/BD/30834/2006 and PTDC/BIA-BDE/65625/2006 from Fundação para a Ciência e Tecnologia, Portugal to T.A.

Conflict of Interest: none declared.

REFERENCES

Antao,T. and Hastings,I.M (2011) Environmental, pharmacological and genetic influences on the spread of drug resistant malaria. *Proc. R. Soc. B*, in press, [Epub ahead of print, doi:10.1098/rspb.2010.1907].

Boni,M.F. *et al*. (2008) Benefits of using multiple first-line therapies against malaria. *Proc. Natl Acad. Sci. USA*, **105**, 14216–14221.

Cock,P.J.A. *et al*. (2009) Biopython: freely available Python tools for computational molecular biology and informatics. *Bioinformatics*, **25**, 1422–1423.

Dondorp,A.M. *et al*. (2009) Artemisinin resistance in Plasmodium Falciparum malaria. *N Eng. J. Med.*, **361**, 455–467.

Hastings,I.M. (1997) A model for the origin of drug-resistant malaria. *Parasitology*, **115**, 133–141.

Hastings,I.M. (2006) Complex dynamics and stability of resistance to antimalarial drugs. *Parasitology*, **132**, 615–624.

Koella,J.C. and Antia,R. (2003) Epidemiological models for the spread of anti-malarial resistance. *Malaria J.*, **2**, 3.

Olliaro,P. (2005) Drug resistance hampers our capacity to roll back malaria. *Clin. Infect. Dis.*, **41** (Suppl. 4), 247–257.

Otto,S.P. (2009) The evolutionary enigma of sex. *Am. Natural.*, **174**, S1–S14.

Peng,B. and Kimmel,M. (2005) simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, **21**, 3686–3687.

Rousset,F. (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Res.*, **8**, 103–106.

Smith,T. *et al*. (2008) Towards a comprehensive simulation model of malaria epidemiology and control. *Parasitology*, **135** 1507–1516.

Wellems,T.E. and Plowe,C.V. (2001) Chloroquine-resistant malaria. *J. Infect. Dis.*, **184**, 770–776.