University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Krzysztof Kamil Gogolewski
Student number: 291538

Computational modeling of stress-induced proliferation of transposable elements in sexual populations

Master's Thesis
in COMPUTER SCIENCE

Supervisor
dr hab. Anna Gambin, prof. UW
Institute of Informatics

September 2014

Oświadczenie kierującego pracą

Potwierdzam, że niniejsza praca została przygotowana pod moim kierunkiem i kwalifikuje się do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data Podpis kierującego pracą

Oświadczenie autora (autorów) pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data Podpis autora (autorów) pracy

Streszczenie

Pomimo tego, że elementy transpozonowe zostały odkryte już w połowie XX w. i wiadomym jest, że stanowią jeden z czynników wpływających na ewolucję genotypu organizmów, to ich dokładna rola wciąż nie jest dokładnie poznana. Niemniej jednak, udowodniono, że mutacje powstające na skutek występowania zdarzeń transpozycji mogą prowadzić do znaczących zmian genetycznych i w naturalny sposób wpłwać na fenotyp organizmu. Zważywszy jednak na fakt, że mutacje mogą okazywać się korzystne, nasuwającym się pytaniem jest, czy w pewnych warunkach aktywność elementów transpozonowych może mieć pozytywny wpływ na ewolucję całych populacji.
W poniższej pracy, podejmujemy próbę modelowania ewolucji diploidalnych populacji płciowych, w których istnieją aktywne rodziny transpozonów. Głównym celem jest zbadanie związków pomiędzy oddziaływaniem stresu środowiskowego na populację a aktywnością elementów transpozonowych, które są obecne na genomach organizmów składających się na rozważaną populację. Analiza stochastycznego modelu obliczeniowego pozwala wnioskować, że w obecności postępującego stresu środowiskowego aktywność elementów transpozonowych, a dokładniej mutacje, które są przez nią mediowane mogą istotnie poprawić efektywność procesu dostosowywania się organizmów do zmiennych warunków i chronią całą populację przed wyginięciem.

Słowa kluczowe

Symulacje stochastyczne, Modelowanie matematyczne, Elementy transpozonowe, Adaptacja, Geometryczny model Fishera, Stres środowiskowy.

Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.3 Informatyka

Klasyfikacja tematyczna

- Informatyka Stosowana
- Nauki przyrodnicze i medyczne
- Genetyka
- Genetyka populacji

Tytuł pracy w języku polskim

Obliczeniowe modelowanie proliferacji elementów transpozonowych indukowanej stresem środowiskowym w populacjach płciowych

Abstract

Although transposable elements were discovered in the middle of the 20^{th} century and are known as one of the factors that drive the evolution of the genomic content of organisms, their exact role has not been fully elucidated yet. However, it was so far shown that mutations induced by transposition events can lead to significant genetic disorders and naturally affect organism's phenotype. Furthermore, since every mutation can also be beneficial, the natural question that arises is whether under certain conditions the activity of TEs can be considered as an evolutionary helper i.e. a mechanism that has positive contribution to the evolution of a population.
In the following thesis, we take a closer look at the evolution of sexual diploid populations which are hosts for active TE families. The purpose is to explore the relationship between the environmental stress, that influences such population and activity of those TEs that are present in genomes composing the population in question. Basing on results obtained from the stochastic computational model we conclude, that in the presence of progressive environmental stress the activity of TEs, and specifically mutations that are mediated by their occurrence, can noticeably improve the process of adaptation to varying conditions and prevent the population from extinction.

Keywords

Stochastic simulations, Mathematical modeling, Transposable elements, Adaptation, Fisher's geometric model, Environmental pressure.

Subject area (codes from Socrates-Erasmus)

11.3 Informatics, Computer Science

Topic classification

- Applied Computing
- Life and medical sciences
- Genetics
- Population genetics

Contents

Chapter 1

Introduction

Evolution and environmental stress. It is commonly known that all species are continuously evolving, and the evolution process itself depends on numbers of factors, such as strength of natural selection, genetic drift, random mutations or recombination (in case of sexual populations). Moreover, populations are constantly exposed to environmental changes (also called environmental stress or pressure) of a various type and intensity which results in adaptation to new conditions or full extinction. As an simple example of changing environmental conditions, one can consider the brood parasitic common cuckoo population that because of desynchronization of the reproductive and migratory cycles with their host species on which they parasite has decreased in size by 6% since 1980 [Antonov et al., 2010].

In general, weak environmental changes are usually overcome by populations thanks to standing genetic variation, while strong changes require more mutation events to happen in a relatively short period of time [Barrett and Schluter, 2008]. Additionally, the size of these mutational effects increases proportionally to the organism's complexity [Miller et al., 1992]. Nevertheless, the number of mutations that support a population in adjusting to new conditions cannot be infinitely large, thus a threshold that describes an upper bound of the number of allowable changes is expected to exist. Even though exact mechanisms controlling the evolvability (i.e. the capacity of the population of individuals for adaptive evolution) are still not well investigated [Partridge and Barton, 2000, Pigliucci, 2008], it is suggested by both theoretical and empirical research that the existence of mutation enhancers can be a part of the adaptive evolution [Taddei et al., 1997]. Since transposable elements are one of the major components of genomes and their nature is mutagenic, their candidacy for one of the enhancers is natural.

Transposable elements. Transposable elements (TEs) are mobile DNA sequences that were discovered in the middle of the 20^{th} century by Barbara McClintock, who conducted the research on maize grains. Specifically, she discovered that few different coloured grains that could have been observed on the maize are result of the mutations driven by the transposition events on its genome [McClintock, 1956]. Although presence of TEs was confirmed by many sequencing programs and the classification with respect to transposition mechanism was derived (see Figure 1.1), their behaviour, dynamics, distribution of copy number in organisms and other general properties are still under investigation [Charlesworth et al., 1994, Le Rouzic and Deceliere, 2005].

However, it is already known that all TEs can undergo the transposition event of two kinds: *copy-paste* and *cut-paste*, both leading to rearrangement on the host's genome and causing mutations changing the host's phenotype. Naturally, these changes can have various nature.

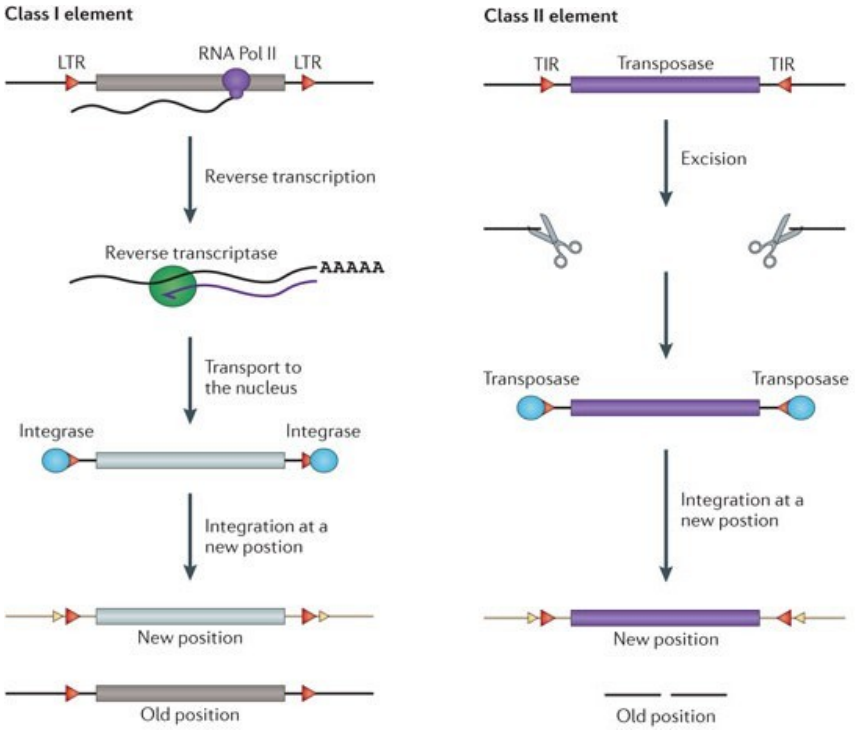


Figure 1.1: TE classification with respect to transposition mechanism: Class I elements (retrotransposons) transpose according to the *copy-paste* scheme. First step is the transcription of the DNA sequence coding the TE onto the RNA sequence. After that, in the presence of the reverse transcriptase enzyme the RNA sequence is back transcribed to the DNA and integrated on the genome. Class II elements (DNA transposons) transpose according to the *cut-paste* scheme. Transposase enzymes are responsible for the excision of the DNA sequence and shifting it to different location. Source: [Lisch, 2013]

It has been shown that the majority of these mutations have deleterious nature [Mackay, 1986] tending to disrupt useful genes [Tachida and Iizuka, 1993]. Nevertheless, the occurrence of adaptive effects of TE-driven mutations with signs of positive selection is observable [Schlenke and Begun, 2004] in the form of insertions, deletions, recombinations, as well as introduction of a new gene or a part of chimeric protein from a TE sequence [Sinzelle et al., 2009]. Even though are mostly considered as *genomic parasites* in sexual populations [Charlesworth and Charlesworth, 1983, Doolittle and Sapienza, 1980, Hickey, 1982], TEs with their, so far explored, nature may be considered as a proper candidate for the evolutionary helpers, in the presence of the environmental pressure. In order to support this theory, it is natural to introduce a computational or analytical stochastic model.

Existing computational models. So far, both computational and analytical models for sexual and asexual populations were proposed and all were meant to describe the complex dynamics of TEs activity. However, some of them are derived on biased assumptions about the dynamics of TEs (i.e. the transposition rate is explicit decreasing function of copy number of TEs [Charlesworth and Charlesworth, 1983] or TE-driven mutations have significantly larger phenotypic effects than background mutations, which is not supported by empirical results [Stoebel and Dorman, 2010]), thus conclusions based on such models ought to be again verified. Nonetheless, the common factor that bonds together all of these approaches is the Fisher's Geometric Model [Fisher, 1930], which is the method used for modeling the mechanism of spontaneous mutations and will be presented more precisely in the chapter 3. The

final model presented in this thesis is a natural extension of the model proposed in [Startek et al., 2013] which considers asexual populations exposed to the environmental pressure and gives theoretical proof that TEs can improve the process of adaptation in varying conditions. Similarly, the main purpose of this thesis and the derived model was to investigate the interplay between TEs activity and the environmental stress. Moreover, we wanted to determine if under certain conditions there exists a transposition-selection equilibrium (TSE) in the sexual populations in question. Here, the TSE should be understood as an equilibrium between increase of TEs copy number in population generated by transposition events and decrease caused by natural selection.

Chapter 2

Theoretical background

In order to start the description of the derived model, we need to recall some notions from the population genetics such as Fisher's Geometric Model (FGM).

2.1. Fisher's Geometric Model

One of the first attempts to describe the effects of random mutations that occur in an organism's genome using mathematical formalism was introduced by Ronald Fisher in 1930 [Fisher, 1930]. The idea was to express an individual as a single point $\theta \in \mathbb{R}^n$, where each coordinate (trait) is the representation of a one phenotypic character such as body size, beak length, petal length, etc. To move on the general definition of fitness needs to be introduced [Orr, 2009]:

Definition 2.1.1 (Fitness)**.** *In the theory of evolution, fitness function (or simply fitness) is the abstract measure of an organism's adaptation to living under certain environmental conditions and can be considered as a description of the ability to both survival and reproduction.*

Next, considering some fixed environmental conditions we assume the existence of the, so called, *phenotypic optimum* $\theta_0 \in \mathbb{R}^n$, meaning such point from the phenotypic space which maximizes the environmental fitness function $\omega : \mathbb{R}^n \to \mathbb{R}$, that is:

$$\theta_0 := \sup_{\theta \in \mathbb{R}^n} \omega(\theta)$$

For simplicity it is assumed that the phenotypic optimum is represented by the origin of the coordinate system in that space.

For such defined environment inhabited by a population of individuals represented by their phenotype the FGM assumes that a single mutation occurring in an organism's genome with the phenotype θ is represented as a vector $\upsilon \in \mathbb{R}^n$ and replaces the former phenotype with the value $\theta + \upsilon$. Furthermore, mutations can have different phenotypic size (i.e. vectors representing mutations can vary in length) and by definition affect every trait of an individual (the phenomenon of the, so-called, universal pleiotropy). Moreover, a mutation is beneficial if it brings the mutant phenotype closer to a nearby (local) optimum (i.e. increases its fitness value), otherwise is called deleterious. For the 2-dimensional example see Figure 2.1.

The most important observation proven by Fisher is the following:

Corollary. *The probability that a random mutation of a given phenotypic size $|\upsilon|$ is beneficial is equal to $1 - \Phi(x)$, where Φ is the cumulative distribution function of a standard normal*

random variable and $x = \frac{|\upsilon|\sqrt{n}}{2z}$ *is a standardized mutational size, where* n *is the number of dimensions of the phenotypic space and* z *is the distance to the optimum.*

Figure 2.2 presents dependency between mutational size and probability of being beneficial.

Although presented formalization is rather simple, it has yielded several robust predictions supported by empirical evidence. In [Orr, 1998] it was proved that the distribution of adaptive substitutions is approximately exponential, meaning the majority of fixed mutations is of a small size and just a few are those large once. Moreover, fixed mutational effects were proven to become on average smaller with increasing organismic complexity [Orr, 2000]. That is why, so far, FGM is probably the best tool for modeling the behavior of random mutations in a given environment.

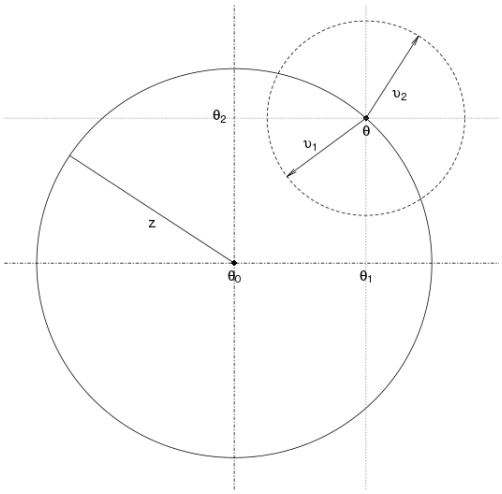


Figure 2.1: 2-dimensional example of the Fisher's Geometric Model: Here, we consider environment in which the phenotypic space is described by two traits and the phenotypic optimum is placed at the point $\theta_0 = (0, 0)$. An individual has the phenotype $\theta = (\theta_1, \theta_2)$, distance from the optimum equal to z and produces mutation of fixed magnitude in random direction. Any mutation that shifts the mutant phenotype inside the large circle is considered beneficial for the individual (υ_1), while others are deleterious (υ_2).

Nevertheless, there are some studies investigating modifications of the pure FGM, e.g. full pleiotropy is considered (i.e. situations where the number of traits that is affected by an occurring mutation event varies) [Matuszewski et al., 2014] or instead of constant phenotypic optimum the moving one is introduced [Waxman and Peck, 1999, Bürger and Gimelfarb, 2002, Nunney, 2003, Matuszewski et al., 2014].

Combining presented theory with the randomness of TEs behaviour it seems reasonable to use the FGM while modeling their activity in context of changing environment, i.e. moving phenotypic optimum.

2.2. Analytical modeling methods

Since the most common approach to modeling adaptation process in terms of fitness and spontaneous mutations was presented, we can move on to the broader revision of existing analytical models that are based on the FGM. Despite the fact, that further on we are not going to focus on derivation of any formal analytical model, some examples will be introduced in order to get the feeling of how, in general, they are constructed.

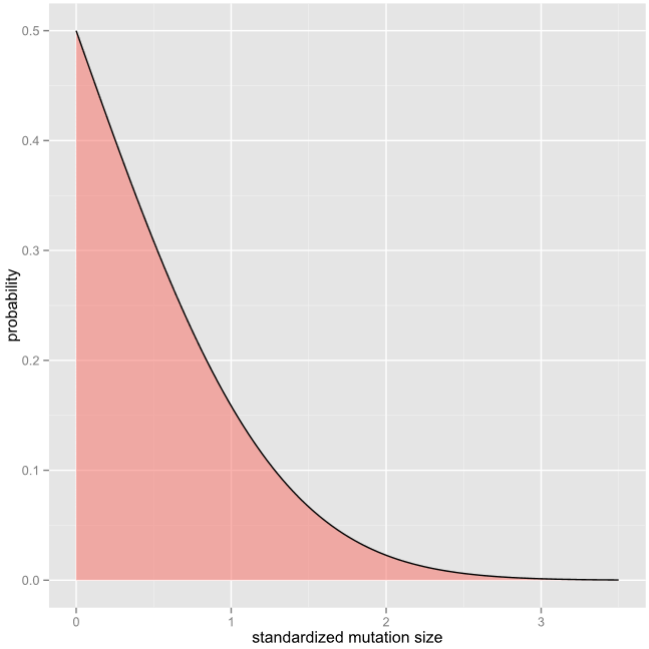


Figure 2.2: Probability of beneficial mutation with respect to its standarized size: Fisher has shown that infinitesimally small mutations have probability 0.5 of being beneficial but this probability falls rapidly for increasing mutational change.

2.2.1. Asexual populations

Here, we present a short description of an analytical model considering the activity of TEs in an asexual population with a moving optimum from [Startek, 2014]. The most important element of the following formalization, in context of this thesis, is the straight-forward usage of the FGM when modeling the effects of mutations in the sense of the population's phenotype.

First, we assume that the population is described with a density ρ on the $\mathbb{R}^n \times \mathbb{N}$ space (see Figure 2.3), and the life cycle of the population is represented by the operator Φ, that evaluated on ρ results with $\Phi(\rho)$, that is the next generation population.

Assumptions. In order to derive the proper operator, the following assumptions have to be made:

- Variable $o = (\phi, n) \in \mathbb{R}^n \times \mathbb{N}$ describes an individual with fitness ϕ and n TEs on its genome.
- In one generation $d \geq 0$ TEs can undergo deletion in a given organism o which is modeled with the binomial distribution:
$$\binom{n}{d} \delta^d (1-\delta)^{n-d}$$
where δ is the deletion rate.
- In one generation $k \geq 0$ TEs can transpose (i.e. copy themselves) in a given organism o which is modeled with the Poisson distribution:
$$\frac{(\mu n)^k}{k!} e^{-\mu n}$$
where μ is the transposition rate.

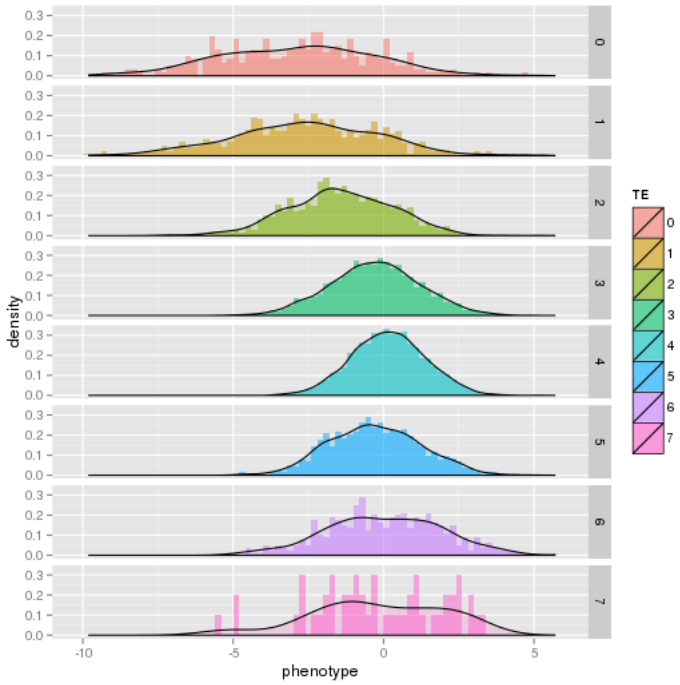


Figure 2.3: Example of a population described by the density ρ: One plot represents phenotype distribution in the sub-population of individuals each of which has exactly $n \in \{0, \ldots, 7\}$ TEs on their genome [Gogolewski, 2014].

- Both deletion and transposition events occur independently.
- Changes in the phenotype distribution of a population driven by mutations are modeled by calculating the convolution of the current phenotype distribution with the density function of the centered normal distribution $\mathcal{N}\left(0, \sigma^2 + \sigma_k^2\right)$, where σ^2, σ_k^2 are variations of the non-TE driven and TE-driven mutations respectively.
- There exists the optimal phenotype $0 \in \mathbb{R}^n$ and the environmental pressure is modeled as a deleterious change of a constant size η in the phenotype of all individuals in the population.
- Natural selection decides which organisms are well-adapted according to the centered normal distribution $\mathcal{N}(0, \xi^2)$, where ξ^2 describes the selection range, and 0 stands for the phenotypic optimum.

Operator Φ. The above assumptions along with the simplification that phenotypic space is one-dimensional ($n = 1$) allowed us to introduce the population operator $\Phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}$ of the form:

$$\Phi(\rho)(n, \phi) = \frac{\hat{\Phi}(\rho)(n, \phi)}{\sum_{n=0}^{\infty} \int_{\mathbb{R}^n} \rho(n, \phi) d\phi}$$

where $\hat{\Phi} : \mathbb{R} \times \mathbb{N} \to \mathbb{R} \times \mathbb{N}$ is defined as:

$$\hat{\Phi}(\rho)(n, \phi) = \nu(0, \xi^2) \cdot \sum_{d=0}^{\infty} \sum_{k=0}^{\infty} \left(\binom{n}{d} \delta^d (1 - \delta)^{n-d} \cdot \frac{(\mu n)^k}{k!} e^{-\mu n} \cdot \rho(n, \cdot) \star \nu(0, \sigma^2 + \sigma_k^2) \right) (\phi - \eta)$$

In this operator's definition the $\nu(0, \sigma^2)$ expression is the density of the centered normal distribution with variation σ^2 and $\star$ stands for the convolution of two functions.

Results. The analysis of the proposed operator involves proving its convergence and the existence of a fixed point, in order to determine if the TSE is attainable by the population. So far, it has been shown that simplified operator:

$$\widetilde{\Phi}(\rho)(\phi) = (\rho \star \nu(0, \sigma^2))(\phi - \eta) \cdot \nu(0, \xi^2)(\phi)$$

which models random mutations converges and has a unique, non-trivial fixed point, which is the density of the normal distribution:

$$\mathcal{N}\left(\frac{\eta\sqrt{4\xi^2+\sigma^2} - \eta\cdot\sigma}{2\sigma}, \frac{\sqrt{2\sigma\left(\sqrt{4\xi^2+\sigma^2}-\sigma\right)}}{2}\right)$$

Considering this result, the most unexpected observation is the fact that the variation of the fixed point does not depend on the intensity of the environmental pressure η. Furthermore, what is worth emphasizing is the simplicity and consistency of the proposed operator.

Moreover, this partial result suggest that the general operator Φ is also likely to have the non-trivial fixed point and therefore the TSE in asexuals confronted with the environmental pressure is attainable. Finally, one will observe that some ideas presented in the above formalization are straightforwardly used when describing the computational model.

2.2.2. Sexual populations

Another example of modeling TEs activity was derived considering a sexual population along with TEs presence and was presented in [Charlesworth and Charlesworth, 1983]. Authors constructed their models in order to investigate the existence of a TSE under certain conditions, however, one can notice that the assumptions that had been made are questionable, because they imposed the expected results.

We are about to follow two methods of modeling, both assuming that:

- populations are infinitely large,
- each organism is able to maintain up to T TEs on its genome,
- environmental conditions are stable (e.g. no stress is introduced).

Regulated transposition. The first model additionally assumed that there is no selection, while the rate of transposition and deletion of TE are a decreasing function $u(n)$ of TEs number and constant value v, respectively.

Authors derived the formula for a value of $\Delta\overline{n}$, that is a difference in a mean copy number of TEs on the genome of an organism between two consecutive generations. This value can be expressed as the expected difference between newly created and deleted TEs, that is $\Delta\overline{n} = \mathbb{E}(n \cdot u(n) - n \cdot v)$ and approximated as:

$$\Delta\overline{n} \approx \overline{n}(u(\overline{n}) - v) + \frac{\mathrm{Var}(n)}{2}\left(2u'(\overline{n}) + \overline{n}u''(\overline{n})\right)$$

where $\mathrm{Var}(n)$ is the variance in the copy numbers in the population and u', u'' are first and second derivatives of the transposition rate function. Then, using the definition of the genotypic variance [Bulmer, 1985], $\mathrm{Var}(n)$ was calculated as:

$$\mathrm{Var}(n) = \overline{n}\left(1 - \frac{\overline{n}}{T}\right) - T\sigma_{\mathcal{G}}^2 + 4\sum_{i<j} C_{ij}$$

From now on, stricter assumptions were introduced. First, it was assumed that the linkage equilibrium between all TEs is negligible, thus $4\sum_{i<j} C_{ij} = 0$ (intuitively the value of C_{ij} describes how the presence of a TE at the i-th locus influences the presence of a TE at the j-th one) and $\sigma^2_{\mathcal{G}} \to 0$, where $\sigma^2_{\mathcal{G}}$ is the genotypic variance in the whole population. As a result of a simple calculation, one obtains:

$$\Delta\overline{n} \approx \overline{n}(u\left(\overline{n}\right) - v) + \frac{\overline{n}}{2}\left(1 - \frac{\overline{n}}{T}\right)\left(2u'(\overline{n}) + \overline{n}u''(\overline{n})\right)$$

Finally, authors assumed that the regulation of transposition is weak, thus the function u is almost constant (consequently its derivatives are $\equiv 0$) and the value of $\Delta\overline{n}$ can be approximated as:

$$\Delta\overline{n} \approx \overline{n}(u\left(\overline{n}\right) - v)$$

With this result it can be concluded that the number $\hat{n}$ of TEs for which the TSE is attained (i.e. there are no changes in the copy number of TEs) is given by the following formula:

$$u(\hat{n}) = v$$

Although, the above calculations are correct, the result seems not to be meaningful. The assumption about the constancy of the transposition rate function allows us to deduce that if there exists a non-trivial number of TEs, $\hat{n}$, such that the TSE exists, the equality

$$\mathbb{E}\left(\hat{n} \cdot u(\hat{n}) - \hat{n} \cdot v\right) = 0$$

has to be satisfied, and so $u(\hat{n}) = v$.

Transpositions and selection. The second approach assumes that the transposition (u) and deletion (v) rates are constant, but the fitness function $\omega(n)$ is the decreasing function of the copy number of TEs present on the individual's genome. In order to prove the existence of the TSE authors calculate the probability Δx_i of the occupancy change at i-th locus in two consecutive generations (i.e. probability that a locus got occupied after being free or got freed after being occupied).

Using the Wright's formula [Wright, 1937], which describes the change in probability that a gene variant is present on a given locus after selection, it can be derived that in case of TSE the value of Δx_i for any locus i can be expressed as:

$$\Delta x = x(1-x)\frac{d\ln\omega(\overline{n})}{d\,\overline{n}} + \frac{u\overline{n}}{T - Tx}(1-x) - vx = x(1-x)\frac{d\ln\omega(\overline{n})}{d\,\overline{n}} + x(u-v)$$

and then transformed into:

$$\Delta\overline{n} = \overline{n}(1 - \frac{\overline{n}}{T})\frac{d\ln\omega(\overline{n})}{d\,\overline{n}} + \overline{n}(u-v) = \overline{n}\left[(1 - \frac{\overline{n}}{T})\frac{d\ln\omega(\overline{n})}{d\,\overline{n}} + (u-v)\right]$$

because of the fact that $\overline{n} = T \cdot x$, when the probability of TE at specific locus is equal to x.

Again, even though the formula for the non-trivial number of TEs in the TSE can be derived, it is not clear why authors postulate the thesis that selection should eliminate organisms that have more TEs on their genome. Not only this assumption appears to be biased, but it also enforce the stability of TEs copy number, because bursts of activity are explicitly muted by natural selection.

Chapter 3

Methods

The stochastic computational model that was used to analyze the evolutionary dynamics of populations in terms of the TEs activity and the ability to adapt to varying environmental conditions is based on the model proposed in [Startek et al., 2013] which was modified and extended in such a way that sexual populations could have been explored.

3.1. Initial conditions

We assume that a population follows the life cycle presented in Figure 3.1. Each cycle begins with the sexual population $\mathcal{P}$ of a constant size $|\mathcal{P}| = m \in \mathbb{N}$ composed of adult organisms representing both sexes: female and male (denoted as 1 and 2 respectively) and capable of reproduction. Moreover, every organism is equipped with $A \in \mathbb{N}$ autonomous transposons of *copy-paste* type, is described by its phenotype $\boldsymbol{\varphi} \in \mathbb{R}^n$ encoding n uncorrelated traits, and has a randomly chosen sex (being male or female has the same probability). Consequently, individual o representing the population $\mathcal{P}$ is of the form:

$$\mathcal{P} \ni o = (A, \boldsymbol{\varphi}, s) \in \mathbb{N} \times \mathbb{R}^n \times \{1, 2\}$$

3.2. TEs and their proliferation

Although only autonomous TEs are present on the genome of the individuals forming the first generation of the population, the model assumes both autonomous and non-autonomous to coexists during the simulation run. For simplicity, each newly created TE is assigned a unique number (*id*), thus it can be tracked during a simulation run, e.g. to detect for how many generations it has been active.

Moreover, according to the experimental and theoretical results, function that describes TEs activity should have the following properties:

- The transposition probability for both kinds of TEs is the same, and should be of the form $\tau_0 \cdot \tau(A, N)$, where τ_0 is the transposition rate parameter and A, N describe number of autonomous and non-autonomous TEs, respectively.
- $\tau(1, 0) = 1$, meaning that τ_0 is the nominal transposition rate when only one autonomous copy is present.
- $\tau(0, N) = 0$, for any $N \in \mathbb{N}$, to ensure that no transposition events occur when there are no autonomous copies.

- $\frac{\partial \tau}{\partial A} \leq 0$, when $A \geq 1$, so the transposition rate per copy is maximal when there is only one copy (although it can be constant).
- $\frac{\partial \tau}{\partial N} \leq 0$, not to allow non-autonomous copies to increase the transposition rate.

One can easily verify that the function:

$$\tau(A, N) = \tau_0 \cdot \frac{A}{A + N}$$

satisfies all of the above conditions, and so function τ is used in the model as the moderator of the dynamics of transposition mechanism (i.e. describes the probability that a given transposon will undergo the transposition event).

Furthermore, since autonomous TEs can loose their ability to produce the mechanism crucial for transposition event to happen (and so become non-autonomous), we assume that such functionality loss happens with frequency Δ_α. Finally, every TE can undergo a permanent deletion with frequency Δ_β. The whole dynamics of TEs is presented in Figure 3.1.

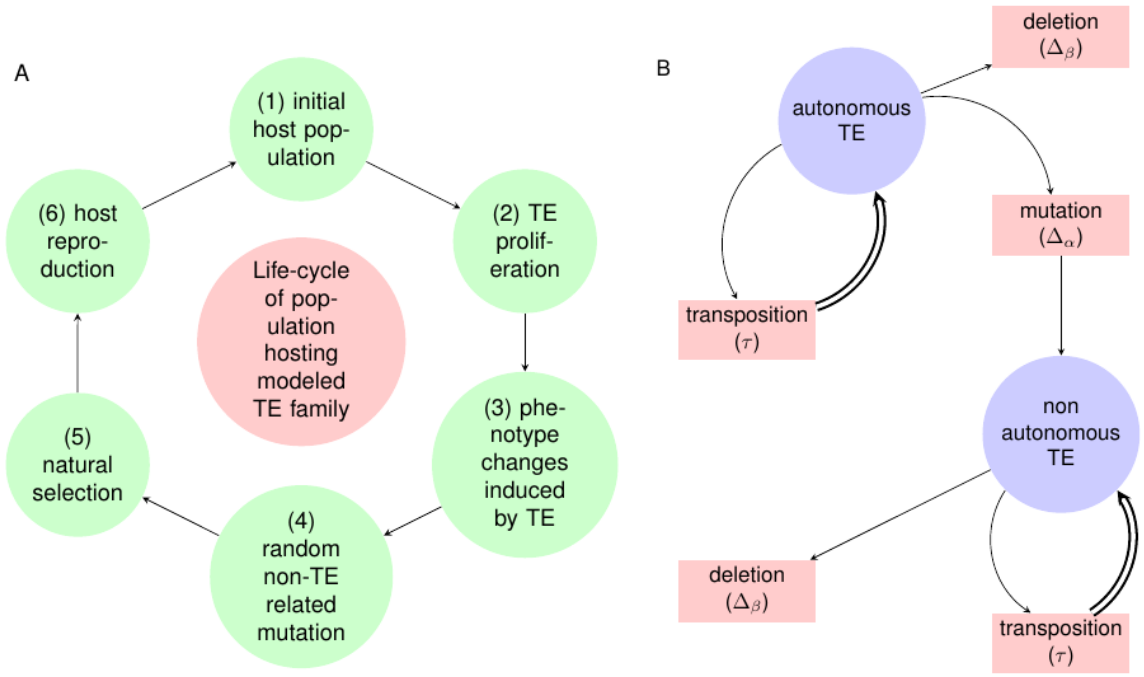


Figure 3.1: (A) The population's life cycle: The graph presents all stages of the simulation that a population has to go through during one generation. (B) TEs proliferation dynamics on a host's genome. Source: [Startek et al., 2013]

3.3. Phenotype and mutational changes

In the model two ways of introducing the mutational changes, both based on the FGM and also two ways of calculating the organism's phenotype were tested.

Mutational changes: multi vs one dimensional As it was stated in the Section 2.1, the pure FGM assumes that a single mutation event affects each phenotypic character of an

individual's phenotype according to multi-dimensional centered normal distribution. Nevertheless, since the problem of pleiotropy is deeply studied, in the model we also introduce the idea of one, random trait being modified by the value drawn from one-dimensional centered normal distribution when mutation event occurs.

Phenotype calculation methods: binding vs non-binding Phenotype calculation was also performed in two different ways.

The first attempt, based on [Startek et al., 2013] assumes that the phenotype (and so fitness) is exposed to changes emerging from both TE and non-TE driven mutations. Specifically, one mutation event changes one (or all, for multidimensional changes) phenotypic trait of the phenotype by a value drawn from the centered normal distribution $\mathcal{N}(0, \mu^2)$, where μ^2 is the mutational variation parameter (or $\mathcal{N}(\mathbf{0}, \mathbb{I} \cdot \mu^2)$, where $\mathbb{I}$ is the identity matrix and $\mathbb{I} \cdot \mu^2$ is the diagonal covariance matrix), for either non-TE related mutations or TE related mutations (i.e. those which occurred as a result of any transposition event). From now on, we will refer to this phenotype calculation method (PCM) as a *non-binding PCM*.

For the purpose of the second approach we need to make few assumptions about the TEs activity.

- When a TE is created a fixed mutational change ξ that will affect the host's phenotype is drawn from the centered normal distribution $\mathcal{N}(0, \mu^2)$ (or $\mathcal{N}(\mathbf{0}, \mathbb{I} \cdot \mu^2)$) with mutational variance μ^2 and assigned to it.

- New TEs that appear as a result of a copy-paste transposition event obtain a new mutational change ξ (i.e. it is not inherited from the parental TE).

- All TEs that were inherited by progeny possess the same mutational change as they used to on the parental genome.

Construction of the second approach assumes that the phenotype, $\boldsymbol{\varphi}$, of an organism is the sum of two components $\boldsymbol{\varphi} = \boldsymbol{\varphi_b} + \boldsymbol{\varphi_t}$ the *base-phenotype*, $\boldsymbol{\varphi_b}$, which changes only due to random mutations not induced by TEs and the *TE-phenotype-contribution*, $\boldsymbol{\varphi_t} = \sum_k \xi_k$, which is the sum of all mutational changes assigned to TEs currently present on the organism's genome [Le Rouzic and Capy, 2005]. By analogy, this method of phenotype calculation will be referred as *binding PCM* and its introduction is motivated by the experimental results concerning the role of TEs in human disease [Callinan and Batzer, 2006].

It is worth emphasizing that both types of mutations, TE and non-TE driven, can have either positive or negative influence on the organism's fitness by definition, thus the model does not favor neither of them, and finally the size of mutational change is drawn from the same distribution regardless of the cause of mutation (i.e. TE or non-TE driven).

3.4. Fitness function and reproduction

Assuming that there exists the *phenotypic optimum* $\mathbb{R}^n \ni \boldsymbol{\varphi^{opt}} = \left[\varphi_k^{opt}\right]_{k=1}^n$ we calculate the fitness of an organism with phenotype $\boldsymbol{\varphi} = [\varphi_k]_{k=1}^n$ using fitness function $f_{\boldsymbol{\varphi^{opt}}} : \mathbb{R}^n \to (0, 1]$ defined as:

$$f_{\boldsymbol{\varphi^{opt}}}(\boldsymbol{\varphi}) = e^{-\|\boldsymbol{\varphi} - \boldsymbol{\varphi^{opt}}\|^2} = e^{-\sum_{k=1}^{n} \left(\varphi_k - \varphi_k^{opt}\right)^2}$$

Further on $f_{\boldsymbol{\varphi^{opt}}}$ is denoted by f for the notation simplification.

Next, we assume that two consecutive populations are disjoint, meaning exactly m times a pair of female and male is chosen from the current generation to reproduce and to introduce a new organism to the next generation. Each time every organism of phenotype $\boldsymbol{\varphi}$ and sex s has probability $p_{\boldsymbol{\varphi},s}$ of being chosen as individual to reproduce, given by following formula:

$$p_{\boldsymbol{\varphi},s} = \begin{cases} \frac{f(\boldsymbol{\varphi})}{f(\mathcal{F})} \text{ if } s = 1 \\ \\ \frac{f(\boldsymbol{\varphi})}{f(\mathcal{M})} \text{ if } s = 2 \end{cases}$$

where $f(\mathcal{A})$ represents the sum of phenotypes of all organisms in sub-population $\mathcal{A} \subseteq \mathcal{P}$, while $\mathcal{F}$ and $\mathcal{M}$ denote all females and males in population respectively.

Having chosen one pair of individuals for reproduction, the creation of a new organism starts. First, each parent generates a gamete with the support of the chromosomal crossing over mechanism, thus the decision which TEs will be passed to an offspring is made. Specifically:

- If there is only one copy of a TE with a given *id* on both chromosomes, it has $\frac{1}{2}$ chance to be inherited.
- If there are two copies of a TE with a given *id*, it is inherited.
- The crossing-over cut does not happen on the sequence coding an active TE (i.e. during crossing-over TEs do not lose their functionality).

Those two generated gametes are combined together to be recognized as a new organism with its phenotype value calculated according to the PCM that is used:

Non-binding PCM For *non-binding PCM* the offspring phenotype $\tilde{\boldsymbol{\varphi}} \in \mathbb{R}^n$ is given as:

$$\tilde{\boldsymbol{\varphi}} = \frac{\boldsymbol{\varphi^1} + \boldsymbol{\varphi^2}}{2} + \varsigma_e + \varsigma_g$$

where $\boldsymbol{\varphi^1}, \boldsymbol{\varphi^2} \in \mathbb{R}^n$ are parental phenotypes and $\varsigma_e, \varsigma_g \in \mathbb{R}^n$ are environmental and genetic values respectively [Bulmer, 1985]. Here, we assume that the environmental value is negligible (i.e. $\varsigma_e \equiv 0$), while ς_g is drawn from the centered normal distribution with variance being the phenotypic variance in the population in question, that is:

$$\varsigma_g \sim \mathcal{N}\left(0, \frac{1}{n}\sum_{k=0}^{m}\left(\phi_k - \overline{\phi_k}\right)^2\right)$$

Finally, each trait of the phenotype is calculated coordinate-wise, meaning:

$$\forall_k \, \tilde{\varphi}_k = \frac{\varphi_k^1 + \varphi_k^2}{2} + \varsigma_{g_k}$$

Binding PCM In the case of the *binding PCM*, only the base-phenotype $\tilde{\boldsymbol{\varphi}}_{\boldsymbol{b}}$ is calculated in accordance with the above formula, while the TE genomic content of the organism (i.e. TEs that were inherited by an individual) determines the TE phenotypic contribution $\tilde{\boldsymbol{\varphi}}_{\boldsymbol{t}}$.

After m reproduction events the new population is composed and the next cycle begins. The process of reproduction is also presented in Figure 3.2.

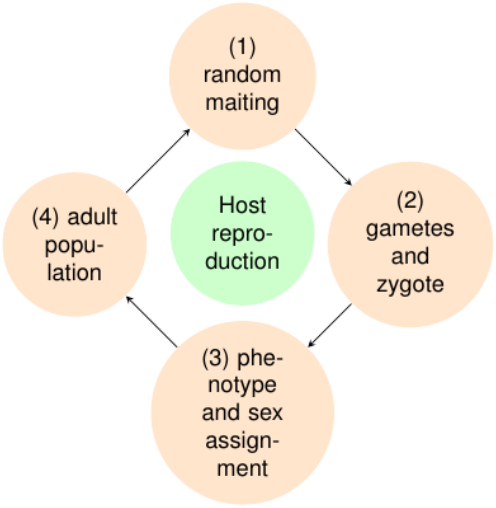


Figure 3.2: Outline of the sexual reproduction in the model: Reproduction phase is composed of m runs of four separate steps; 1. Random mating - parents are chosen with probabilities proportional to their fitness. 2. Each parent produces gamete with the support of an crossing-over mechanism. 3. Sex selection and generation of the offspring's phenotype. 4. Combined gametes create zygote which is assumed to be an adult organism in the next generation.

3.5. Environmental stress

Finally, the model assumes that the phenotypic optimum is time dependent and q traits change every Γ generations, while remaining $n - q$ traits stay invariant. The idea of such partition is meant to provide more natural concept of environmental changes, where stress does not necessarily affect each trait, but is likely to affect only a particular subset. In presented simulations we consider two scenarios of the optimum shift.

- *Global warming* scenario: Slow, gradual environmental changes that affect the host population are modeled as a directed shift of the *optimal phenotype* in each consecutive generation by a small amount.
- *Meteor Impact* scenario: Strong, instantaneous environmental changes that might eliminate vast number of individuals in the population are modeled as a single shift of the *optimal phenotype* by a large value every Γ generations

Chapter 4

Results and discussion

Having introduced the whole model, we can now proceed to the presentation of sample results obtained from the computational model for specific scenarios. Since the model is substantially an implementation of a stochastic process, each simulation run is just one possible trajectory of this process. That is why all graphs that are included below represent the most common outputs for a given scenario with specified set of parameters (see Table 4.1).

4.1. Regular environmental pressure

The dynamics of population in case of environmental changes presented in the Section 3.5 are first investigated. The set of parameters for these simulation runs is chosen in such a way that the model performance is close to the natural dynamics of TEs proliferation in sexual populations observed by means of experimental research.

Parameter name	General GW	General MI (rare hits)	General MI (frequent hits)	Boundary (no transp.)	Boundary (no env. change)	Pleiotropy (one-dimension)
General parameters						
Niche size (m)	1000	1000	1000	1000	1000	1000
Traits number (N)	10	10	10	10	10	10
Transposition/ mutation rate (τ_0)	0.003	0.003	0.003	0	0.003	0.003
Mutational variance (μ^2)[1]	0.01	0.01	0.01	0	0.01	0.01
Deletion rate (Δ_β)	0.003	0.003	0.003	0.003	0.003	0.003
Autonomy loss rate (Δ_α)	0.003	0.003	0.003	0.003	0.003	0.003
Environmental change parameters						
Shift size (η)	0.006	0.45	0.45	0.006	0	0.006
Shift frequency (Γ)	1	500	300	1	-	1
No. traits affected (q)	4	10	10	4	-	1

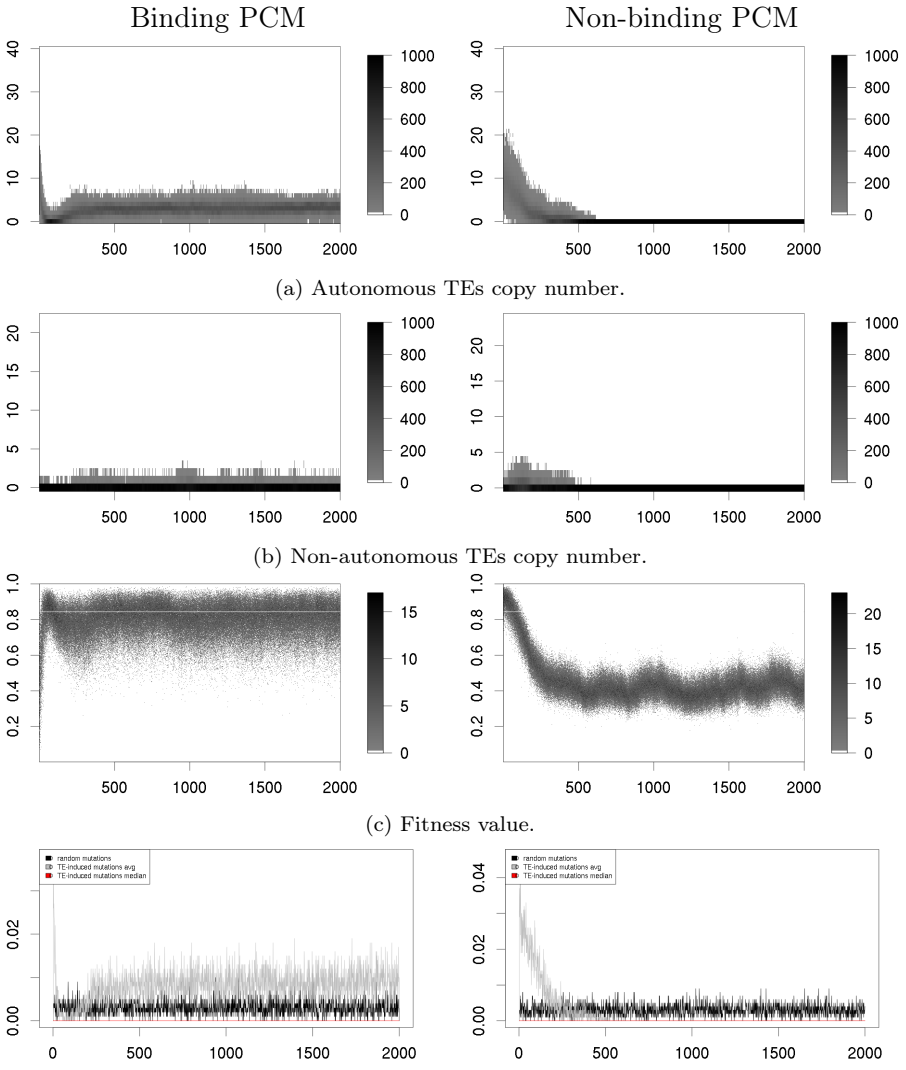
[1] In case of multidimensional mutations we consider covariance matrix of the form: $\mathbb{1} \cdot \mu^2$.

Table 4.1: Crucial parameters of the model: The table presents the values of parameters that were used for specific scenarios presented in this chapter.

4.1.1. *Global warming* scenario

Let us recall that the global warming scenario is the gradual, directed shift of the optimal phenotype in each generation by a constant rate. Figure 4.1 presents the most common outputs of the simulation run for two different phenotype calculation methods: *binding* and *non-binding PCM* described in the Section 3.3.

It can be observed that the method change leads to significant difference in the behavior of the system. *Non-binding PCM* results in the mediocre adaptation, while *binding PCM* yields to the progressively adapting population characterized by the high fitness value (see figure: 4.1c). It is reasonable to conclude that such a difference in the fitness value between populations arises from the difference in TEs activity (see Figure 4.1(d)) and their presence in the population (see Figure 4.1(a),(b)). We can conclude that the complete randomness



(a) Autonomous TEs copy number.

(b) Non-autonomous TEs copy number.

(c) Fitness value.

(d) Mutational activity.

Figure 4.1: The global warming scenario: Examples of the most common results of the simulation run for both *binding PCM* (left side) and *non-binding PCM* (right side). The (a) and (b) sub-figures present the number of autonomous and non-autonomous TEs (y-axis) in each generation (x axis) in population, respectively. The (c) plot is a distribution of the fitness value (y-axis) in population in each generation (x-axis), while the (d) is a comparison of the size of mutational changes invoked by non-TE and TE-induced mutations. In case of plots (a-c) the gray scale describes the number of individuals in population with a given value.

of mutational effects of transposition events (*non-binding*) is not beneficial enough to maintain TEs presence in the population. On the other hand, the persistent attachment of the mutational change to the specific TE (*binding*) allows selection to eliminate deleterious TEs

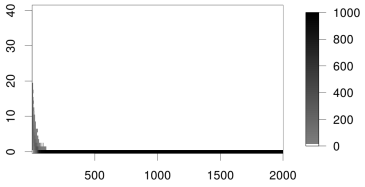
and preserve only those which are beneficial and keep the track of the optimum phenotype shift. That behaviour is a strong evidence for the existence of the TSE in the scenario under consideration. However, the formal proof of that fact is not a part of this thesis.

4.1.2. *Meteor impact* scenario

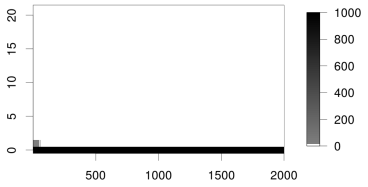
The second scenario assumes large, directed shifts of the phenotypic optimum every Γ generations. Preliminary results showed that the importance of TEs increases with the frequency of the meteor impact (i.e. optimum shift). As a justification of this fact we can use following observation: a scenario when both the Γ parameter and the shift size, η, are relatively small can be well-approximated by the *global warming* scenario. Let us look at the results of meteor impact scenario for two different Γ values: 500 and 150.

Sporadic meteor hits: $\Gamma = 500$. When the optimum shift is too rare event, TEs activity is redundant because environmental conditions are stable for the long period of time and there is no need for adaptation. In this case both PCMs have the same result: before the first meteor hit the number of TEs tends to zero (see Figures 4.2(a) and 4.2(b)), and therefore there are no transposition related mutations involved for the rest of the simulation run (see Figure 4.2(d)). Nevertheless, no TEs activity does not mean that the population is poorly adjusted and has low fitness value. In fact, because of long periods of stability random mutations are able to reach the new phenotypic optimum and the population is well-adapted before the next meteor impact (see Figure 4.2(c)).

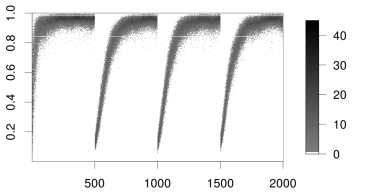
These observations allow us to postulate the hypothesis that in the case of no environmental stress, the activity of TEs decreases rapidly and since they are not beneficial for the population are permanently eliminated. Moreover, spontaneous mutations are sufficient tool for the population to reach the static phenotypic optimum. This fact is also an evidence that the computational model is correct because activity of TEs is neither necessary in the process of adaptation nor favoured over the typical random mutations.



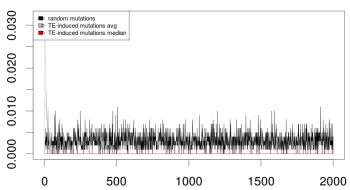
(a) Autonomous TEs copy number.



(b) Non-autonomous TEs copy number.



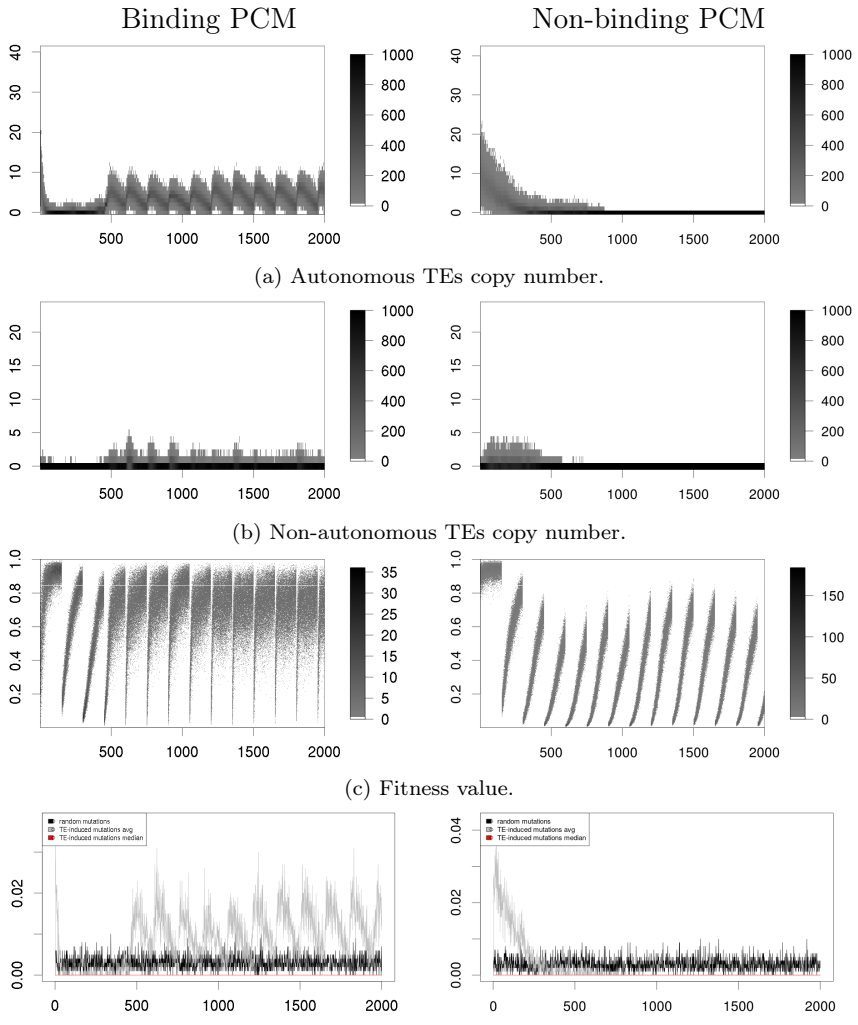
(c) Fitness value.



(d) Mutational activity.

Figure 4.2: The meteor impact scenario in case of sporadic shifts: (a) and (b) present the rapid fall in the copy number of TEs in the population. Consequently, (d) shows that mutations that drive the evolution of the population are only non-TE induced. Finally, (c) proves that random mutations are able to follow the stable phenotypic optimum.

Frequent meteor hits: $\Gamma = 150$**.** In contrast to rare shifts in the optimal phenotype, when it comes to frequently occurring changes, TEs have an important role in the adaptation process to varying conditions. However, these observations again can be made only when we assume *binding PCM*, otherwise TEs are eliminated from the population, even though they are active for more than three meteor hits (see Figure 4.3(a), (b)). In particular, when considering the amount of time that is needed for population to adapt to a distant phenotypic optimum we can observe that in case of the *binding PCM* TEs activity noticeably shortens it (see Figure 4.3(c)). When phenotype changes in population are driven only by random mutations, they indeed lead the population closer to the optimum, but the complete adjustment is impossible because of the appearance of the next meteor impact.



(a) Autonomous TEs copy number.

(b) Non-autonomous TEs copy number.

(c) Fitness value.

(d) Mutational activity.

Figure 4.3: The meteor impact scenario in case of frequent shifts (*binding* vs *non-binding* PCM): All graphs present the cyclic behaviour of transposition dynamics in the case of *binding PCM*. Furthermore, the comparison of (c) plots shows the beneficial contribution of TEs in adaptation process.

On the other hand, cyclic bursts in both the copy number of TEs and their activity provide faster adjustment, resulting in well-adapted population dozens of generations before the upcoming next shift of the phenotypic optimum. Moreover, periodicity of those activity bursts is the next evidence supporting the fact that TEs are muted in case of longer periods

of no environmental conditions change.

Finally, we can draw a general conclusion that the *non-binding* PCM that was used in [Startek et al., 2013] for modeling the phenotype changes induced by TEs activity does not give any meaningful results in the case of sexual populations. Regardless of the type of environmental stress, the random contribution to the phenotype change generated by TEs activity is equally beneficial as spontaneous mutational changes, thus they are likely to be eliminated by natural selection. That is why the rest of results will only take *binding PCM* into consideration.

4.2. Boundary conditions

No environmental changes. This scenario is meant to confirm the, already mentioned, lack of significance of TEs in the case of no changes in the phenotypic optimum throughout the complete simulation run. As it was already postulated, when the environmental conditions are stable TEs are expected to be removed by natural selection. The following plots (see Figure 4.4) prove that in the initial phase of simulations TEs are indeed excluded since their activity is not required for population to be distributed around the phenotypic optimum. Moreover, it can be observed that the presence of natural selection ensures that individuals affected by number of deleterious mutations are excluded from the population. As an effect of such control the highly adjusted population is maintained until the end of a simulation run.

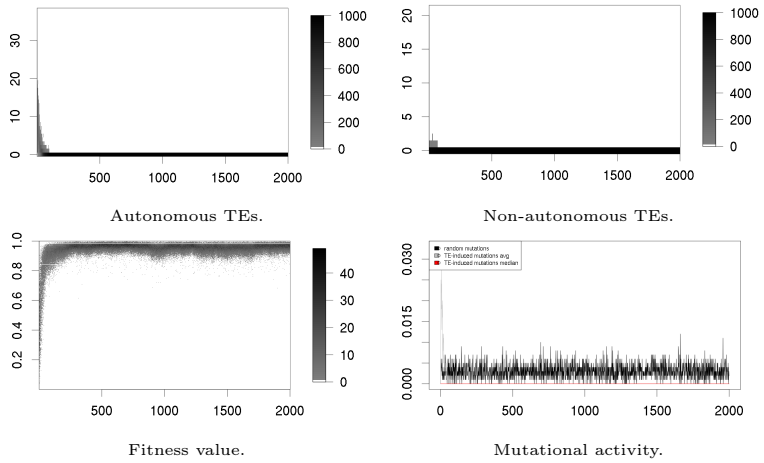


Figure 4.4: No environmental changes: Above graphs present the immediate exclusion of TEs from the population because of their redundancy.

No transposition events. Another property of the model is connected directly with the sexuality of populations in question. This scenario is designed in order to expose the fact that even if the transposition rate is equal to zero (i.e. transposition events does not occur and new TEs won't be introduced) the mean number of TEs in population can fluctuate only because of the support of the natural selection and reproduction (crossing-over, in particular). Figure 4.5 shows that TEs can exist for a long time before the total exclusion by natural selection, especially when their starting phenotypic contribution is highly beneficial.

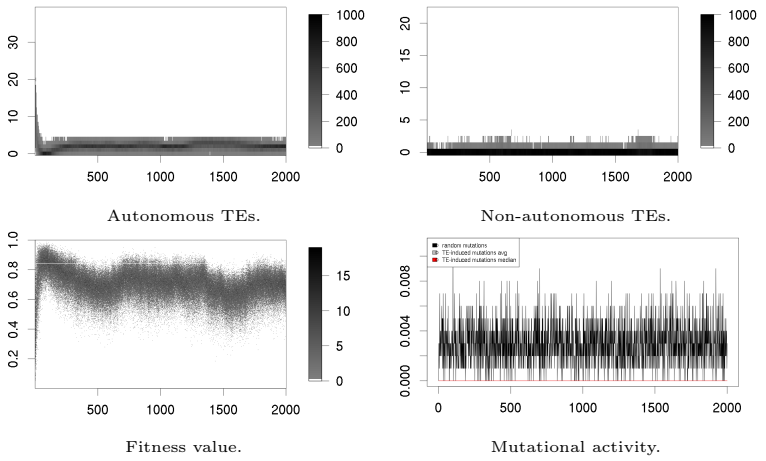


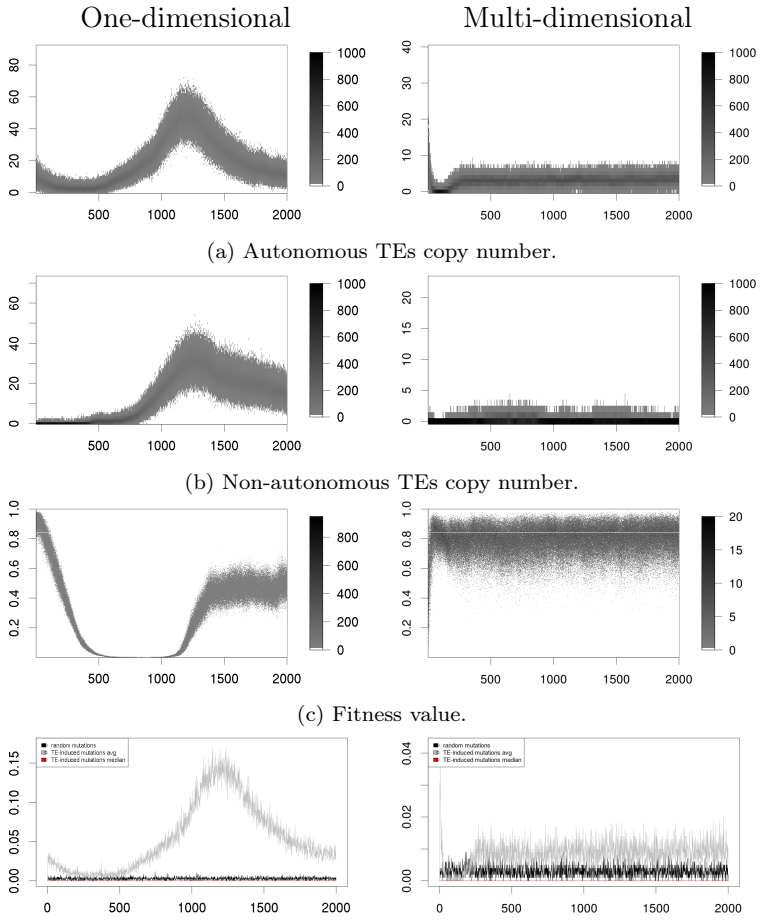
Figure 4.5: No transposition events in changing environment: No transposition events does not imply immediate exclusion of TEs. Above results prove that the stochastic process stitched in the model allows TEs to exists only thanks to inheritance.

4.3. The significance of pleiotropy

In this last scenario we want to investigate the relationship between the number of traits affected by mutational changes and the rate of adaptation. We want to challenge if the increase in the number of traits that describe the mutagenic change results in a longer adaptation process and so it is more likely that a population will go extinct.

The following results are meant to present a significant difference in the amount of time that is needed for population to adapt to changing conditions for two different ways of introducing the mutational changes: the first one assuming the universal isotropic pleiotropy (i.e. all traits are influenced by the mutational change as in previous simulations) and the second one assuming the one-dimensional mutational changes. Figure 4.6 shows how the adaptability of populations change for those two different approaches.

Note that in case of one-dimensional mutational changes random mutations are unable to keep the track of moving phenotypic optimum. In the first phase of simulations the mean fitness in population is sharply decreasing near to extinction (i.e. value of fitness tends to 0), while the number of both types of TEs increases up to the point where their summary phenotypic contribution allows the population to get at least half of the way to the optimum. Moreover, comparing to the Figure 4.1(c), where TEs were quickly removed, one can observe that the presence of TEs slightly improves the global fitness in the population and again gives us an evidence for their beneficial activity.



(a) Autonomous TEs copy number.

(b) Non-autonomous TEs copy number.

(c) Fitness value.

(d) Mutational activity.

Figure 4.6: Multi-dimensional vs one-dimensional introduction of mutational changes: (a, b, d) show the difference in copy numbers of TEs and their activity that is required by the population to start the proper adaptation process. The (c) plot describes the time delay before the revival of the population.

Chapter 5

Conclusions

In this thesis a computational model of TEs evolution in sexual populations exposed to environmental stress was developed, and the model itself is sufficiently realistic for broad analysis of real-world phenomena. Moreover, dynamics of TEs was implemented in accordance with the current theoretical and experimental knowledge, inculding the explicit effect of TE mobility on the host's phenotype and fitness. Presented simulations evidence that the activity of TEs in sexuals can be a very complex process including bursts, re-invasions, losses and apperances of TE copies. However, these events occur mainly when spontanuous mutations are not able to follow the moving phenotypic optimum. In that case, TEs can significantly improve the process of adaptation to varying conditions, and individuals that are hosts for them are favored by the natural selection. It was also investigated that the pleiotropy of mutational changes can also affect the quality and tempo of adjustment of populations. Finally, it was shown that the transposition-selection equilibrium is attainable in sexual populations.

Various extensions of the proposed model are possible. The phenomenon of horizontal transfer can be modeled inside our framework, e.g. to determine how it influences the process of adaptation. The spatial distribution of organisms' phenotypes can be condiered giving the posibility to study the influence of transposons activity on population adapting to new niches. Furthermore, it seems reasonable to introduce new types of environmental stress, e.g. brownian motion or cyclic changes and test if the TSE is still attainable. Lastly, the model can be used for the analysis of the full history of transposition events that occur in population throughout the whole simulation, which may help to understand the dynamics of the system even better.

Chapter 6

Acknowledgments

This research and work was supported by the National Science Centre Poland grant: Modelling stress-induced transposon activity, number: 2012/06/M/ST6/00438.

Bibliography

[Antonov et al., 2010] Antonov, A., Stokke, B. G., Vikan, J. R., Fossøy, F., Ranke, P. S., Røskaft, E., Moksnes, A., Møller, A. P., and Shykoff, J. A. (2010). Egg phenotype differentiation in sympatric cuckoo cuculus canorus gentes. *J. Evol. Biol.*, 23:1170–1182.

[Barrett and Schluter, 2008] Barrett, R. D. and Schluter, D. (2008). Adaptation from standing genetic variation. *Trends Ecol. Evol.*, 23(1):38–44.

[Bulmer, 1985] Bulmer, M. G. (1985). *The mathematical theory of quantitative genetics*. Oxford University Press.

[Bürger and Gimelfarb, 2002] Bürger, R. and Gimelfarb, A. (2002). Fluctuating environment and the role of mutation in maintaining quantitative genetic variation. *Genet. Res.*, 80:31–46.

[Callinan and Batzer, 2006] Callinan, P. A. and Batzer, M. A. (2006). Retrotransposable elements and human disease. *Genome Dynamics*, 1:104–115.

[Charlesworth and Charlesworth, 1983] Charlesworth, B. and Charlesworth, D. (1983). The population dynamics of transposable elements. *Genet. Res.*, 42:1–27.

[Charlesworth et al., 1994] Charlesworth, B., Sniegowski, P., and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 371:215–220.

[Doolittle and Sapienza, 1980] Doolittle, W. and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature*, 284(5757):601–603.

[Fisher, 1930] Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford University Press.

[Gogolewski, 2014] Gogolewski, K. (2014). Modeling stress-induced transposons' proliferation. Bachelor's thesis, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw.

[Hickey, 1982] Hickey, D. A. (1982). Selfish DNA : a sexually-transmitted nuclear parasite. *Genetics*, 101:519–531.

[Le Rouzic and Capy, 2005] Le Rouzic, A. and Capy, P. (2005). The first steps of transposable elements invasion. *Genetics*, 169:1033–1043.

[Le Rouzic and Deceliere, 2005] Le Rouzic, A. and Deceliere, G. (2005). Models of the population genetics of transposable elements. *Genet. Res.*, 85:171–181.

[Lisch, 2013] Lisch, D. (2013). How important are transposons for plant evolution? *Nature Reviews Genetics*, 14:49–61.

[Mackay, 1986] Mackay, T. (1986). Transposable element-induced fitness mutations in drosophila melanogaster. *Genet. Res.*, 48:77–87.

[Matuszewski et al., 2014] Matuszewski, S., Hermisson, J., and Kopp, M. (2014). Fisher's geometric model with a moving optimum. *Evolution*, 105.

[McClintock, 1956] McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harbor Symp. Quant. Biol.*, 21:197–216.

[Miller et al., 1992] Miller, W. J., Hagemann, S., Reiter, E., and Pinsker, W. (1992). P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche*. *Proc. Natl. Acad. Sci. USA*, 89(9):4018–4022.

[Nunney, 2003] Nunney, L. (2003). The cost of natural selection revisited. *Ann. Zool. Fennici*, 40:185–194.

[Orr, 1998] Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, 52:935–949.

[Orr, 2000] Orr, H. A. (2000). Adaptation and the cost of complexity. *Evolution*, 54:13–20.

[Orr, 2009] Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nat. Rev. Genet.*, 10:531–539.

[Partridge and Barton, 2000] Partridge, L. and Barton, N. (2000). Evolving evolvability. *Nature*, 407:457–458.

[Pigliucci, 2008] Pigliucci, M. (2008). Is evolvability evolvable? *Nat. Rev. Genet.*, 9(1):75–82.

[Schlenke and Begun, 2004] Schlenke, T. and Begun, D. (2004). Strong selective sweep associated with a transposon insertion in drosophila simulans. *Proc. Natl. Acad. Sci. USA*, 101:1626–1631.

[Sinzelle et al., 2009] Sinzelle, L., Izsvák, Z., and Ivics, Z. (2009). Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cell. Mol. Life. Sci.*, 66(6):1073–1093.

[Startek, 2014] Startek, M. (2014). Modelling the evolution of mobile genetic elements. *PhD Thesis*, The Faculty of Mathematics, Informatics and Mechanics, University of Warsaw.

[Startek et al., 2013] Startek, M., Le Rouzic, A., Capy, P., Grzebelus, D., and Gambin, A. (2013). Genomic parasites or symbionts? modeling the effects of environmental pressure on transposition activity in asexual populations. *Theor. Popul. Biol.*, 90:145–151.

[Stoebel and Dorman, 2010] Stoebel, D. M. and Dorman, C. J. (2010). The effect of mobile element IS10 on experimental regulatory evolution in *Escherichia coli*. *Mol. Biol. Evol.*, 27(9):2105–2112.

[Tachida and Iizuka, 1993] Tachida, H. and Iizuka, M. (1993). A population genetic study of the evolution of sines. polymorphism with regard to the presence or absence of an element. *Genetics*, 133:1023–1030.

[Taddei et al., 1997] Taddei, F., Radman, M., Maynard-Smith, J., et al. (1997). Role of mutator alleles in adaptive evolution. *Nature*, 387(6634):700–702.

[Waxman and Peck, 1999] Waxman, D. and Peck, J. R. (1999). Sex and adaptation in a changing environment. *Genetics*, 153:1041–1053.

[Wright, 1937] Wright, S. (1937). The distribution of gene frequencies in populations. *Proc. Natl. Acad. Sci. USA*, 23:307–320.