Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

On the waiting time until coordinated mutations get fixed in regulatory sequences

Ola Hössjer [a,*], Günter Bechly [b], Ann Gauger [b]

[a] Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden
[b] Biologic Institute, 16310 NE 80th Street, Redmond, WA 98052, USA

ARTICLE INFO

Article history:
Received 4 August 2020
Revised 13 February 2021
Accepted 24 February 2021
Available online 04 March 2021

Keywords:
Binding site
Continuous time Markov process
Coordinated mutations
Moran model
Phase-type distribution
Regulatory sequence
Waiting time

ABSTRACT

In this paper we consider the time evolution of a population of size N with overlapping generations, in the vicinity of m genes. We assume that this population is subject to point mutations, genetic drift, and selection. More specifically, we analyze the statistical distribution of the waiting time T_m until the expression of these genes have changed for all individuals, when transcription factors recognize and attach to short DNA-sequences (binding sites) within regulatory sequences in the neighborhoods of the m genes. The evolutionary dynamics is described by a multitype Moran process, where each individual is assigned a $m \times L$ regulatory array that consists of regulatory sequences with L nucleotides for all m genes. We study how the waiting time distribution depends on the number of genes, the mutation rate, the length of the binding sites, the length of the regulatory sequences, and the way in which the targeted binding sites are coordinated for different genes in terms of selection coefficients. These selection coefficients depend on how many binding sites have appeared so far, and possibly on their order of appearance. We also allow for back mutations, whereby some acquired binding sites may be lost over time. It is further assumed that the mutation rate is small enough to warrant a fixed state population, so that all individuals have the same regulatory array, at any given time point, until the next successful mutation arrives in some individual and spreads to the rest of the population. We further incorporate stochastic tunneling, whereby successful mutations get mutated before their fixation. A crucial part of our approach is to divide the huge state space of regulatory arrays into a small number of components, assuming that the array component varies as a Markov process over time. This implies that T_m is the time until this Markov process hits an absorbing state, with a phase-type distribution. A number of interesting results can be derived from our general setup, for instance that the expected waiting time increases exponentially with m, for a selectively neutral model, when back-mutations are possible.

1. Introduction

A classical problem of population genetics is to study the time until new genetic variants first appear through germline mutations and then get fixed, i.e. spread to all individuals of a species, as it adapts to a new environment and evolves over time. Early work focused on genes or portions of DNA with only two variants, the wildtype allele and the mutated allele. By modeling the fraction of the wildtype allele as a Markov process, expressions were found for the probability of fixation of the mutant allele and conditionally on this, the expected waiting time until fixation. An overview of some of these results can be found in the monographs of Crow and Kimura (1970), Ewens (2004) and Durrett (2008).

More recently, researchers have addressed the more challenging problem of analyzing evolution of whole DNA sequences of nucleotides of length L, written on the four letter alphabet A, C, G, T. The evolutionary process is then a random walk on the large fitness landscape of DNA strings (Gillespie, 1984), whose steps are initiated by point mutations at single nucleotides. In order to simplify the analysis, Gillespie assumed a small mutation rate and a fixed state model, whereby all individuals in the population at any time point have the same DNA string. The transition rates between two neighboring fixed states of the random walk can be then derived from the fixation probability and waiting time formulas of the classical theory, treating the original and mutated sequence as the wildtype and mutant allele. For instance, Chatterjee et al. (2014) studied the waiting time until a certain

[*] Corresponding author.
E-mail addresses: ola@math.su.se (O. Hössjer), bechly@mac.com (G. Bechly), ann.gauger@gmail.com (A. Gauger).

https://doi.org/10.1016/j.jtbi.2021.110657
0022-5193/© 2021 The Author(s). Published by Elsevier Ltd.

target of DNA sequences is reached; all sequences whose fraction of mismatches with one or a few fixed DNA sequences is at most c. For a neutral model, they showed that this waiting time increases either polynomially or exponentially with L, depending on whether c is larger or smaller than $3/4$.

For the evolution of gene expression, it is of interest to consider changes of non-coding DNA, more specifically evolution of regulatory regions within the enhancer or promoter region of a gene, whose length L is of the order 1000. Transcriptional regulation is perhaps the most studied form of gene regulation, where transcription factor proteins recognize and bind to short subsequences of the regulatory region, so that the expression of the nearby gene changes. These subsequences are called binding sites, and typically they have a length of 6–10 nucleotides. The target of the random walk then consists of all regulatory sequences along which at least one of a few pre-specified binding sites appears somewhere along the sequence, possibly allowing for mismatches at one or a few nucleotides between a binding site and a substring of the regulatory sequence (Stone and Wray, 2001; MacArthur and Brockfield, 2004; Yona et al., 2017). The waiting time until the expression of the gene changes, is modeled as the time until the random walk hits the target, and it depends on the mutation rate, the selective advantage of the mutated regulatory sequence, the size of the population, the length of the regulatory sequence and the length of the binding sites (Durrett and Schmidt, 2007, 2008; Behrens and Vingron, 2010; Behrens et al., 2012; Nicodéme, 2012; Tuğrul et al., 2015; Sanford et al., 2007, 2015). For more complex adaptations of a species, it is necessary that *several genes* are modified in a *coordinated* manner, either through mutations in the coding sequence, or through changed expression of these m genes. For instance, the fossil record is often interpreted as having long periods of stasis (Voje et al., 2018), interrupted by more abrupt changes and "explosive" origins (Bechly and Meyer, 2017). These changes include, for instance, the evolution of life (Bell et al., 2015), photosynthesis (Hecht, 2013), multicellularity and the "Avalon Explosion" (Shen et al., 2008), animal body plans and the "Cambrian Explosion" (Erwin and Valentine, 2013), complex eyes (Paterson et al., 2011), vertebrate jaws and teeth (Fraser et al., 2010), terrestrialization (e.g., in vascular plants, arthropods, and tetrapods) (Bateman et al., 1998), insect metamorphosis (Labandeira, 2011), animal flight and feathers (Wu et al., 2018; Yang et al., 2019), reproductive systems, including angiosperm flowers, amniote eggs, and the mammalian placenta (Chuong, 2013; Doyle, 2012; Roberts et al., 2016; Sauquet, 2017; Specht and Bartlett, 2009), echolocation in whales (Churchill et al., 2016; Park et al., 2016) and bats (Simmons et al., 2008), and even cognitive skills of modern man (Neubauer et al., 2018). Based on radiometric dating of the available windows of time in the fossil record, these genetic changes are believed to have happened very quickly on a macroevolutionary timescale. In order to evaluate the chances for a neo-Darwinian process to bring about such major phenotypic changes, it is important to give rough but reasonable estimates of the time it would take for a population to evolve so that the required multiple genetic changes occur.

In this paper we focus on the coordinated evolution of gene expression of existing genes, and ask the question how long time T_m it would take for a species to change the expression of m distinct genes. This corresponds to the time it would take for the required binding sites, in the regulatory sequences of m distinct genes, to evolve in a coordinated way. The microevolutionary process is then a random walk on a fitness landscape of *regulatory arrays*, that is, a random walk on $m \times L$ matrices, whose rows are the regulatory sequences of all m genes. The target of this process consists of all regulatory arrays for which at least one of the required binding sites has been reached (possibly with some mismatches) in the enhancer or promoter regions of all m genes, or equivalently, the time until the gene-specific targets of all m regulatory sequences (the rows of the array) have been reached.

In order to find the statistical distribution of T_m, we model the microevolutionary dynamics as a population of regulatory arrays, according to a multitype Moran model (Moran, 1958a, Section 3.4 of Ewens, 2004). This is a continuous time Markov process with overlapping generations, point mutations and selection, where each individual is represented by its regulatory array. The selective fitness of an individual is a function of its regulatory array; more specifically it depends on how many of the m required gene-specific targets that have been reached, in a pre-specified or arbitrary order. Since the state space of a Moran process of regulatory arrays is huge, we simplify the problem in three steps. First, we adopt a fixed state population assumption, whereby all individuals at any time point have the same regulatory array, the so-called consensus array. Second, we clump these consensus arrays into a smaller number of components, each of which has its own selective fitness. This gives rise to an array component process, where the component to which the consensus array process belongs is monitored over time. Third, we assume that this array component process is a continuous time Markov process. From this it follows that T_m has a phase-type distribution (Neuts, 1981; Asmussen et al., 1996), since it is the waiting time until a Markov process reaches an absorbing state, where all m targets have been reached. Our work is a follow-up paper of Hössjer et al. (2018), where we studied the waiting time for coordinated mutations to appear, without focusing on regulatory sequences.

The paper is organized as follows: In Section 2 we first introduce the model more casually, with all its relevant input parameters, and present the waiting time distribution in some special cases. Then in the following sections we develop our model in more detail. First we present the $m \times L$ regulatory array of nucleotides in Section 3, for each individual of the studied population, and define the time dynamics of the population in terms of a Moran model in Section 4. The fixed state population assumption is introduced in Section 5, the components of regulatory arrays are defined in Section 6, and the array component process in Section 7. This makes it possible to present the main result of the paper in Section 8, the phase-type distribution formulas for the waiting time T_m. In Section 9 we give more explicit formulas for the initial distribution and intensity matrix of the Markovian array component process, in terms of the input parameters of the model. These results are used in Section 10 in order to compute the waiting time distribution numerically for some example models. Possible extensions are discussed in Section 11, whereas further numerical and mathematical results are provided in Appendices A-D.

2. Problem definition

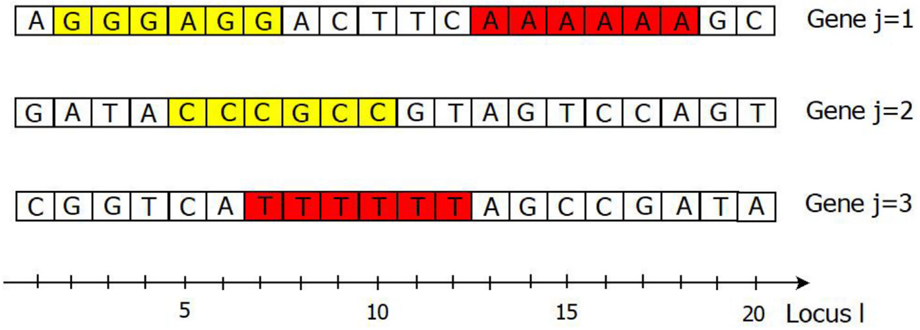
2.1. Model and its input/output parameters

Consider a haploid population with N individuals, each of which has one copy of the genome of a particular species. We will analyze the evolution of m regulatory sequences in the enhancer/promoter regions of m distinct genes in this population. Each regulatory sequence consists of L nucleotides, written in the four letter[1] alphabet

$$\mathcal{A} = \{A = \text{adenine}, C = \text{cytosine}, G = \text{guanine}, T = \text{thyamine}\}. \tag{1}$$

Typically L is of the order 1000, whereas m is a number between 1 and 10. The population has overlapping generations, and the

[1] Our framework is easily modified to accommodate other alphabets, such as binary strings with $\mathcal{A} = \{0, 1\}$.



A G G G A G G A C T T C A A A A A A G C Gene j=1

G A T A C C C G C C G T A G T C C A G T Gene j=2

C G G T C A T T T T T T A G C C G A T A Gene j=3

Fig. 1. Illustration of $m = 3$ regulatory sequences of length $L = 20$. There are $K_1 = 2$ targeted binding sites of length $W = 6$ at gene 1 ($\boldsymbol{b}_{11} = (G, G, G, G, G, G)$ and $\boldsymbol{b}_{12} = (A, A, A, A, A, A)$), $K_2 = 1$ targeted binding site at gene 2 ($\boldsymbol{b}_{21} = (C, C, C, C, C, C)$) and $K_3 = 1$ targeted binding site at gene 3 ($\boldsymbol{b}_{31} = (T, T, T, T, T, T)$). A word of length $W = 6$ that perfectly matches a targeted binding site of that row is marked with red, whereas a word with one mismatch to a targeted binding site of that row, is marked with yellow. If the order 1,2,3 of target appearance is fixed, it follows that 1 target has been reached globally if no mismatches are allowed ($\delta_{\max} = 0$), whereas 3 targets have been reached globally if one mismatch is allowed ($\delta_{\max} = 1$). On the other hand, if the order or target appearance is arbitrary, it follows that 2 targets have been reached globally if $\delta_{\max} = 0$, whereas 3 targets have been reached globally if $\delta_{\max} = 1$.

genetic composition evolves in continuous time $t \geqslant 0$ (counted in units of generations) according to a Moran model with genetic drift, selection, and mutations. Point mutations occur with a probability μ per individual, generation and nucleotide. Whenever a mutation at a particular nucleotide occurs, the probability is p_{xy} that an allele x changes into y, where $x, y \in \mathcal{A}$. Our goal is to study the waiting time T_m until transcription factors have recognized and attached to at least one of $K_j \geqslant 1$ substrings of enhancer/promoter region j, for all genes $j = 1, \ldots, m$ and all individuals in the population. Such a substring is referred to as a binding site, and its length W is typically a number between 6 and 10. For the purpose of illustration, we depict in Fig. 1 the regulatory sequences and targeted binding sites for a system with $m = 3, L = 20$, and $W = 6$.

Whenever a transcription factor attaches to one of all K_j possible binding sites of an enhancer/promoter region j, we say that the target of the corresponding gene has been reached. We will allow for a maximal mismatch $\delta_{\max}$ between a substring of the regulatory sequence of gene j, and one of its targeted binding sites, in order for the target of this gene to be reached. Typically $\delta_{\max} \in \{0, 1, 2, 3, 4\}$, assuming that the occupancy of a transcription factor on its binding site is in thermodynamic equilibrium. The transcription factors will bind to a DNA string of length W, if the binding free energy between the factor and the string is small enough, or if the binding affinity is large enough. It is often assumed that the binding free energy is proportional to, or some other increasing function of the number of mismatches (Berg and von Hippel, 1987; Fields et al., 1997, Chapter 11 of Phillips et al., 2013), and therefore binding might occur, in spite of a small number of mismatches (Durrett and Schmidt, 2007; Tuğrul et al., 2015).

We will distinguish between a local target of a particular gene j, that a transcription factor attaches to one of its binding sites, and the number of targets that have been reached for the whole system of m genes. The latter will also depend on the order of target appearance (TA). If the order of target appearance is arbitrary (TA = arbitrary), the global number of reached targets is simply the sum of the number of local targets that have been reached. On the other hand, if the targets have to appear in a fixed order $1, \ldots, m$ (TA = fixed), then the global number of reached targets is the largest number j for which all local targets $1, \ldots, j$ have been reached. A mutation is referred to as a forward mutation if it increases or does not change the number of globally reached targets, whereas a backward mutation decreases the number of globally reached targets. Forward mutations are always allowed, whereas a backward mutation is allowed with probability $0 \leqslant \gamma \leqslant 1$. We will further assume that the selective advantage or reproductive success of an individual is s_h if h of its binding site targets have been reached globally.[2] These selection (or fitness) coefficients satisfy $s_0 = 1$ and $s_h \in (0, \infty]$ for $h = 1, \ldots, m$, corresponding either to negative selection ($0 < s_h < 1$), neutrality ($s_h = 1$) and positive selection ($s_h > 1$). It is inherently difficult to estimate s_h, although it is well known that only a very small fraction of mutations are beneficial in the sense that they increase the fitness coefficients of an individual (Kimura, 1979; Betancourt, 2007; Orr, 2010). In this paper, however, we have a pre-specified target of m possible binding sites. For this reason, the selection coefficients should rather be chosen specifically for each possible application of the model, based on the predicted fitness for individuals who have $0, 1, \ldots, m$ of the binding sites fixed globally.

The input parameters of the model are listed in the upper part of Table 1. Our goal is to find how these parameters impact the distribution function of the waiting time T_m, in particular the expected value and variance. In order to get tractable expressions for the waiting time distribution, a number of approximations are developed in Sections 3–7. Readers who want to focus on the main results of this paper may first go through the examples of Section 2.2 and then proceed to Section 8, where a general phase-type distribution formula for T_m is presented.

2.2. Some introductory waiting time formulas

In the rest of Section 2 we will approximate the waiting time distribution in some special cases, in order to provide an intuitive understanding of how the input parameters of the model affect this distribution. It is possible to deduce these formulas from the general theory of Section 8, as will be shown in Section 10 and Appendix B.

In all examples of this section we will assume that mutations between all pairs of nucleotides are equally likely (Jukes and Cantor, 1969), corresponding to a transition matrix

$$\boldsymbol{P} = (p_{xy}) = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}, \tag{2}$$

with a uniform stationary distribution

$$\pi_A = \pi_C = \pi_G = \pi_T = 1/4 \tag{3}$$

of all four nucleobases.

[2] This definition of selection coefficients will be generalized at the end of Section 6.2.

Table 1
Input parameters (upper part) and output parameters (lower part).

Symbol	Definition
N	Population size.
m	Required number of coordinated targets to appear, in each of the enhancer or promoter regions of m genes, in order for a protein to be expressed.
L	Length of the regulatory sequence in each enhancer/promoter region.
W	Word length, or length of a binding site ($\in \{6, 7, 8, 9, 10\}$).
μ	Mutation probability per individual, generation and nucleotide.
$\boldsymbol{P}$	Matrix with mutation probabilities between all pairs of alleles ($= \left(p_{xy}; x, y \in \{A, C, G, T\}\right)$).
s_j	Selection coefficient when $j \in \{0, 1, \ldots, m\}$ binding site targets have been reached ($\in (0, \infty]$, with $s_0 = 1$).
γ	Probability that a backward mutation is allowed ($\in [0, 1]$).
K_j	Number of possible binding site targets in enhancer/promoter region j.
$\boldsymbol{b}_{jk}$	Binding site target number k in enhancer/promoter region j ($= (b_{jk1}, \ldots, b_{jkW})$).
$\delta_{\max}$	Maximal number of mismatching nucleotides, in order for a target to be reached ($\in \{0, 1, 2\}$).
T_m	Waiting time until m targets have been fixed in the population.
$E(T_m)$	Expected value of T_m.
$\mathrm{Var}(T_m)$	Variance of T_m.
F_{T_m}	Distribution function of T_m.
f_{T_m}	Density function of T_m.

2.2.1. One gene ($m = 1$)

To start with, we will analyze the waiting time T_1 until a single binding site at one gene is reached with perfect match ($K_1 = m = 1, \delta_{\max} = 0$). Let H_0 be the number of substrings of the regulatory sequence that perfectly match the targeted binding site at time $t = 0$. We will refer to H_0 as a hit variable, since it counts the number of substrings that hit (or belong to) a certain set of words of length W, in this case the single targeted binding site. Since the stationary distribution of the Juke-Cantor model is uniform (3), each word of length W has the same probability 4^{-W}. From this it follows that the expected value of the hit variable is $E(H_0) = L_0 4^{-W}$, where $L_0 = L - W + 1$ is the number of possible starting points of a binding site along the regulatory sequence. We have that $T_1 = 0$ if $H_0 > 0$, whereas if $H_0 = 0$ our first approximation entails that T_1 is exponentially distributed with rate

$$\lambda = L_0 3W4^{-W} \cdot \frac{N\mu}{3} \cdot \beta(s_1) \tag{4}$$

The probability that a word of length W has one mismatch with the targeted binding site is $3W4^{-W}$ for the Juke-Cantor model. Therefore, the first term $E(H_1) = L_0 3W4^{-W}$ of (4) is the expected value of H_1, the number of substrings of length W at time $t = 0$, of the regulatory sequence, with only one mismatch to the targeted binding site. We will refer to H_1 as a minus 1 hit variable, since it records how many substrings that lack one match (in this case $W - 1$ nucleotides rather than W) with the targeted binding site. The second term $N\mu/3$ of (4) is the rate at which one single word of length W, with one mismatch to the targeted binding site, mutates in *some* individual at the nucleotide where it differs from the target (at rate $N\mu$) to the targeted allele (with probability $1/3$ for the Juke-Cantor model). Finally, the third term

$$\beta(s) = \beta_N(s) = \begin{cases} 1/N, & s = 1, \\ (1 - s^{-1})/(1 - s^{-N}), & s \neq 1, \end{cases} \tag{5}$$

of (4) is a fixation probability of a Moran model (Komarova et al., 2003, Section 6.1 of Durrett, 2008), i.e. the probability that a targeted mutation with selection coefficient s gets fixed in the population, if all the other $N - 1$ individuals have selection coefficient 1 at the time when the mutation first appears. Assuming that H_0 has a Poisson distribution, it follows that T_1 is zero and positive with probabilities $P(H_0 > 0) = 1 - e^{-L_0 4^{-W}}$ and $P(H_0 = 0) = e^{-L_0 4^{-W}}$ respectively. Putting things together, the sought for waiting time

$$T_1 \overset{\mathcal{L}}{\in} \left(1 - e^{-L_0 4^{-W}}\right)\delta_0 + e^{-L_0 4^{-W}} \mathrm{Exp}(\lambda) \tag{6}$$

is approximately a mixture of a one-point distribution δ_0 at 0 and an exponential distribution with rate λ. The corresponding expected waiting time is

$$E(T_1) = \left[\mu N \beta(s_1) L_0 W 4^{-W}\right]^{-1} e^{-L_0 4^{-W}}. \tag{7}$$

A more refined approximation of T_1 is obtained by conditioning not only on whether H_0 is zero or positive, but also on the value of the minus one hit variable H_1 when $H_0 = 0$. Assume that H_1 is approximately Poisson distributed (with expected value as defined below (4)) and independent of H_0. If also the conditional distribution of T_1 satisfies $T_1 | H_1 = n \overset{\mathcal{L}}{\in} \mathrm{Exp}(n \cdot N\mu/3 \cdot \beta(s_1))$, it follows that

$$\begin{aligned} T_1 \quad & \overset{\mathcal{L}}{\in} \quad \left(1 - e^{-L_0 4^{-W}}\right)\delta_0 \\ & + \quad e^{-L_0 4^{-W}} \sum_{n=1}^{\infty} \frac{e^{-L_0 3W4^{-W}} \left(L_0 3W4^{-W}\right)^n}{n!} \mathrm{Exp}\left(\frac{n \cdot N\mu\beta(s_1)}{3}\right), \end{aligned} \tag{8}$$

a formula obtained by Durrett and Schmidt (2007) when $s_1 = \infty$ and $\beta(s_1) = 1$. Whereas (6) approximates the distribution of T_1 well when $E(H_1)$ is large (i.e. when L is large or W is small), (8) is even more accurate if s_1 is large as well. However, when $E(H_1)$ gets small and s_1 is not very large, we need a more general phase-type distribution formula for T_1. This formula should incorporate the possibility that at least two mutations are needed to reach the target ($H_0 = H_1 = 0$), as well as the number $H_2 \overset{\mathcal{L}}{\in} \mathrm{Po}\left(\frac{9}{2} W(W-1)4^{-W}\right)$ of substrings of the regulatory sequence of length W that differ from the targeted binding site at two positions at time $t = 0$. If two mutations are needed to reach the target they are either fixed one by one or through stochastic tunneling (Carter and Wagner, 2002; Komarova et al., 2003; Iwasa et al., 2004), where the second mutation occurs before the first one has been fixed. The latter option has an impact on the waiting time distribution only when s_1 is large. If s_1 is large, the resulting distribution of T_1 is a approximately a mixture of exponentials that, apart from the terms of (8), also incorporates a positive weight for the $H_0 = H_1 = 0$ event that at least two mutations are required to reach the targeted binding site (Durrett and Schmidt, 2007). On the other hand, if s_1 is not very large and two mutations are required to reach the target, then a more general phase-type distribution is needed for T_1.

The accuracy of formulas (6) and (8), as approximations of the waiting time T_1, will be further discussed in Section 10.1 and Appendix B.1.

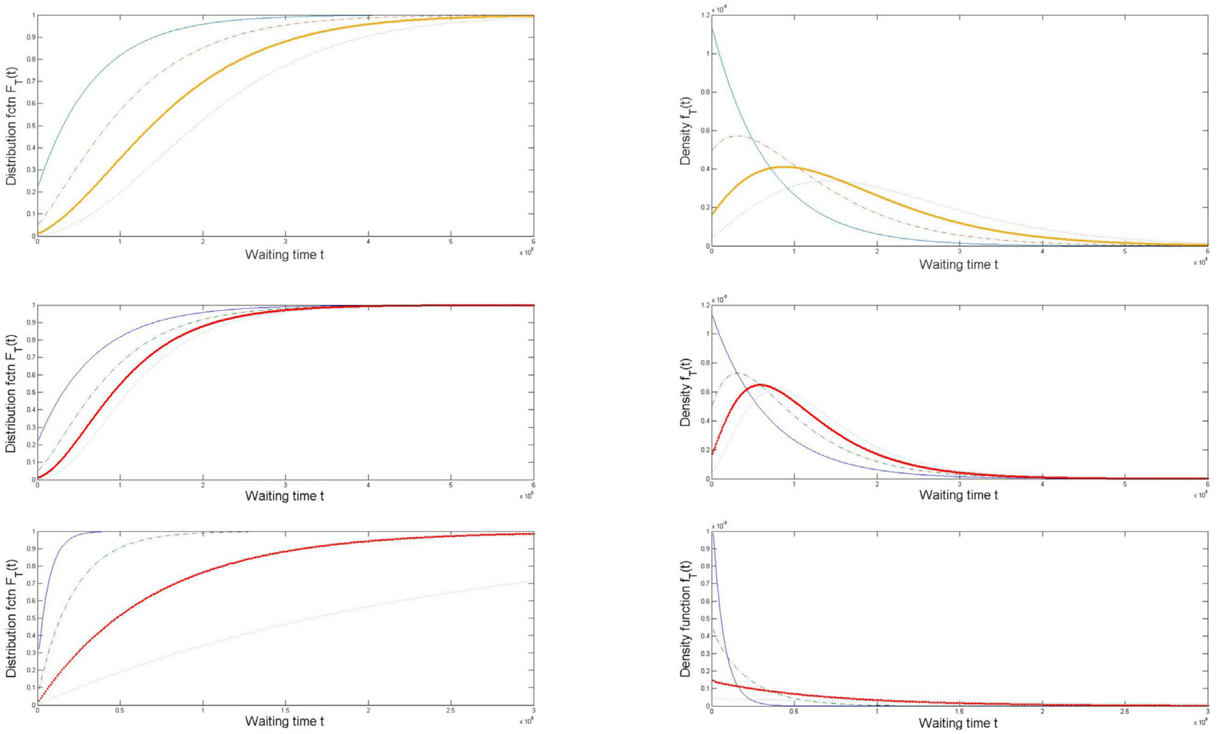


Fig. 2. Distribution functions (left) and density functions (right) for the waiting time T_m when $m = 1$ (solid), $m = 2$ (dash-dotted), $m = 3$ (thick solid) and $m = 4$ (dotted). The backward mutation probability γ and the target appearance rule (TA) are defined as follows. Upper: $\gamma = 0$ and TA = fixed order, Middle: $\gamma = 0$ and TA = arbitrary order, and Lower: $\gamma = 1$ and TA = fixed or arbitrary order. The other parameters are the same as in Table 3. In particular, the model is selectively neutral and stochastic tunneling is accounted for. Since the waiting time is larger when $\gamma = 1$, the lower row has a different scale along the x-axis ($3 \cdot 10^9$ generations) than the upper and lower subplots ($6 \cdot 10^8$ generations).

2.2.2. Several genes ($m > 1$)

Next we consider the waiting time T_m until the targets of $m > 1$ genes are reached with perfect match ($\delta_{\max} = 0$) when there is only one targeted binding site per gene ($K_j = 1$) and Juke-Cantor's mutation model (2) is used. We will assume that the model is neutral ($s_1 = \ldots = s_m = 1$). In Appendix B.1 we motivate in detail that for neutral models it suffices to summarize information about each regulatory sequence in terms of whether there is at least one hit ($H_0 > 0$) or not ($H_0 = 0$). Making use of this approximation, it can be shown very explicitly that the back mutation probability γ influences the waiting time a lot, in particular when the order of target appearance is arbitrary (TA = arbitrary). This is illustrated in the middle and lower parts of Fig. 2, with plots of the distribution and density functions of T_m, for different values of m.

In order to quantify how the parameters of the model affect T_m, we will present some formulas for $E(T_m)$. If back mutations are not allowed, $E(T_m)$ increases linearly or logarithmically with m depending on whether the order of target appearance is fixed or arbitrary. On the other hand, if back mutations are allowed, then $E(T_m)$ increases exponentially with m. In order to motivate this we will first assume that the order of target appearance TA = arbitrary, and let $\kappa_1 = \exp\left(-L_0 4^{-W}\right) = P(H_0 = 0)$ be the probability that at each regulatory sequence has no subsequence at time $t = 0$ that matches the targeted binding site. The number regulatory sequences that have not yet reached their local targets at $t = 0$ has a binomial distribution $\mathrm{Bin}(m, \kappa_1)$. Conditioning on how many genes that have reached their local targets at time $t = 0$, we motivate in Section 10.2 that

$$
\begin{aligned}
E(T_m) \quad &= \quad \left(\mu L_0 W 4^{-W}\right)^{-1} \sum_{c=1}^{m} \binom{m}{c} (1-\kappa_1)^{m-c} \kappa_1^c \\
&\quad \cdot \sum_{h=m-c}^{m-1} \sum_{k=0}^{h} \frac{\binom{m-1}{k}}{(m-k)\binom{m-1}{h}} \cdot r^{h-k} \\
&\sim \begin{cases} \left(\mu L_0 W 4^{-W}\right)^{-1} (\log(m) + 0.5772), & \gamma = 0, \\ \left(\mu L_0 W 4^{-W}\right)^{-1} (1+r)^m/(rm), & 0 < \gamma \leqslant 1, \end{cases}
\end{aligned}
\tag{9}
$$

for large m, where $a_m \sim b_m$ means that $a_m/b_m \to 1$ as $m \to \infty$, $\mu L_0 W 4^{-W}$ is the rate in (4) at which each local target appears for a neutral model, and

$$
r = \frac{\gamma P(H_0 = 0)}{P(H_0 > 0)} = \frac{\gamma e^{-L_0 4^{-W}}}{1 - e^{-L_0 4^{-W}}} \tag{10}
$$

is the ratio between the rates at which local targets are lost and acquired. Formula (9) also holds for TA = fixed when $\gamma = 1$, whereas

$$
E(T_m) = \left(\mu L_0 W 4^{-W}\right)^{-1} e^{-L_0 4^{-W}} \cdot m, \quad \text{when } \gamma = 0. \tag{11}
$$

Consequently, the mode of target appearance has no impact on the waiting time T_m when all back mutations are allowed ($\gamma = 1$). On the other hand, when no back mutations are allowed ($\gamma = 0$), it takes longer time to reach all m targets when they have to appear in a fixed order.

Table 2
List of mathematical notation. The table is divided into four parts, separated by empty lines. They correspond to generic notation (G), notation for regulatory sequences (RS), notation for regulatory arrays (RA) and notation for regulatory array components (RAC).

Category	Symbol	Definition
G	i	Number of an individual ($\in \{1,\ldots,N\}$).
	j	Number of a gene, and its associated regulatory sequence ($\in \{1,\ldots,m\}$).
	l	Nucleotide locus number ($\in \{1,\ldots,L\}$).
	k	Number of a specific binding site target within enhancer/promoter region j ($\in \{1,\ldots,K_j\}$).
	t	Time, counted in units of generations ($\geqslant 0$).
	w	Position of a letter within a word ($\in \{1,\ldots,W\}$).
	δ	Number of mismatches between a regulatory sequence and the closest targeted binding site.
	x, y	Allele, i.e. type of letter or nucleobase ($\in \mathcal{A} = \{A,C,G,T\}$).
	π_x	Probability that a nucleotide has allele x.
	$\beta(s)$	Fixation probability, for one regulatory array with selection coefficient s, when all other $N-1$ arrays in the population have selection coefficient 1.
RS	$\boldsymbol{x}, \boldsymbol{y}$	Regulatory sequence ($\boldsymbol{x} = (x_1,\ldots,x_L), \boldsymbol{y} = (y_1,\ldots,y_L)$).
	$\delta_j(\boldsymbol{x})$	Number of mismatches between regulatory sequence $\boldsymbol{x}$ and the closest targeted binding site of enhancer/promoter region j.
	$\mathcal{X}$	State space of regulatory sequences ($= \{A,C,G,T\}^L$).
	$h_j(\boldsymbol{x})$	Binary variable that indicates whether the target for regulatory sequence $\boldsymbol{x}$ in promoter region j has been reached or not.
	$\boldsymbol{X}_{tj}(i)$	Regulatory sequence of individual i in enhancer/promoter region j at time t (=$(X_{tj1}(i),\ldots,X_{tjL}(i))$).
	$\boldsymbol{X}_{tj}$	Consensus sequence of individual i in enhancer/promoter region j at time t (=$(X_{tj1},\ldots,X_{tjL})$).
RA	$\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}}$	Regulatory array ($\underline{\boldsymbol{x}} = (\boldsymbol{x}_1^T,\ldots,\boldsymbol{x}_m^T)^T = (x_{jl}), \underline{\boldsymbol{y}} = (\boldsymbol{y}_1^T,\ldots,\boldsymbol{y}_m^T)^T = \left(y_{jl}\right)$, with $1 \leqslant j \leqslant m, 1 \leqslant l \leqslant L$).
	$\underline{\mathcal{X}}$	State space of regulatory arrays ($= \mathcal{X}^m$).
	$h(\underline{\boldsymbol{x}})$	Type of array $\underline{\boldsymbol{x}}$, i.e. how many targets that have been reached globally within $\underline{\boldsymbol{x}}$ ($\in \{0,\ldots,m\}$).
	$\underline{\boldsymbol{X}}_t(i)$	Regulatory array of individual i at time t ($= \left(\boldsymbol{X}_{t1}(i)^T,\ldots,\boldsymbol{X}_{tm}(i)^T\right)^T$).
	$\underline{\boldsymbol{X}}_t$	Consensus array at time t ($= \left(\boldsymbol{X}_{t1}^T,\ldots,\boldsymbol{X}_{tm}^T\right)^T$).
	$\pi_{\underline{\boldsymbol{x}}}$	Stationary probability that the consensus array process $\underline{\boldsymbol{X}}_t$ equals $\underline{\boldsymbol{x}}$ (for a neutral model).
	$\sigma_{\underline{\boldsymbol{x}\boldsymbol{y}}}$	Transition rate, for consensus array process $\underline{\boldsymbol{X}}_t$, to jump from $\underline{\boldsymbol{x}}$ to $\underline{\boldsymbol{y}}$, when $\underline{\boldsymbol{x}} \neq \underline{\boldsymbol{y}}$.
	$\boldsymbol{\Sigma}$	Intensity matrix of the consensus array process $\underline{\boldsymbol{X}}_t$ ($= \left(\sigma_{\underline{\boldsymbol{x}},\underline{\boldsymbol{y}}}\right)_{\underline{\boldsymbol{x}},\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}}$).
RAC	$\overline{C}$	Number of components of the set $\underline{\mathcal{X}}$ of regulatory arrays ($\geqslant 2$).
	$\underline{C}$	Number of components of the set $\underline{\mathcal{X}}$ of regulatory arrays that represent a final state, where all targets have been reached ($\in \left\{1,\ldots,\overline{C}-1\right\}$).
	$\mathcal{C}$	Set of all array components ($= \left\{0,\ldots,\overline{C}-1\right\}$).
	C_j	Number of components of the set $\mathcal{X}$ of regulatory sequences for enhancer/promoter region j.
	$\underline{\mathcal{X}}_{\boldsymbol{c}}$	Component number $\boldsymbol{c}$ among all regulatory arrays (with order number $\in \mathcal{C}$, and $\boldsymbol{c} = (c_1,\ldots,c_m)$) for distance-based components).
	$g(\boldsymbol{c})$	Type of all arrays in component $\underline{\mathcal{X}}_{\boldsymbol{c}}$.
	$Z_{t\boldsymbol{c}}$	Fraction of individuals with regulatory array in component $\boldsymbol{c}$ at time t.
	$\boldsymbol{Z}_t$	Array component distribution at time t ($= \left(Z_{t0},\ldots,Z_{t,\overline{C}-1}\right)$).
	$\boldsymbol{C}_t$	Component to which the consensus array $\underline{\boldsymbol{X}}_t$ belongs at time t.
	$\kappa_{\boldsymbol{c}}$	Stationary probability (for neutral model) that the array component process $\boldsymbol{C}_t$ equals $\boldsymbol{c}$.
	$\lambda_{\boldsymbol{cd}}$	Transition rate, for array component process $\boldsymbol{C}_t$, to jump from $\boldsymbol{c}$ to $\boldsymbol{d}$, when $\boldsymbol{c} \neq \boldsymbol{d}$.
	$\boldsymbol{\Lambda}$	Intensity matrix of the array component process $\boldsymbol{C}_t$ ($= \lambda_{\boldsymbol{cd}})_{\boldsymbol{c},\boldsymbol{d}=0}^{\overline{C}-1}$).

3. Regulatory arrays

In this section we will introduce regulatory arrays and some of the notation used in the rest of the paper (see Table 2 for a summary of the most important symbols). This will make it possible to formulate in more detail the waiting time T_m until all m targets have been reached globally. At each time point $t \geqslant 0$, we denote the regulatory sequence of individual $i \in \{1,\ldots,N\}$ that influences gene j, as a vector

$$\boldsymbol{X}_{tj}(i) = \left(X_{tj1}(i),\ldots,X_{tjL}(i)\right) \in \mathcal{X} = \mathcal{A}^L,$$

whose components $X_{tjl}(i) \in \mathcal{A}$ are nucleotides (1) at L consecutive loci $l = 1,\ldots,L$. We also represent all m regulatory sequences of individual i at time t as one single regulatory array or matrix

$$\underline{\boldsymbol{X}}_t(i) = \left(\boldsymbol{X}_{t1}(i)^T,\ldots,\boldsymbol{X}_{tm}(i)^T\right)^T \in \underline{\mathcal{X}} = \mathcal{X}^m$$

of dimension $m \times L$, where T denotes matrix transposition. Each entry (j,l) of this array will be referred to as a position, and Fig. 1 illustrates a regulatory array with $m = 3, L = 20$ and $3 \times 20 = 60$ positions. In order to define the waiting time T_m, we assign for each regulatory array $\underline{\boldsymbol{x}} = \left(\boldsymbol{x}_1^T,\ldots,\boldsymbol{x}_m^T\right)^T \in \underline{\mathcal{X}}$ a number $h(\underline{\boldsymbol{x}}) \in \{0,1,\ldots,m\}$ that specifies how many gene-specific targets that have been reached globally, also referred to as the type of the array. We can also think of $h(\underline{\boldsymbol{x}})$ as the type of an individual whose DNA at the m regions is represented by the regulatory array $\underline{\boldsymbol{x}}$, and whose selection coefficient is $s_{h(\underline{\boldsymbol{x}})}$. The time until all m targets have been reached globally, for all individuals in the population, is

$$T_m = \min\{t \geqslant 0; h(\underline{\boldsymbol{X}}_t(i)) = m \text{ for } i = 1,\ldots,N\}. \tag{12}$$

For any sequence $\boldsymbol{x} = (x_1,\ldots,x_L) \in \mathcal{X}$ of enhancer/promoter region j, we define a binary function $h_j(\boldsymbol{x}) \in \{0,1\}$ that specifies whether the target has been reached (1) or not (0) locally within this sequence. The type of an individual will depend on its regulatory array $\underline{\boldsymbol{x}}$ and the order of target appearance, as

$$h(\underline{\boldsymbol{x}}) = \begin{cases} \min\{j;\, 1 \leqslant j \leqslant m, h_j(\boldsymbol{x}_j) = 0\} - 1, & \text{TA} = \text{fixed}, \\ \sum_{j=1}^{m} h_j(\boldsymbol{x}_j), & \text{TA} = \text{arbitrary}, \end{cases} \tag{13}$$

with the convention $\min \varnothing = m + 1$ (and hence $h(\underline{\boldsymbol{x}}) = m$) in the upper row of (13), when all targets have been fixed locally. It is also possible to choose h in other ways, for instance so that some gene-specific targets have to appear in a pre-determined order, whereas others may arrive in any order. In order to define the region specific target functions $h_j(\boldsymbol{x})$, we let

$$\delta_j(\boldsymbol{x}) = \min_{1\leqslant k\leqslant K_j} \min_{1\leqslant l\leqslant L_0} |\boldsymbol{x}_{l:l+W-1} - \boldsymbol{b}_{jk}| \qquad (14)$$

be the number of mismatches between $\boldsymbol{x}$ and the closest of the targeted binding sites $\boldsymbol{b}_{j1}, \ldots, \boldsymbol{b}_{jK_j}$ of region j. The vector $\boldsymbol{x}_{l:l+W-1} = (x_l, \ldots, x_{l+W-1})$ is the substring of $\boldsymbol{x}$ of length W that starts at locus l, and $|\boldsymbol{x}_{l:l+W-1} - \boldsymbol{b}_{jk}|$ the Hamming distance between $\boldsymbol{x}_{l:l+W-1}$ and $\boldsymbol{b}_{jk} = (b_{jk1}, \ldots, b_{jkW})$, i.e. the number of nucleotides at which these two words of length W differ. Since we assume that a target in $\boldsymbol{x}$ is reached when there are at most $\delta_{\max}$ mismatches between $\boldsymbol{x}$ and the closest targeted binding site, it follows that[3]

$$h_j(\boldsymbol{x}) = 1\left(\delta_j(\boldsymbol{x}) \leqslant \delta_{\max}\right). \qquad (15)$$

4. A Moran model for the population dynamics

We will use a haploid, multitype Moran model with mutations and selection in order to study the time dynamics of the regulatory arrays $\{\underline{\boldsymbol{X}}_t(i);\, i = 1, \ldots, N\}$ of all individuals in the population. More specifically, we assume the following:

1. Each individual dies independently according to a Poisson process with rate 1.
2. An individual's type $h(\underline{\boldsymbol{x}})$ is a function of its regulatory array $\underline{\boldsymbol{x}}$. When an individual dies, an offspring of some randomly chosen individual (including the one that dies) replaces it. The parent is chosen among the N individuals in the population, with probabilities proportional to their selection coefficients $s_{h(\underline{\boldsymbol{x}})}$.
3. If the parent has a regulatory array $\underline{\boldsymbol{x}} = (x_{jl})$, then the offspring in ii) is assigned an array $\underline{\boldsymbol{y}} = (y_{jl})$, where mutations $(x_{jl} \to y_{jl})$ occur independently between all loci (j, l) with probability μ. Whenever a mutation occurs, the nucleotide is changed from x_{jl} to y_{jl} with probability $p_{x_{jl}y_{jl}}$.

Although generations are overlapping, we notice from step i) that the lifetime of each individual is exponentially distributed with expected value 1, so that t is counted in units of generations and therefore μ in step iii) is the mutation rate per locus and generation. The matrix $\boldsymbol{P} = (p_{xy})$ is the transition matrix of a Markov chain, whose state space consists of the four alleles $x \in \mathcal{A}$ in (1). This matrix has zeros along the diagonal, since p_{xy} is the probability of having a change from allele x to y at a particular locus, when an offspring is formed. Assuming that $\boldsymbol{P}$ is irreducible, it has a unique stationary distribution

$$\boldsymbol{\pi} = (\pi_A, \pi_C, \pi_G, \pi_T), \qquad (16)$$

i.e. a unique probability vector satisfying $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$. The simplest example is the Jukes-Cantor model in (2), with a uniform stationary distribution (3). But it is also possible to choose

$$\boldsymbol{P} = \begin{pmatrix} 0 & \alpha_1\pi_C & \alpha_2\pi_G & \alpha_3\pi_T \\ \alpha_1\pi_A & 0 & \alpha_4\pi_G & \alpha_5\pi_T \\ \alpha_2\pi_A & \alpha_4\pi_C & 0 & \alpha_6\pi_T \\ \alpha_3\pi_A & \alpha_5\pi_C & \alpha_6\pi_G & 0 \end{pmatrix},$$

as the transition matrix of a time-reversible Markov chain in such a way that its stationary distribution coincides with a prescribed probability vector $\boldsymbol{\pi}$. There are six parameters $\alpha_1, \ldots, \alpha_6$ that can be varied with two degrees of freedom, since the four row sums of $\boldsymbol{P}$ must equal 1 (Lanave et al., 1984).

5. Fixed state approximation

Our framework in this paper is a small or intermediately sized population, such that

$$NL\mu \ll 1. \qquad (17)$$

For instance, if $L = 10^3$ and the human mutation rate $\mu = 10^{-8}$ of nuclear DNA is used (Scally and Durbin, 2012), then (17) translates to an effective size $N \ll 10^5$, which is reasonable assumption for the human population (Tenesa et al., 2007) as well as other populations of small or intermediate size. Formula (17) implies a weak-mutation condition $N\mu \ll 1$, so that the mutation rate μ is small in comparison to the inverse population size. This implies that most regulatory arrays $\underline{\boldsymbol{X}}_t(i)$ of the population will look very similar, at any time point t. By this we mean that one allele $x = X_{tjl}$ will dominate at most positions (j, l). This can be motivated in several ways. First, the nucleotide diversity π_{nucl} is the proportion of nucleotides that differ between two randomly chosen individuals, and it is well known that whenever $N\mu$ is small, π_{nucl} is of order $N\mu$, see for instance Durrett (2008, p. 39) Second, most new mutations $X_{tjl} \to y$ will be unsuccessful, that is, they will soon die out and not spread to the whole population. Moreover, whenever (17) holds, the time between two successful mutations at position (j, l) will be large in comparison to the time it takes for such mutations to spread after their first appearance, see for instance Hössjer et al. (2018) and references therein. When studying the evolution of the population as a whole, it is therefore reasonable to look at the time dynamics of a consensus array

$$\underline{\boldsymbol{X}}_t = \left(\boldsymbol{X}_{t1}^T, \ldots, \boldsymbol{X}_{tm}^T\right)^T = (X_{tjl};\, 1 \leqslant j \leqslant m, 1 \leqslant l \leqslant L), \qquad (18)$$

rather than the dynamics of the genetic composition of all arrays $\{\underline{\boldsymbol{X}}_t(i);\, i = 1, \ldots, N\}$ in the population. We will further assume that all individuals share this consensus array, a so-called *fixed state population model*. By this we mean that at any time point t, all individuals $i = 1, \ldots, N$ have the same state or regulatory array $\underline{\boldsymbol{X}}_t(i) = \underline{\boldsymbol{X}}_t \in \underline{\mathcal{X}}$. This gives a simplified definition

$$T_m = \min\{t > 0;\, h(\underline{\boldsymbol{X}}_t) = m\} \qquad (19)$$

of the waiting time in (12), until all targets have been fixed globally in the population.

We will postulate that $\{\underline{\boldsymbol{X}}_t;\, t \geqslant 0\}$ is a continuous time Markov process with state space $\underline{\mathcal{X}}$. Because of the memoryless property of Markov processes, it follows that the distribution of $\{\underline{\boldsymbol{X}}_t;\, t \geqslant 0\}$ is determined by the initial distribution of $\underline{\boldsymbol{X}}_0$ and its intensity matrix. We will assume that all loci of $\underline{\boldsymbol{X}}_0$ vary independently, with marginal distribution (16), so that for any regulatory array $\underline{\boldsymbol{x}} = \left(\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_m^T\right)^T$, we have that

$$P(\underline{\boldsymbol{X}}_0 = \underline{\boldsymbol{x}}) = \prod_{j,l} \pi_{x_{jl}} =: \prod_{j=1}^{m} \pi_{\boldsymbol{x}_j} =: \pi_{\underline{\boldsymbol{x}}}. \qquad (20)$$

The intensity matrix $\boldsymbol{\Sigma} = \left(\sigma_{\underline{\boldsymbol{x}}\underline{\boldsymbol{y}}}\right)$ of $\underline{\boldsymbol{X}}_t$ has non-diagonal entries $\sigma_{\underline{\boldsymbol{x}}\underline{\boldsymbol{y}}}$ that correspond to the rate of jumping from $\underline{\boldsymbol{x}}$ to $\underline{\boldsymbol{y}}$ when $\underline{\boldsymbol{y}} \neq \underline{\boldsymbol{x}}$, whereas $\sigma_{\underline{\boldsymbol{x}}\underline{\boldsymbol{x}}} = -\sum_{\underline{\boldsymbol{y}};\underline{\boldsymbol{y}}\neq\underline{\boldsymbol{x}}} \sigma_{\underline{\boldsymbol{x}}\underline{\boldsymbol{y}}}$ is minus the rate of leaving array $\underline{\boldsymbol{x}}$. We further assume that it is only possible to have a transition from regulatory array $\underline{\boldsymbol{x}}$ to those arrays $\underline{\boldsymbol{y}}$ that differ from $\underline{\boldsymbol{x}}$ at one or two positions.In order to motivate the first type of transition, suppose

[3] More generally, one may assume that a transcription factor binds to a binding site within regulatory sequence $\boldsymbol{x}$ with a probability that is modeled by a Fermi function $h_j(\boldsymbol{x}) = \left(1 + \exp\left(\varepsilon\left(\delta_j(\boldsymbol{x}) - \delta_0\right)\right)\right)^{-1}$, where δ_0 is the number of mismatches for which the binding probability equals 1/2, whereas $0 < \varepsilon < \infty$ quantifies how sensitive the binding probability is to the number of mismatches (Berg and von Hippel, 1987). We may regard (15) as a limiting Fermi function when $\delta_0 = \delta_{\max} + 1/2$ and $\varepsilon \to \infty$.

that the current consensus array of the population is $\underline{\boldsymbol{X}}_t = \underline{\boldsymbol{x}}$. If a mutation occurs, then the array $\underline{\boldsymbol{y}}$ of the offspring satisfies $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 1$, with $x_{jl} \neq y_{jl}$ at precisely one position (j, l).The probability for a mutation $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{y}}$ to happen in *some* individual is $N\mu p_{x_{jl}y_{jl}}$, so that $N - 1$ individuals have type $\underline{\boldsymbol{x}}$ and 1 individual has type $\underline{\boldsymbol{y}}$. Ignoring the possibility of other mutations, we then use that the fixation probability of $\underline{\boldsymbol{y}}$, when initially all other individuals carry array $\underline{\boldsymbol{x}}$, is $\beta\left(s_{h(\underline{\boldsymbol{y}})}/s_{h(\underline{\boldsymbol{x}})}\right)$, with $\beta(\cdot)$ as defined in (5). Finally,if $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{y}}$ is a back mutation ($h\left(\underline{\boldsymbol{y}}\right) < h(\underline{\boldsymbol{x}})$), it is accepted with probability γ, whereas a forward mutation ($h\left(\underline{\boldsymbol{y}}\right) \geqslant h(\underline{\boldsymbol{x}})$) is always accepted. In Appendix C.1 we motivate that the above-mentioned three steps (mutation, fixation, and acceptance of a back mutation) can be combined into one single rate

$$\sigma_{\underline{\boldsymbol{xy}}} = N\mu\beta\left(\frac{s_{h(\underline{\boldsymbol{y}})}}{s_{h(\underline{\boldsymbol{x}})}}\right)\gamma^{1\left(h\left(\underline{\boldsymbol{y}}\right)<h\left(\underline{\boldsymbol{x}}\right)\right)}p_{x_{jl}y_{jl}} \tag{21}$$

of changing the fixed state of the population from $\underline{\boldsymbol{x}}$ to $\underline{\boldsymbol{y}}$ whenever $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 1$ and $x_{jl} \neq y_{jl}$, using a convention $0^0 = 1$ in (21) when $\gamma = 0$.

In order to motivate the second type of transition of the consensus array process, from fixed state $\underline{\boldsymbol{X}}_t = \underline{\boldsymbol{x}}$ to another fixed state $\underline{\boldsymbol{y}}$ with $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 2$, assume that $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ differ at the two positions (j_1, l_1) and (j_2, l_2). A transition from $\underline{\boldsymbol{x}}$ to $\underline{\boldsymbol{y}}$ must therefore occur in two steps, along either of the two paths $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{v}} \to \underline{\boldsymbol{y}}$ or $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{w}} \to \underline{\boldsymbol{y}}$, where $\underline{\boldsymbol{v}}$ and $\underline{\boldsymbol{w}}$ differ from $\underline{\boldsymbol{x}}$ at positions (j_1, l_1) and (j_2, l_2) respectively. It follows that the transition rate $\sigma_{\underline{\boldsymbol{xy}}}$ from $\underline{\boldsymbol{x}}$ to $\underline{\boldsymbol{y}}$ will be the sum of the rates at which changes appear along each of these two paths. In order to find the appearance rate of these two paths we need to incorporate stochastic tunneling. Since there is yet no general theory of stochastic tunneling for multitype Moran processes, we will give rather crude estimates of the appearance rate of $\underline{\boldsymbol{y}}$ along each one of these two paths. Starting with path $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{v}} \to \underline{\boldsymbol{y}}$, we must first have a mutation $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{v}}$ at rate $N\mu p_{x_{j_1 l_1} y_{j_1 l_1}}$, so that the composition of the population changes to $N - 1$ individuals with array $\underline{\boldsymbol{x}}$ and one individual with array $\underline{\boldsymbol{v}}$. Then, before the newly mutated array $\underline{\boldsymbol{v}}$ either dies out or increases to a high frequency, it must have at least one mutated offspring $\underline{\boldsymbol{v}} \to \underline{\boldsymbol{u}}$ (the tunneling event), whose descendants then spread to the whole population. The array $\underline{\boldsymbol{u}}$ that spreads to the whole population must then equal $\underline{\boldsymbol{y}}$ and finally, be accepted with probability $\gamma^{1\left(h\left(\underline{\boldsymbol{y}}\right)<h\left(\underline{\boldsymbol{x}}\right)\right)}$. Our expression for the appearance rate of $\underline{\boldsymbol{y}}$ along path $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{v}} \to \underline{\boldsymbol{y}}$ will involve $r_{\underline{\boldsymbol{xv}}}$, which is the ratio of the probability that $\underline{\boldsymbol{v}}$ has *some* offspring $\underline{\boldsymbol{u}}$ that first appears, gets mutated at position $(j, l) \neq (j_1, l_1)$ and then gets fixed, divided by the probability that a *randomly chosen* such mutated offspring $\underline{\boldsymbol{u}}$ gets fixed, *given* that $\underline{\boldsymbol{u}}$ first appeared and was mutated. In simpler words, $r_{\underline{\boldsymbol{xv}}}$ quantifies how simple it is for a stochastic tunneling event $\underline{\boldsymbol{v}} \to \underline{\boldsymbol{u}}$ to occur, once $\underline{\boldsymbol{v}}$ exists in the population, in comparison to the chances of a single copy of $\underline{\boldsymbol{u}}$ to get fixed, once $\underline{\boldsymbol{u}}$ exists in the population. A similar argument applies for the appearance rate of $\underline{\boldsymbol{y}}$ along the other path $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{w}} \to \underline{\boldsymbol{y}}$, and this rate will involve the analogous quantity $r_{\underline{\boldsymbol{xw}}}$. By adding the rates of the two paths $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{v}} \to \underline{\boldsymbol{y}}$ and $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{w}} \to \underline{\boldsymbol{y}}$, we motivate in Appendix C.1 that

$$\sigma_{\underline{\boldsymbol{xy}}} = N\mu(mL-1)^{-1}\left(r_{\underline{\boldsymbol{xv}}} + r_{\underline{\boldsymbol{xw}}}\right)\beta\left(\frac{s_{h(\underline{\boldsymbol{y}})}}{s_{h(\underline{\boldsymbol{x}})}}\right)\gamma^{1\left(h\left(\underline{\boldsymbol{y}}\right)<h\left(\underline{\boldsymbol{x}}\right)\right)}p_{x_{j_1 l_1} y_{j_1 l_1}} p_{x_{j_2 l_2} y_{j_2 l_2}} \tag{22}$$

when $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 2$ and $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ differ at positions (j_1, l_1) and (j_2, l_2).

For a neutral model ($s_0 = s_1 = \ldots = s_m = 1$) with all back mutations allowed ($\gamma = 1$), we show in Appendix C.1 that (21)-(22) simplify to

$$\sigma_{\underline{\boldsymbol{xy}}} = \begin{cases} \mu p_{x_{jl}y_{jl}}, & |\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 1, \\ 2N^{1/2}\mu^{3/2}(mL-1)^{-1/2}p_{x_{j_1 l_1} y_{j_1 l_1}} p_{x_{j_2 l_2} y_{j_2 l_2}}, & |\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 2, \end{cases} \tag{23}$$

since $\beta\left(s_{h(\underline{\boldsymbol{y}})}/s_{h(\underline{\boldsymbol{x}})}\right) = 1/N$ and $r_{\underline{\boldsymbol{xv}}} = r_{\underline{\boldsymbol{xw}}} = \sqrt{N\mu(mL-1)}$. It follows from (17) and (23) that double mutations can be ignored for neutral models. Moreover, we notice from (23) that $\underline{\boldsymbol{X}}_t$ is a stationary Markov process for a neutral model where all back mutations are allowed, with marginal distribution (20) for all $t \geqslant 0$. If some back mutations are not allowed ($\gamma < 1$) or if the regulatory arrays have a fitness that varies with type ($s_h \neq 1$ for at least one type h), (20) will no longer be the marginal distribution for $t > 0$. However, we will still use this equation as an approximation of $P(\underline{\boldsymbol{X}}_t = \underline{\boldsymbol{x}})$.

6. Regulatory array components

The state space $\underline{\mathcal{X}}$ of regulatory arrays is huge, of size 4^{mL}. It is therefore convenient to decompose it into a smaller number of disjoint *array components*. Let $\overline{C}$ refer to the total number of such components, and write

$$\underline{\mathcal{X}} = \cup_{\boldsymbol{c}=0}^{\overline{C}-1}\underline{\mathcal{X}}_{\boldsymbol{c}}, \tag{24}$$

where $\boldsymbol{c}$ is a vector (see below) that indexes the partition. It is assumed in (24) that these vectors are ordered in some way into a set

$$\mathcal{C} = \left\{0, \ldots, \overline{C} - 1\right\}. \tag{25}$$

With some abuse of notation, we will therefore refer to $\boldsymbol{c}$ in two ways; as a vector that summarizes information from the m regulatory sequences, and as a scalar order number of these vectors. We further assume that all arrays within each component of (24) are of the same type $g(\boldsymbol{c})$, i.e.

$$\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{c}} \Rightarrow h(\underline{\boldsymbol{x}}) = h\left(\underline{\boldsymbol{y}}\right) = g(\boldsymbol{c}), \tag{26}$$

and moreover, that the arrays within the first $\underline{C}$ components ($1 \leqslant \underline{C} < \overline{C}$) have reached all targets, i.e.

$$\begin{array}{rcll} g(\boldsymbol{c}) & = & m, & \text{for } 0 \leqslant \boldsymbol{c} \leqslant \underline{C} - 1, \\ g(\boldsymbol{c}) & < & m, & \text{for } \underline{C} \leqslant \boldsymbol{c} \leqslant \overline{C} - 1. \end{array} \tag{27}$$

We will refer to $g(\boldsymbol{c})$ as the type of array component $\underline{\mathcal{X}}_{\boldsymbol{c}}$.

6.1. A coarse regulatory array decomposition

The coarsest possible decomposition (24) of $\underline{\mathcal{X}}$ that satisfies (26)–(27) has only $\overline{C} = m + 1$ components, one for each type of array, and only $\underline{C} = 1$ of these components consists of arrays for which all targets have been reached. This corresponds to

$$\underline{\mathcal{X}}_{\boldsymbol{c}} = \{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}};\, h(\underline{\boldsymbol{x}}) = m - \boldsymbol{c}\}, \tag{28}$$

for $\boldsymbol{c} = 0, \ldots, m$, so that $\boldsymbol{c}$ represents the number of remaining targets to be reached globally. We will use the notation

$$\mathcal{C}_{\text{red}} = \{0, 1, \ldots, m\} \tag{29}$$

for this maximally reduced set of array component indices.

6.2. A regulatory array decomposition based on number of mismatches to closest binding sites

In this subsection we consider another decomposition (24) of $\underline{\mathcal{X}}$, which is more complex but also more useful than (28). We first write the state space of regulatory sequences, in the enhancer/promoter region of each gene j, as a union

$$\mathcal{X} = \cup_{c=1}^{C_j} \mathcal{X}_{jc} \tag{30}$$

of C_j components. By taking the direct product of (30), for all genes j, we then decompose the space $\underline{\mathcal{X}}$ of all regulatory arrays into $\overline{C} = \prod_{j=1}^{m} C_j$ components

$$\underline{\mathcal{X}}_{\boldsymbol{c}} = \underline{\mathcal{X}}_{(c_1,\ldots,c_m)} = \mathcal{X}_{1c_1} \times \ldots \times \mathcal{X}_{mc_m}, \tag{31}$$

that correspond to all vectors in the set

$$\mathcal{C}_{\text{mism}} = \left\{\boldsymbol{c} = (c_1, \ldots, c_m);\ 0 \leqslant c_j \leqslant C_j - 1\text{for } j = 1, \ldots, m\right\} \tag{32}$$

that is a vectorized version of (25). Typically, the smaller $0 \leqslant c \leqslant C_j - 1$ in (30) is, the fewer mismatches the sequences in $\mathcal{X}_{jc}$ have with the local target of this region. We will make this more precise by dividing $\{0, 1, \ldots, W\}$ into C_j disjoint, ordered, and connected subsets $\Delta_{j0} < \Delta_{j1} < \ldots < \Delta_{j,C_j-1}$ of distances to the local target of region j, with $0 \in \Delta_{j0}$ and $W \in \Delta_{j,C_j-1}$. This is achieved by introducing numbers $0 = \delta_{j0} < \delta_{j1} < \ldots < \delta_{jC_j} = W + 1$ that define the end points

$$\Delta_{jc} = \left\{\delta_{jc}, \ldots, \delta_{j,c+1} - 1\right\} \tag{33}$$

of these subsets, for $c = 0, \ldots, C_j - 1$. Then

$$\mathcal{X}_{jc} = \left\{\boldsymbol{x};\ \delta_j(\boldsymbol{x}) \in \Delta_{jc}\right\} \tag{34}$$

consists of all regulatory sequences $\boldsymbol{x}$ whose number of mismatches (14) with the closest targeted binding site of enhancer/promoter region j, belongs to the set Δ_{jc}. We further assume that the regulatory sequences within the first $C'_j < C_j$ components $(\mathcal{X}_{jc}, c = 0, \ldots, C'_j - 1)$ of region j have reached the target of this region. In view of (15), this is equivalent to requiring that $\delta_{jC'_j} = \delta_{\max} + 1$, so that $\Delta_{jC'_j} = \left\{\delta_{\max} + 1, \ldots, \delta_{jC'_j+1} - 1\right\}$. From this it follows that there are $\underline{C} = \prod_{j=1}^{m} C'_j$ array components for which the targets of all m genes have been reached.

Let $g_j(c) = 1\left(0 \leqslant c < C'_j\right)$ be the indicator function for whether the target has been reached or not, locally in region j. Because of (15), we then have that $h_j(\boldsymbol{x}) = g_j(c)$ for all $\boldsymbol{x} \in \mathcal{X}_{jc}$. It follows from (13) and (26) that the type of array component $\underline{\mathcal{X}}_{\boldsymbol{c}}$ can be written as

$$g(\boldsymbol{c}) = \begin{cases} \min\{j;\ 1 \leqslant j \leqslant m, g_j(c_j) = 0\} - 1, & \text{TA} = \text{fixed}, \\ \sum_{j=1}^{m} g_j(c_j), & \text{TA} = \text{arbitrary}. \end{cases} \tag{35}$$

An important special case of (33)–(34) is when component 0 of enhancer/promoter region j consists of those regulatory regions $\boldsymbol{x}$ that reached the target within region, i.e.

$$\mathcal{X}_{j0} = \left\{\boldsymbol{x};\ h_j(\boldsymbol{x}) = 1\right\}. \tag{36}$$

This corresponds to having $C'_j = 1$ and $\Delta_{j0} = \{0, \ldots, \delta_{\max}\}$. Then $\underline{C} = 1$, since only component $\boldsymbol{c} = (0, \ldots, 0)$ will consist of arrays for which all m targets have been reached (see Fig. 3 for an illustration).

The gene-specific decomposition (31) of regulatory arrays has another advantage. It allows for a more general definition of selection coefficients, whereby all regulatory arrays $\underline{\boldsymbol{x}}$ within component $\underline{\mathcal{X}}_{\boldsymbol{c}} = \underline{\mathcal{X}}_{(c_1,\ldots,c_m)}$ are assigned the same selection coefficient $s_{\boldsymbol{c}} = s_{c_1,\ldots,c_m}$. Such a model is biologically very flexible, allowing for up to $\overline{C} = \prod_{j=1}^{m} C_j$ different selection coefficients, whose values not only depend on which local targets that have been reached, but also how far away the regulatory sequences of the other genes are from reaching their local target. In particular, when $C_j \equiv 2$ we get a fitness landscape of binary strings, with analogues to spin glass models of physics (Kauffman and Levin, 1987).

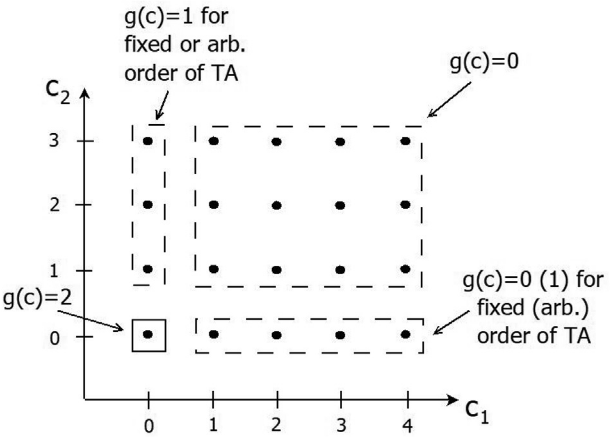


Fig. 3. Illustration of fixed and arbitrary order of target appearance (TA) for a system of $m = 2$ genes, with $C_1 = 5$ components for the regulatory sequence of gene 1, and $C_2 = 4$ components for the regulatory sequence of gene 2. Only the first component of each gene corresponds to a reached target, so that $C'_1 = C'_2 = 1$. Values of the type $g(\boldsymbol{c})$ of the array component vector $\boldsymbol{c} = (c_1, c_2)$ are shown for each mode of TA, as defined in (35).

6.3. A regulatory array decomposition based on number of mismatches to closest binding sites and hit variables

Recall from Section 2.2 that in order to improve the approximation of the distribution of the waiting time T_m, it was possible to condition on hit variables. These variables

$$H_{\mathcal{B}}(\boldsymbol{x}) = \sum_{l=1}^{L_0} 1(\boldsymbol{x}_{l:l+W-1} \in \mathcal{B}) \tag{37}$$

count the number of substrings of length W of a regulatory sequence $\boldsymbol{x}$, that belong to a certain set of words $\mathcal{B} \subset \mathcal{B}_{\text{all}} = \mathcal{A}^W = \{A, C, G, T\}^W$. In particular, we will focus on the sets

$$\mathcal{B}_{jc} = \left\{\boldsymbol{b} \in \mathcal{B}_{\text{all}};\ \min_{1 \leqslant k \leqslant K_j} |\boldsymbol{b} - \boldsymbol{b}_{jk}| \in \Delta_{jc}\right\} \tag{38}$$

of words of length W, whose distance to the closest targeted binding site of regulatory region j belong to interval (33). Notice that $\boldsymbol{x} \in \mathcal{X}_{jc}$ if and only if $H_{\mathcal{B}_{jd}}(\boldsymbol{x}) = 0$ for $d = 0, \ldots, c - 1$ and $H_{\mathcal{B}_{jc}}(\boldsymbol{x}) > 0$. It is possible to obtain a finer decomposition of the regulatory sequences of gene j than in (34), by also recording information about the hit variable $H_{\mathcal{B}_{jc}}(\boldsymbol{x})$. For any (j, c), let $\mathcal{H}_{jc1} < \ldots < \mathcal{H}_{jcn_{jc}}$ refer to an ordered decomposition of the positive integers into a finite union of n_{jc} connected subsets. This gives rise to sets

$$\mathcal{X}_{jcn} = \left\{\boldsymbol{x};\ \delta_j(\boldsymbol{x}) \in \Delta_{jc} \text{ and } H_{\mathcal{B}_{jc}}(\boldsymbol{x}) \in \mathcal{H}_{jcn}\right\}$$

of regulatory sequences $\boldsymbol{x}$ of gene j whose distance to the closest targeted binding site of this gene belongs to Δ_{jc}, whereas the number of substrings of $\boldsymbol{x}$ of length W with such a distance to the target, belongs to $\mathcal{H}_{jcn}$. Then

$$\mathcal{X} = \cup_{c=0}^{C_j-1} \cup_{n=1}^{n_{jc}} \mathcal{X}_{jcn} \tag{39}$$

is identical to (30) when $n_{jc} = 1$ and $\mathcal{H}_{jc1} = \{1, 2, \ldots\}$ for $c = 0, \ldots, C_j - 1$, but it gives a finer decomposition of $\mathcal{X}$ than (30) if at least one n_{jc} exceeds 1. When hit variables are taken into account, the set of possible indeces of the array components, is

$$\begin{aligned}\mathcal{C}_{\text{mismhit}} \quad = \quad & \{\boldsymbol{c} = (\boldsymbol{c}', \boldsymbol{n}); \boldsymbol{c}' = (c_1, \ldots, c_m), 0 \leqslant c_j \leqslant C_j - 1, \\ & \boldsymbol{n} = (n_1, \ldots, n_m), 1 \leqslant n_j \leqslant n_{jc_j} \text{ for } j = 1, \ldots, m\}.\end{aligned} \tag{40}$$

The corresponding decomposition

$$\underline{\mathcal{X}}_{\boldsymbol{c}} = \underline{\mathcal{X}}_{\boldsymbol{c}',\boldsymbol{n}} = \underline{\mathcal{X}}_{(c_1,\ldots,c_m),(n_1,\ldots,n_m)} = \mathcal{X}_{1c_1n_1} \times \ldots \times \mathcal{X}_{mc_mn_m} \tag{41}$$

into array components $\boldsymbol{c} \in \mathcal{C}_{\text{mismhit}}$ becomes quite complicated when the number of genes m gets large. However, it is sometimes needed in order to improve the approximation of the distribution for the waiting time T_m.

In order to illustrate the usefulness of the finer decomposition (39) of regulatory sequences, in comparison to the coarser decomposition (30), we return to the example of Section 2.2.1, where the waiting time T_1 until a single binding site at $m = 1$ gene was studied. We presented two formulas (6) and (8) for the distribution of T_1, of which the latter is more accurate. Both of these formulas correspond to having $C_1 = 3$ distance intervals $\Delta_{10} = \{0\}, \Delta_{11} = \{1\}$ and $\Delta_{12} = \{2, \ldots, W\}$ to the targeted binding site. Neither (6) nor (8) make use of any further information about the hit variable that records exact hits to the target, corresponding to Δ_{10} ($n_{10} = 1$), nor about the hit variable that records a distance of at least two from the targeted binding site, corresponding to Δ_{12} ($n_{12} = 1$). However, the two formulas treat the number hits to the set Δ_{11} of one mismatch differently. Whereas (6) simply records whether there is at least one such hit or not ($n_{11} = 1$), a truncated version of formula (8) corresponds to an exact recording ($\mathcal{H}_{11n} = \{n\}$ for $n = 1, \ldots, n_{11} - 1$) of the hit variable $H_{\mathcal{B}_{11}}(\boldsymbol{x})$ up to a truncating threshold n_{11} ($\mathcal{H}_{11n_{11}} = \{n_{11}, \ldots\}$). This will example be further discussed in Appendix B.1.

7. Regulatory array component process

In this section we will investigate how the array components of Section 6 can be used to simplify the Moran model $\{\underline{\boldsymbol{X}}_t(i); i = 1, \ldots, N\}$ of Section 4 as well as the consensus array process $\underline{\boldsymbol{X}}_t$ of Section 5.

7.1. No fixed state assumption

We will start with the general case, without any fixed state population assumption. It is helpful then to introduce the array component distribution

$$\boldsymbol{Z}_t = \left(Z_{t0}, \ldots, Z_{t,\overline{C}-1}\right) \in \mathcal{Z} \tag{42}$$

at time t, whose coordinate

$$Z_{t\boldsymbol{c}} = N^{-1}|\{i;\ 1 \leqslant i \leqslant N, \underline{\boldsymbol{X}}_t(i) \in \underline{\mathcal{X}}_{\boldsymbol{c}}\}| \tag{43}$$

refers to the fraction of individuals at time t that have their regulatory arrays in $\underline{\mathcal{X}}_{\boldsymbol{c}}$. The state space $\mathcal{Z}$ of the $\boldsymbol{Z}_t$-process in (42) is the intersection between the $\left(\overline{C} - 1\right)$-simplex (see Fig. 4)

$$\Xi = \left\{ \boldsymbol{z} = (z_0, \ldots, z_{\overline{C}-1}); z_{\boldsymbol{c}} \geqslant 0, \sum_{\boldsymbol{c}=0}^{\overline{C}-1} z_{\boldsymbol{c}} = 1 \right\} \tag{44}$$

spanned by $\boldsymbol{e}_1 = (1,0,\ldots,0), \boldsymbol{e}_2 = (0,1,0,\ldots,0), \ldots, \boldsymbol{e}_{\overline{C}} = (0,\ldots,0,1)$, and the set $\mathbb{N}^{\overline{C}}/N$ of vectors $\boldsymbol{z}$ whose coordinates are natural numbers divided by N. It follows from (43) that (12) is equivalent to

$$T_m = \inf\{t \geqslant 0; \boldsymbol{Z}_{t\boldsymbol{c}} = 0 \text{ for all } \boldsymbol{c} \ \text{ with } g(\boldsymbol{c}) < m\}. \tag{45}$$

7.2. Fixed state assumption

As a next step, we will combine (24), the collection of regulatory arrays into components, with the fixed state population model (18), whereby all individuals have the same regulatory array $\underline{\boldsymbol{X}}_t$ at each time point. To this end, we introduce a continuous time process

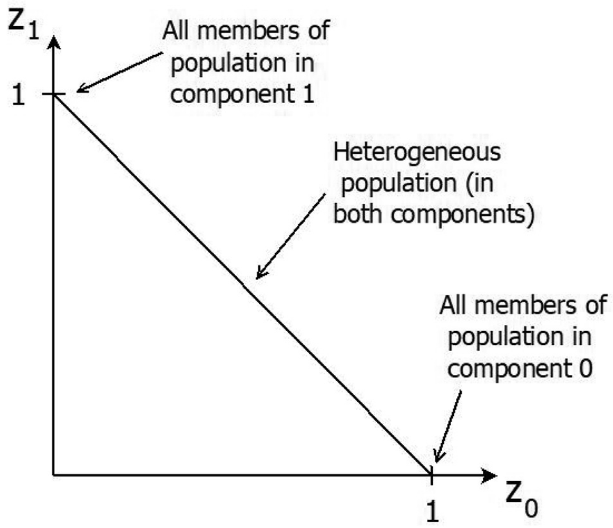

Z1
1
All members of population in component 1
Heterogeneous population (in both components)
All members of population in component 0
1
Z0

Fig. 4. Illustration of the 1-simplex $\Xi = \{\boldsymbol{z} = (z_0, z_1); z_0, z_1 \geqslant 0, z_0 + z_1 = 1\}$ in (44), the set possible states of the array component distribution process $\boldsymbol{Z}_t$, when there are $\overline{C} = 2$ array components $\underline{\mathcal{X}}_0$ and $\underline{\mathcal{X}}_1$. The two states $\boldsymbol{e}_1 = (1,0)$ and $\boldsymbol{e}_2 = (0,1)$ correspond to all individuals having their regulatory arrays in $\underline{\mathcal{X}}_0$ and $\underline{\mathcal{X}}_1$ respectively. All individuals have reached the target (belong to $\underline{\mathcal{X}}_0$) when $\boldsymbol{Z}_t = \boldsymbol{e}_1$.

$$\boldsymbol{C}_t \in \mathcal{C} \tag{46}$$

that for each time point $t \geqslant 0$ monitors which array component $\underline{\mathcal{X}}_{\boldsymbol{c}}$ in (24) that $\underline{\boldsymbol{X}}_t$ belongs to, with the state space $\mathcal{C}$ of $\boldsymbol{C}_t$ defined in (25). This can be phrased as

$$\underline{\boldsymbol{X}}_t \in \underline{\mathcal{X}}_{\boldsymbol{C}_t}. \tag{47}$$

It is also possible to express $\boldsymbol{C}_t$ in terms of the $\boldsymbol{Z}_t$-process in (42). Indeed, since the fixed state approximation implies that all individuals have the same regulatory array at any time point, $\boldsymbol{Z}_t$ will jump between the $\overline{C}$ edge points $\boldsymbol{e}_0, \ldots, \boldsymbol{e}_{\overline{C}-1}$ of $\mathcal{Z}$. The array component process $\boldsymbol{C}_t$ tells which of these edge points that is visited at time t, i.e.

$$\boldsymbol{Z}_t = \boldsymbol{e}_{\boldsymbol{C}_t}. \tag{48}$$

If the distance-based array components of Section 6.2 are used, then $\boldsymbol{c} = (c_1, \ldots, c_m)$ as in (31) and (34). Consequently, $\underline{\boldsymbol{x}} = \left(\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_m^T\right)^T \in \underline{\mathcal{X}}_{\boldsymbol{c}}$ whenever the number of mismatches between regulatory sequence $\boldsymbol{x}_j$ and the closest targeted binding site of enhancer/promoter region j belongs to the set Δ_{jc_j}, for $j = 1, \ldots, m$. This makes it possible to express

$$\boldsymbol{C}_t = (C_{t1}, \ldots, C_{tm}) \in \mathcal{C}_{\text{mism}} \tag{49}$$

as a vector-valued process that keeps track of which component $\mathcal{X}_{j,c_j}$ the regulatory sequence $\boldsymbol{X}_{tj}$ of region j belongs to at time t, for $j = 1, \ldots, m$, with $\mathcal{C}_{\text{mism}}$ as defined in (32). As in Section 6, we will also assume that the $\overline{C} = \prod_j C_j$ states of $\boldsymbol{C}_t$ are linearly ordered with a state space $\mathcal{C}$ as in (25) (see Fig. 5 for an illustration).

If the finer partition (40)–(41) of array components from Section 6.3 is used, then

$$\boldsymbol{C}_t = (C_{t1}, \ldots, C_{tm}, n_{t1}, \ldots, n_{tm}) \in \mathcal{C}_{\text{mismhit}} \tag{50}$$

also conveys information about the number of substrings of length W each regulatory sequence j has that hit the sets $\mathcal{B}_{jC_{tj}}$.

Recall, from the end of Section 5, the assumption that the array component process $\underline{\boldsymbol{X}}_t$ is a stationary Markov process. Although $\boldsymbol{C}_t$

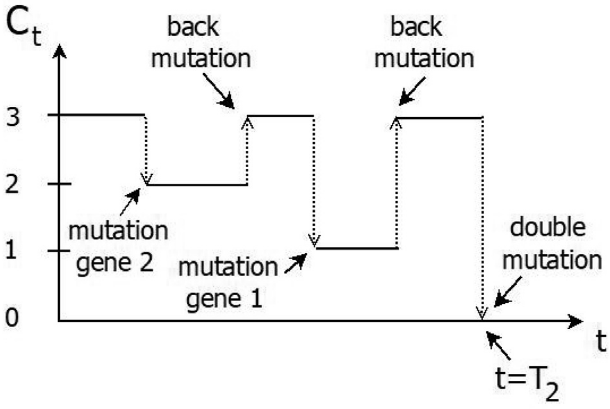


Fig. 5. Illustration of the array component process $\boldsymbol{C}_t = (C_{t1}, C_{t2})$, for a system of $m = 2$ genes, with $C_1 = C_2 = 2$ regulatory sequence components per gene. This implies that $\boldsymbol{C}_t$ has $\overline{C} = \prod_{j=1}^2 C_j = 4$ states, which are numbered as $0 = (0,0), 1 = (0,1), 2 = (1,0)$, and $3 = (1,1)$. The target is reached locally at gene j when $C_{tj} = 0$, and for the whole system when $C_{t1} = C_{t2} = 0$. Thus there is $\underline{C} = 1$ absorbing state $(0,0)$ and three non-absorbing states $(0,1), (1,0), (1,1)$. The waiting time T_2 is the time until the absorbing state is reached.

is a function $\underline{\boldsymbol{X}}_t$, it is not itself a Markov process, unless a certain rather strong lumpability condition is fulfilled (Kemeny and Snell, 1976). We will however make the simplifying assumption that $\boldsymbol{C}_t$ is also a stationary Markov process. The accuracy of this approximation will depend on how much the transition rates of $\underline{\boldsymbol{X}}_t$ vary within the different array components (see Appendix C.4 for details). In order to find the marginal distribution of $\boldsymbol{C}_t$, we recall from (20), (23), and the discussion below this equation, that the marginal distribution of $\underline{\boldsymbol{X}}_t$ is $\pi_{\underline{\boldsymbol{x}}} = \prod_j \pi_{\boldsymbol{x}_j}$. This gives a marginal distribution of $\boldsymbol{C}_t$ that equals

$$\kappa_{\boldsymbol{c}} = P(\boldsymbol{C}_t = \boldsymbol{c}) = \sum_{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}} \pi_{\underline{\boldsymbol{x}}} = \begin{cases} \prod_{j=1}^m \pi\left(\mathcal{X}_{jc_j}\right), & \text{with } \boldsymbol{c} \text{ as in } (31), \\ \prod_{j=1}^m \pi\left(\mathcal{X}_{jc_jn_j}\right), & \text{with } \boldsymbol{c} \text{ as in } (41), \end{cases} \tag{51}$$

where $\pi(\mathcal{X}') = \sum_{\boldsymbol{x} \in \mathcal{X}'} \pi_{\boldsymbol{x}}$ for any subset $\mathcal{X}' \subset \mathcal{X}$ of regulatory sequences. The intensity matrix $\boldsymbol{\Lambda} = (\lambda_{\boldsymbol{cd}})_{\boldsymbol{c},\boldsymbol{d}=0}^{\overline{C}-1}$ contains transition rates of $\boldsymbol{C}_t$, with

$$\lambda_{\boldsymbol{cd}} = \frac{1}{\kappa_{\boldsymbol{c}}} \sum_{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}} \pi_{\underline{\boldsymbol{x}}} \sum_{\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}}} \sigma_{\underline{\boldsymbol{xy}}} \tag{52}$$

the rate of $\boldsymbol{C}_t$ jumping from array component $\boldsymbol{c}$ to another component $\boldsymbol{d} \neq \boldsymbol{c}$, whereas $\lambda_{\boldsymbol{cc}} = -\sum_{\boldsymbol{d};\boldsymbol{d}\neq\boldsymbol{c}} \lambda_{\boldsymbol{cd}}$ is minus the rate of leaving $\boldsymbol{c}$. Equivalently, we interpret (52) as the approximate rate at which $\underline{\boldsymbol{X}}_t$ switches from $\underline{\mathcal{X}}_{\boldsymbol{c}}$ to $\underline{\mathcal{X}}_{\boldsymbol{d}}$ at stationarity.

The array component process $\boldsymbol{C}_t$ can be used in order to study the waiting time (19) until all targets have been reached in the population. Indeed, it follows from (45) and (48) that this waiting time can be rephrased as

$$T_m = \min\{t > 0; g(\boldsymbol{C}_t) = m\}. \tag{53}$$

In the next section we will use (53) in order to give an explicit formula for the distribution of T_m.

8. Phase-type distribution approximation of T_m

In this section we will use the framework of Section 7.2 in order to find the distribution of T_m, the waiting time until the targets of all m genes have been fixed. Assume that the states of the array component process $\boldsymbol{C}_t$ in (46) are linearly ordered $\left\{0, \ldots, \overline{C} - 1\right\}$, corresponding to the decomposition (24) of all regulatory arrays. It follows from (27) and (53) that

$$T_m = \min\{t \geqslant 0; \boldsymbol{C}_t \leqslant \underline{C} - 1\}, \tag{54}$$

since the type of the array component satisfies $g(\boldsymbol{C}_t) = m$ if and only if $0 \leqslant \boldsymbol{C}_t \leqslant \underline{C} - 1$. This gives rise to a decomposition of $\mathcal{C}$ into two groups $\mathcal{C}_{\mathrm{a}} = \{0, \ldots, \underline{C} - 1\}$ and $\mathcal{C}_{\mathrm{n}} = \left\{\underline{C}, \ldots, \overline{C} - 1\right\}$ of states, with T_m the time until $\boldsymbol{C}_t$ reaches the first group $\mathcal{C}_{\mathrm{a}}$. It is therefore convenient to decompose the intensity matrix of $\boldsymbol{C}_t$ into four blocks

$$\boldsymbol{\Lambda} = \begin{pmatrix} **\boldsymbol{\Lambda}_{\mathrm{a}} & \boldsymbol{\Lambda}_{\mathrm{an}} \\ \boldsymbol{\Lambda}_{\mathrm{na}} & \boldsymbol{\Lambda}_{\mathrm{n}} \end{pmatrix}, \tag{55}$$

where $\boldsymbol{\Lambda}_{\mathrm{a}}$ $(\boldsymbol{\Lambda}_{\mathrm{n}})$ contains the transition rates among the states for which all targets have (have not) been reached, and similarly $\boldsymbol{\Lambda}_{\mathrm{an}}$ and $\boldsymbol{\Lambda}_{\mathrm{na}}$ contain the transition rates between these two groups of states. We will regard the first group of states $\mathcal{C}_{\mathrm{a}}$ as absorbing, and the second group $\mathcal{C}_{\mathrm{n}}$ of states as non-absorbing.[4] Let $\boldsymbol{\kappa}_{\mathrm{n}} = \left(\kappa_{\underline{C}}, \ldots, \kappa_{\overline{C}-1}\right)$ be a row vector, whose coordinates represent the initial probabilities (51) of the $\boldsymbol{C}_t$-process at time $t = 0$, for all of the non-absorbing states, and let $\kappa_{\mathrm{n}} = \sum_{\boldsymbol{c}=\underline{C}}^{\overline{C}-1} \kappa_{\boldsymbol{c}}$ be the probability that $\boldsymbol{C}_0$ belongs to some non-absorbing state. Then, since T_m is the time until one of the absorbing states of the continuous time Markov process $\boldsymbol{C}_t$ is reached, it has a phase-type distribution when $\boldsymbol{C}_0$ belongs to a non-absorbing state (Neuts, 1981). Incorporating the possibility that $\boldsymbol{C}_0$ belongs to an absorbing state as well, we find that the waiting time distribution

$$T_m \overset{\mathcal{L}}{\in} (1 - \kappa_{\mathrm{n}})\delta_0 + \kappa_{\mathrm{n}} \mathrm{PD}\left(\frac{\boldsymbol{\kappa}_{\mathrm{n}}}{\kappa_{\mathrm{n}}}, \boldsymbol{\Lambda}_{\mathrm{n}}\right) \tag{56}$$

is a mixture of a one point distribution δ_0 at $t = 0$, and a phase-type distribution, whose first argument is the initial distribution of the $\boldsymbol{C}_t$-process among the non-absorbing states, given that the process starts in such a state, whereas the second argument is the intensity matrix among the non-absorbing states. From (56), and the theory of phase-type distributions, we get very explicit approximate expressions for the distribution function

$$F_{T_m}(t) = 1 - \boldsymbol{\kappa}_{\mathrm{n}} \exp(\boldsymbol{\Lambda}_{\mathrm{n}} t) \boldsymbol{1}_{\overline{C}-\underline{C}}, \quad t \geqslant 0 \tag{57}$$

and density function

$$f_{T_m}(t) = (1 - \kappa_{\mathrm{n}})\delta_0(t) + \boldsymbol{\kappa}_{\mathrm{n}} \exp(\boldsymbol{\Lambda}_{\mathrm{n}} t) \boldsymbol{\Lambda}_{\mathrm{na}} \boldsymbol{1}_{\underline{C}-\underline{C}}, \quad t \geqslant 0 \tag{58}$$

of the waiting time T_m, with $\boldsymbol{1}_n = (1, \ldots, 1)^T$ a column vector of n ones. From this we deduce formulas

$$E(T_m) = -\boldsymbol{\kappa}_{\mathrm{n}} \boldsymbol{\Lambda}_{\mathrm{n}}^{-1} \boldsymbol{1}_{\overline{C}-\underline{C}} \tag{59}$$

and

$$\mathrm{Var}(T_m) = 2\boldsymbol{\kappa}_{\mathrm{n}} \boldsymbol{\Lambda}_{\mathrm{n}}^{-2} \boldsymbol{1}_{\overline{C}-\underline{C}} - \left(\boldsymbol{\kappa}_{\mathrm{n}} \boldsymbol{\Lambda}_{\mathrm{n}}^{-1} \boldsymbol{1}_{\overline{C}-\underline{C}}\right)^2 \tag{60}$$

for the expected value and variance of T_m. Formulas (57)–(60) are the main results of this paper. However, in order to make use of these expressions for the waiting time distribution, we need to know the initial probability vector $\boldsymbol{\kappa}_{\mathrm{n}}$ and the intensity matrix $\boldsymbol{\Lambda}_{\mathrm{n}}$

[4] Strictly speaking, $\mathcal{C}_a$ corresponds to the set of absorbing states of $\boldsymbol{C}_t$ when we change the intensity matrix in (55) by putting $\boldsymbol{\Lambda}_{\mathrm{a}} = \boldsymbol{\Lambda}_{\mathrm{an}} = \boldsymbol{0}$. In any case, formula (57), for the distribution of T_m, does not involve any of $\boldsymbol{\Lambda}_{\mathrm{a}}$ and $\boldsymbol{\Lambda}_{\mathrm{an}}$.

among the non-absorbing states of the array component process. In Section 9 and Appendix A we will describe how (51) and (52) can be used in order to express $\boldsymbol{\kappa}_{\rm n}$ and $\boldsymbol{\Lambda}_{\rm n}$ as functions of the input parameters of Table 1. Then this will be further illustrated in Section 10 and Appendix B.

9. Marginal distribution and transition rates of the array component process

In this section we will find more explicit expressions for the marginal distribution $\boldsymbol{\kappa} = (\kappa_{\boldsymbol{c}})$ and transition matrix $\boldsymbol{\Lambda} = (\lambda_{\boldsymbol{cd}})$ of the array component process $\boldsymbol{C}_t$ that was introduced in Section 7.2. Recall that these quantities are needed in order to derive the distribution function (57) of the waiting time T_m. Then, in Appendix A we apply the results of this section and illustrate how $\boldsymbol{\kappa}$ and $\boldsymbol{\Lambda}$ are computed for the Jukes-Cantor model (2).

9.1. Marginal distribution of array component process

In order to find the marginal distribution $\boldsymbol{\kappa} = (\kappa_{\boldsymbol{c}})$ of the array component process $\boldsymbol{C}_t$ we will make use of the hit variables (37). Since the consensus array process $\underline{\boldsymbol{X}}_t$ in (18) is assumed to be stationary, we may without loss put $t = 0$. Let $H_{jc} = H_{\mathcal{B}_{jc}}(\boldsymbol{X}_{0j})$ refer to the number of substrings of length W of the regulatory sequence $\boldsymbol{X}_{0j}$ of gene j at time 0 that belong to the set $\mathcal{B}_{jc}$. Recall from (38) that $\mathcal{B}_{jc}$ is the set of words of length W with a distance Δ_{jc} in (33) to the closest targeted binding site of gene j.

We will first consider the distance-based components of regulatory sequences from Section 6.2. By the definition of $\mathcal{X}_{cj}$ in (34), and of the hit variables H_{cj}, it follows that

$$\pi(\mathcal{X}_{jc}) = P(\boldsymbol{X}_{0j} \in \mathcal{X}_{cj}) = P(H_{j0} = \ldots = H_{j,c-1} = 0, H_{jc} > 0). \tag{61}$$

In order to evaluate (61), we will need to approximate the joint distribution of $H_{j0}, \ldots, H_{jc}$. To this end, it is convenient to introduce the indicator variable $1_{jcl} = 1(\boldsymbol{X}_{0j;l:l+W-1} \in \mathcal{B}_{jc})$ for the event that a substring of $\boldsymbol{X}_{0j}$ of length W, with leftmost locus at l, belongs to $\mathcal{B}_{jc}$. Since the L components of $\boldsymbol{X}_{0j}$ are independent and identically distributed, it follows that

$$\Pi_{jc} = E(1_{jcl}) = \sum_{\boldsymbol{b} \in \mathcal{B}_{jc}} \pi_{\boldsymbol{b}} \tag{62}$$

does not depend on l, with $\pi_{\boldsymbol{b}} = \prod_{w=1}^{W} \pi_{b_w}$ for any $\boldsymbol{b} = (b_1, \ldots, b_W)$. Now $H_{jc} = \sum_{l=1}^{L_0} 1_{jcl}$, since there are $L_0 = L - W + 1$ possible subwords of $\boldsymbol{X}_{0j}$ of length W, and therefore

$$E(H_{jc}) = L_0 \Pi_{jc}. \tag{63}$$

In order to find the variance of H_{jc}, consider two subsequences of $\boldsymbol{X}_{0j}$ whose leftmost loci l and $l + \eta$ are $\eta \in \{0, \ldots, W-1\}$ positions apart. Let

$$\Pi_{jc\eta} = E(1_{jcl} 1_{jc,l+\eta}) = \sum_{\substack{\boldsymbol{b},\boldsymbol{b}' \in \mathcal{B}_{jc} \\ \boldsymbol{b}_{1+\eta:W} = \boldsymbol{b}'_{1:W-\eta}}} \prod_{w=1}^{W} \pi_{b_w} \cdot \prod_{w=W-\eta+1}^{W} \pi_{b'_w} \tag{64}$$

be the probability that the corresponding substrings of $\boldsymbol{X}_{0j}$ both belong to $\mathcal{B}_{jc}$. In particular, we have that $\Pi_{jc0} = \Pi_{jc}$. Because of the stationarity of the sequence $\boldsymbol{X}_{0j}$ along the L loci, we deduce that

$$\begin{aligned} \mathrm{Var}(H_{jc}) &= L_0 \mathrm{Var}(1_{jcl}) + 2\sum_{\eta=1}^{W-1} (L_0 - \eta) \mathrm{Cov}(1_{jcl}, 1_{jc,l+\eta}) \\ &= L_0 \left(\Pi_{jc} - \Pi_{jc}^2\right) + 2\sum_{\eta=1}^{W-1} (L_0 - \eta)\left(\Pi_{jc\eta} - \Pi_{jc}^2\right). \end{aligned} \tag{65}$$

When the substrings of $\boldsymbol{X}_{0j}$ of length W that match $\mathcal{B}_{jc}$ occur in isolation along regulatory sequence j, there is typically no overdispersion $(E(H_{jc}) \approx \mathrm{Var}(H_{jc}))$ and then we approximate the law of H_{jc} by a Poisson distribution

$$P(H_{jc} = n) = \frac{E(H_{jc})^n}{n!} e^{-E(H_{jc})} \tag{66}$$

for $n = 0, 1, 2, \ldots$, with expected value (63). On the other hand, when the substrings of $\boldsymbol{X}_{0j}$ of length W that belong to $\mathcal{B}_{jc}$ occur in clusters along regulatory sequence j, there is typically overdispersion $(E(H_{jc}) < \mathrm{Var}(H_{jc}))$. Durrett and Schmidt (2007) use a Poisson clumping heuristic in order to handle overdispersion, whereas we will use a simpler approach and approximate the distribution of H_{jc} by a negative binomial (or equivalently, a mixed Poisson) distribution with expected value (63) and variance (65). From properties of the negative binomial distribution we find that

$$P(H_{jc} = n) = \frac{E(H_{jc})^n}{n!} \cdot \frac{\Gamma(n+a)}{\Gamma(a) a^n} \left(1 + \frac{E(H_{jc})}{a}\right)^{-(a+n)}, \tag{67}$$

for $n = 0, 1, 2, \ldots$, with $a = E(H_{jc})^2 / [\mathrm{Var}(H_{jc})) - E(H_{jc}))] > 0$ whenever overdispersion occurs, and with $\Gamma(\cdot)$ the gamma function, whereas equation (66) is used when there is no overdispersion $(\mathrm{Var}(H_{jc}) \leqslant E(H_{jc}))$. Notice in particular that the Poisson distribution in (66) corresponds to the limit $a \to \infty$ in (67). Regardless of whether (66) or (67) is used, we will assume that the hit variables H_{jc} of region j are independent for $c = 0, \ldots, C_j - 1$. In conjunction with (51) and (61), it follows that the stationary distribution of the array component process $\boldsymbol{C}_t$ is

$$\kappa_{\boldsymbol{c}} = \kappa_{(c_1, \ldots, c_m)} \propto \prod_{j=1}^{m} \left\{ P\left(H_{jc_j} > 0\right) \prod_{d=0}^{c_j - 1} P(H_{jd} = 0) \right\}, \tag{68}$$

when these components are defined as in Section 6.2. The proportionality constant of (68) is added in order to ensure that $\sum_{\boldsymbol{c}} \kappa_{\boldsymbol{c}} = 1$, and the probabilities on the right hand side of (68) are obtained from the Poisson or negative binomial distribution approximations, defined in (66) and (67) respectively.

Since $\sum_{c=0}^{C_j - 1} H_{jc} = L_0$, it follows that the independence assumption of $\{H_{jc}\}_{c=0}^{C_j - 1}$ is wrong. However, we can still justify it by first regarding the C_j hit variables $\{H_{jc}\}_{c=0}^{C_j - 1}$ as independent, and then conditioning on their sum being L_0. Such a conditioning is reflected by the fact that we normalize all $\kappa_{\boldsymbol{c}}$ to sum to 1 in (68).

By a similar argument as above, it is possible obtain the stationary distribution of the array component process $\boldsymbol{C}_t$ based on the finer partition of Section 6.3. One finds that

$$\kappa_{\boldsymbol{c}} = \kappa_{(c_1, \ldots, c_m), (n_1, \ldots, n_m)} \propto \prod_{j=1}^{m} \left\{ P\left(H_{jc_j} \in \mathcal{H}_{jc_j n_j}\right) \prod_{d=0}^{c_j - 1} P(H_{jd} = 0) \right\}. \tag{69}$$

9.2. Transition rates of array component process

In this subsection we will obtain explicit expressions for the transition intensity $\lambda_{\boldsymbol{cd}}$ in (52), of the array component process $\boldsymbol{C}_t$, for any two array component indeces $\boldsymbol{c}$ and $\boldsymbol{d}$. We will mostly consider the array decomposition (31) of Section 6.2, with $\boldsymbol{c} = (c_1, \ldots, c_m)$ and $\boldsymbol{d} = (d_1, \ldots, d_m)$. The finer decomposition (41) will also be treated, with $\boldsymbol{c} = (\boldsymbol{c}', \boldsymbol{n}) = (c_1, \ldots, c_m, n_1, \ldots, n_m)$ and $\boldsymbol{d} = (\boldsymbol{d}', \boldsymbol{q}) = (d_1, \ldots, d_m, q_1, \ldots, q_m)$, where $\boldsymbol{c}'$ and $\boldsymbol{d}'$ index distances to targeted binding sites, whereas $\boldsymbol{n}$ and $\boldsymbol{q}$ index information about the size of hit variables. Let $|\boldsymbol{d} - \boldsymbol{c}| = \sum_{j=1}^{m} |d_j - c_j|$ denote to the size of the jump from $\boldsymbol{c}$ to $\boldsymbol{d}$ that corresponds to the distance (in terms of number of mismatches) between regulatory arrays in $\underline{\mathcal{X}}_{\boldsymbol{c}}$ and $\underline{\mathcal{X}}_{\boldsymbol{d}}$. Recall from (21) and (22) that any jump of the consensus array process $\underline{\boldsymbol{X}}_t$ from $\underline{\boldsymbol{x}}$ to $\underline{\boldsymbol{y}}$ will alter it at one or two positions (j, l). From this it follows that the jumps of $\boldsymbol{C}_t$ have a size of at most 2, i.e.

$$\lambda_{\boldsymbol{cd}} = 0 \text{ if } |\boldsymbol{d} - \boldsymbol{c}| > 2. \tag{70}$$

In order to deal with the nonzero jump rates of $\boldsymbol{C}_t$, when $|\boldsymbol{d} - \boldsymbol{c}|$ equals 1 or 2, we divide the transition rate

$$\lambda_{\boldsymbol{cd}} = \lambda^{(1)}_{\boldsymbol{cd}} + \lambda^{(2)}_{\boldsymbol{cd}} \tag{71}$$

in (52) from $\boldsymbol{c}$ to $\boldsymbol{d}$ into two terms, where $\lambda^{(1)}_{\boldsymbol{cd}}$ is the rate of single mutational events, whereas $\lambda^{(2)}_{\boldsymbol{cd}}$ is the rate of double mutations. In the next two subsections, we study these two transition rates separately.

9.2.1. Single mutations

In this subsection we will find explicit expressions for the transition rates $\lambda^{(1)}_{\boldsymbol{cd}}$ of single mutations, that cause $\underline{\boldsymbol{X}}_t$ to change from $\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}$ to $\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}}$. Since we have no tunneling event, only one regulatory sequence $\boldsymbol{X}_{tj}$ of $\underline{\boldsymbol{X}}_t$ will change, and consequently, only one distance-related component C_{tj} of the array component process $\boldsymbol{C}_t$ will be altered. From this it follows that $\lambda^{(1)}_{\boldsymbol{cd}} = 0$ when $|\boldsymbol{d} - \boldsymbol{c}| > 1$, and therefore $|\boldsymbol{d} - \boldsymbol{c}| = 1$ is assumed, with $d_j = c_j \pm 1$. In order to find an explicit expression for $\lambda^{(1)}_{\boldsymbol{cd}}$ when $|\boldsymbol{d} - \boldsymbol{c}| = 1$ we recall the definition of Π_{jc} in (62) and introduce

$$\theta^{(1)}_{jcd} = \frac{1}{\Pi_{jc}} \sum_{\boldsymbol{b} \in \mathcal{B}_{jc}} \pi_{\boldsymbol{b}} \sum_{\boldsymbol{b}' \in \mathcal{B}_{jd}; |\boldsymbol{b}' - \boldsymbol{b}| = 1} p_{b_w b'_w}, \tag{72}$$

the one-step transition rate between the two disjoint sets $\mathcal{B}_{jc}$ and $\mathcal{B}_{jd}$ of binding sites, caused by one single changed letter, with $w = w(\boldsymbol{b}, \boldsymbol{b}')$ the position where $\boldsymbol{b} = (b_1, \ldots, b_W)$ and $\boldsymbol{b}' = (b'_1, \ldots, b'_W)$ differ. Since there are W possible letters of $\boldsymbol{b}$ to change by one mutation, we may view $W^{-1}\theta^{(1)}_{jcd}$ as a transition probability and $\boldsymbol{\theta}^{(1)}_j = \left\{ W^{-1}\theta^{(1)}_{jcd} \right\}_{c,d=0}^{C_j - 1}$ as the transition matrix of a discrete time Markov chain for words of length W, when these words are grouped into C_j states $\mathcal{B}_{j0}, \ldots, \mathcal{B}_{j,C_j-1}$ and the W letters change independently (but one at a time) with a transition matrix $\boldsymbol{P} = (p_{xy})$.

With these preliminaries we are ready to give expressions for the single mutation transition rates of $\boldsymbol{C}_t$. Starting with the distance-based array components (31), we motivate in Appendix C.2 that

$$\begin{aligned} \lambda^{(1)}_{\boldsymbol{cd}} &= N\mu\beta\left(\tfrac{s_{g(\boldsymbol{d})}}{s_{g(\boldsymbol{c})}}\right)\gamma^{1(g(\boldsymbol{d})<g(\boldsymbol{c}))} \cdot \theta^{(1)}_{jc_jd_j} \\ &\cdot \quad P\Big(H_{jc_j} = 1\Big)^{1\left(d_j = c_j + 1\right)} E\Big(H_{jc_j}\Big)^{1\left(d_j = c_j - 1\right)} / P\Big(H_{jc_j} > 0\Big). \end{aligned} \tag{73}$$

where $N\mu$ is the total rate at which mutations appear at each nucleotide within some individual in the population, $\beta\left(s_{g(\boldsymbol{d})}/s_{g(\boldsymbol{c})}\right)$ involves the selection coefficients in terms of the probability[5] that a mutation gets fixed if it changes a regulatory sequence $\boldsymbol{x}_j \in \mathcal{X}_{jc_j}$ into $\boldsymbol{y}_j \in \mathcal{X}_{jd_j}$, whereas $\gamma^{1(g(\boldsymbol{d})<g(\boldsymbol{c}))}$ reduces the transition rate of back mutations by a factor $0 \leqslant \gamma \leqslant 1$. The term $\theta^{(1)}_{jc_jd_j}$ is a transition rate for words of length W, whereas the second row of (73) takes hit variables into account, and thereby it adjusts the transition rate $\lambda^{(1)}_{\boldsymbol{cd}}$ for the fact that $L \geqslant W$ (in particular, the second row equals 1 when $L = W$).

Next we consider the finer partition (41) of array components. Recall the definitions $\boldsymbol{c} = (\boldsymbol{c}', \boldsymbol{n}) = (c_1, \ldots, c_m, n_1, \ldots, n_m)$, $\boldsymbol{d} = (\boldsymbol{d}', \boldsymbol{q}) = (d_1, \ldots, d_m, q_1, \ldots, q_m)$, and $d_j \in \{c_j - 1, c_j, c_j + 1\}$ for a single mutation at some gene j. By an argument analogous to (73), it can be shown that

$$\lambda^{(1)}_{\boldsymbol{cd}} = N\mu\beta\left(\frac{S_{g(\boldsymbol{d}')}}{S_{g(\boldsymbol{c}')}}\right)\gamma^{1\left(g\left(\boldsymbol{d}'\right)<g(\boldsymbol{c}')\right)} \cdot \rho_{\boldsymbol{cd}}, \tag{74}$$

where

$$\rho_{\boldsymbol{cd}} = \begin{cases} \theta^{(1)}_{jc_j,c_j-1} E\Big(H_{jc_j} | H_{jc_j} \in \mathcal{H}_{jc_jn_j}\Big), & d_j = c_j - 1, q_j = 1, \\ \theta^{(1)}_{jc_j,c_j+1} P\Big(H_{j,c_j+1} + 1 \in \mathcal{H}_{j,c_j+1,q_j}\Big) & \\ \qquad \cdot P\Big(H_{jc_j} = 1 | H_{jc_j} \in \mathcal{H}_{jc_j1}\Big), & d_j = c_j + 1, n_j = 1, \\ \theta^{(1)}_{j,c_j+1,c_j} P\Big(H_{jc_j} = \overline{\mathcal{H}}_{jc_jn_j} | H_{jc_j} \in \mathcal{H}_{jc_jn_j}\Big) & \\ \qquad \cdot E\Big(H_{j,c_j+1} | H_{j,c_j+1} > 0\Big), & d_j = c_j, q_j = n_j + 1, \\ \theta^{(1)}_{j,c_j,c_j+1} E\Big(H_{jc_j} 1\Big(H_{jc_j} = \underline{\mathcal{H}}_{jc_jn_j}\Big) | H_{jc_j} \in \mathcal{H}_{jc_jn_j}\Big), & d_j = c_j, q_j = n_j - 1, \\ 0, & \text{otherwise}, \end{cases} \tag{75}$$

and $\mathcal{H}_{cjn} = \{\underline{\mathcal{H}}_{cjn}, \ldots, \overline{\mathcal{H}}_{cjn}\}$. Among the four nonzero transition rates in (75), the first two modify the distance class of regulatory sequence j. In contrast, the last two nonzero transitions of (75) leave the mismatch-based distance class of regulatory sequence j fixed, but instead there is a change in the number of substrings of length W that match this distance class.

9.2.2. Double mutations

In this subsection we will derive explicit expressions for the transition rates $\lambda^{(2)}_{\boldsymbol{cd}}$ of double mutations when $|\boldsymbol{d} - \boldsymbol{c}| = 2$. We will restrict ourselves to the distance-based regulatory array decomposition (31) and consider two choices of $\boldsymbol{c} = (c_1, \ldots, c_m)$ and $\boldsymbol{d} = (d_1, \ldots, d_m)$ separately.

In the first case, $\boldsymbol{c}$ and $\boldsymbol{d}$ differ at one single component j with $d_j = c_j \pm 2$, so that both mutations that change $\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}$ into $\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}}$ at positions (j_1, l_1) and (j_2, l_2) occur within the same enhancer/promoter region j $(j_1 = j_2 = j)$, and also within the same binding site $(|l_2 - l_1| < W)$ of that region. In order to proceed, we will need some definitions. Let $\boldsymbol{d}' = (\boldsymbol{c} + \boldsymbol{d})/2$ be the index of the array component through which the $\boldsymbol{C}_t$ process tunnels on its way from $\boldsymbol{c}$ to $\boldsymbol{d}$. This corresponds to a transition of $\boldsymbol{X}_{tj}$ from $\boldsymbol{x} \in \mathcal{X}_{jc_j}$ to $\boldsymbol{y} \in \mathcal{X}_{jd_j}$, since both of the intermediate regulatory arrays $\underline{\boldsymbol{v}}$ and $\underline{\boldsymbol{w}}$ (defined below (21)) will have their j:th regulatory sequence in $\mathcal{X}_{jd'_j}$. Notice in particular that $|\boldsymbol{d}' - \boldsymbol{c}| = 1$ and $d'_j = c_j \pm 1$. Then, similarly as in (72), we introduce

$$\theta^{(2)}_{jcd} = \frac{1}{\Pi_{jc}} \sum_{\boldsymbol{b} \in \mathcal{B}_{jc}} \pi_{\boldsymbol{b}} \sum_{\boldsymbol{b}' \in \mathcal{B}_{jd}; |\boldsymbol{b}' - \boldsymbol{b}| = 2} p_{b_v b'_v} p_{b_w b'_w}, \tag{76}$$

the two-step transition rate between the two disjoint sets $\mathcal{B}_{jc}$ and $\mathcal{B}_{jd}$ of binding sites, with $v = v(\boldsymbol{b}, \boldsymbol{b}')$ and $w = w(\boldsymbol{b}, \boldsymbol{b}')$ the two positions where $\boldsymbol{b} = (b_1, \ldots, b_W)$ and $\boldsymbol{b}' = (b'_1, \ldots, b'_W)$ differ. Since there are $W(W-1)/2$ pairs of letters in $\boldsymbol{b}$ to change through a double mutation, we may view $2\theta^{(2)}_{jcd}/[W(W-1)]$ as a transition probability and $\boldsymbol{\theta}^{(2)}_j = \left\{ 2\theta^{(2)}_{jcd}/[W(W-1)] \right\}_{c,d=0}^{C_j-1}$ as the transition matrix of a discrete time Markov chain for words of length W where two randomly chosen letters change at a time, and the words are clustered into C_j groups $\mathcal{B}_{j0}, \ldots, \mathcal{B}_{j,C_j-1}$.

With these definitions we are ready to formulate the double mutation transition rates of $\boldsymbol{C}_t$. In Appendix C.3 we motivate that

[5] This is the part of the transition rate in (73) that includes the selection coefficient $s_{g(\boldsymbol{c})}$ of regulatory arrays $\boldsymbol{c}$ for which $g(\boldsymbol{c})$ targets have been reached. As mentioned at the end of Section 6.2, this can be generalized by allowing each regulatory array $\boldsymbol{c}$ to have its own selection coefficient $s_{\boldsymbol{c}}$. The fixation probability of (73) then changes to $\beta(s_{\boldsymbol{d}}/s_{\boldsymbol{c}})$.

$$\begin{aligned}\lambda_{\boldsymbol{cd}}^{(2)} &= 2N\mu(mL-1)^{-1}\beta\left(\tfrac{s_{g(\boldsymbol{d})}}{s_{g(\boldsymbol{c})}}\right)\gamma^{1(g(\boldsymbol{d})<g(\boldsymbol{c}))}r_{\boldsymbol{cd}'}\cdot\theta_{jc_jd_j}^{(2)}\\ &\cdot\quad\left[P\left(H_{jc_j}=1\right)P\left(H_{j,c_j+1}=0\right)\right]^{1\left(d_j=c_j+2\right)}E\left(H_{jc_j}\right)^{1\left(d_j=c_j-2\right)}/P\left(H_{jc_j}>0\right),\end{aligned} \tag{77}$$

when $|\boldsymbol{d}-\boldsymbol{c}|=2$ and $d_j=c_j\pm 2$, where $r_{\boldsymbol{cd}'}$ is the constant value of the tunneling probability $r_{\underline{\boldsymbol{xv}}}=r_{\underline{\boldsymbol{xw}}}$, defined above (22), for any two regulatory arrays with $\underline{\boldsymbol{x}}\in\underline{\mathcal{X}}_{\boldsymbol{c}}$ and $\underline{\boldsymbol{v}},\underline{\boldsymbol{w}}\in\underline{\mathcal{X}}_{\boldsymbol{d}'}$. Notice in particular the similarity between the transition rates (73) and (77) due to single and double mutations. Two major differences is that (77) adds a tunneling probability $r_{\boldsymbol{cd}'}$, and replaces $\theta_{jc_jd_j}^{(1)}$ by $\theta_{jc_jd_j}^{(2)}$, the transition intensity for words of length W based on two (rather than one) changed letters. The second row of (77) takes hit variables into account, and thereby adjusts for the fact that $L\geqslant W$ (in particular, the second row equals 1 when $L=W$).

Next we consider the case when the two mutations occur at different regulatory regions j_1 and j_2, so that $\boldsymbol{c}$ and $\boldsymbol{d}$ will differ at two positions, i.e. $|\boldsymbol{d}-\boldsymbol{c}|=2, d_{j_1}=c_{j_1}\pm 1$, and $d_{j_2}=c_{j_2}\pm 1$. We motivate in Appendix C.3 that

$$\begin{aligned}\lambda_{\boldsymbol{cd}}^{(2)} &= N\mu(mL-1)^{-1}\beta\left(\tfrac{s_{g(\boldsymbol{d})}}{s_{g(\boldsymbol{c})}}\right)\gamma^{1(g(\boldsymbol{d})<g(\boldsymbol{c}))}\left(r_{\boldsymbol{cd}_1}+r_{\boldsymbol{cd}_2}\right)\cdot\theta_{j_1c_{j_1}d_{j_1}}^{(1)}\,\theta_{j_2c_{j_2}d_{j_2}}^{(1)}\\ &\cdot\quad P\left(H_{j_1c_{j_1}}=1\right)^{1\left(d_{j_1}=c_{j_1}+1\right)}E\left(H_{j_1c_{j_1}}\right)^{1\left(d_{j_1}=c_{j_1}-1\right)}/P\left(H_{j_1c_{j_1}}>0\right)\\ &\cdot\quad P\left(H_{j_2c_{j_2}}=1\right)^{1\left(d_{j_2}=c_{j_2}+1\right)}E\left(H_{j_2c_{j_2}}\right)^{1\left(d_{j_2}=c_{j_2}-1\right)}/P\left(H_{j_2c_{j_2}}>0\right),\end{aligned} \tag{78}$$

where $r_{\boldsymbol{cd}_1}$ and $r_{\boldsymbol{cd}_2}$ correspond to tunneling probabilities for the two states that $\boldsymbol{C}_t$ tunnels through on its way from $\boldsymbol{c}$ to $\boldsymbol{d}$.

When comparing the two double mutation rates (77) and (78) with the single mutation rate (73), we notice that they both contain an extra tunneling probability term. It turns out that stochastic tunneling has a negligible impact for neutral models ($s_j\equiv 1$), but it is often important if the intermediate state(s) between $\boldsymbol{c}$ and $\boldsymbol{d}$ has (have) a selective disadvantage compared to $\boldsymbol{c}$ and $\boldsymbol{d}$. The reason is that the fixation probability $\beta(\cdot)$ will then be much larger for the double mutation rates than for the single mutation rate.

We end this subsection by noting that it is also possible to have double mutations when $|\boldsymbol{d}-\boldsymbol{c}|=1$ and $d_j=c_j\pm 1$. This corresponds to a transition from $\underline{\boldsymbol{x}}\in\underline{\mathcal{X}}_{\boldsymbol{c}}$ to $\underline{\boldsymbol{y}}\in\underline{\mathcal{X}}_{\boldsymbol{d}}$ that involves two mutations ($|\underline{\boldsymbol{y}}-\underline{\boldsymbol{x}}|=2$), one of which is silent in the sense that it does not affect the array component process $\boldsymbol{C}_t$. The silent mutation either appears before or after the non-silent mutation. In any case, we will not include these types of double mutations in the transition matrix $\boldsymbol{\Lambda}=(\lambda_{\boldsymbol{cd}})$, and therefore put $\lambda_{\boldsymbol{cd}}^{(2)}=0$ when $|\boldsymbol{d}-\boldsymbol{c}|=1$. This is due to the fact $\boldsymbol{C}_t$ is assumed to be a Markov process (cf. Section 7), where $\boldsymbol{C}_t=\boldsymbol{c}$ means that the consensus array process $\underline{\boldsymbol{X}}_t$, with distribution (18), is imperfectly observed ($\underline{\boldsymbol{X}}_t\in\underline{\mathcal{X}}_{\boldsymbol{c}}$). Whenever a silent mutation occurs, it will neither change the distribution of $\underline{\boldsymbol{X}}_t$ nor the fact that $\underline{\boldsymbol{X}}_t\in\underline{\mathcal{X}}_{\boldsymbol{c}}$. Therefore, it has at most a minor impact on the transition rate in (52), which is a prediction of the actual transition rate from $\underline{\boldsymbol{x}}$ into $\underline{\mathcal{X}}_{\boldsymbol{d}}$, given the information that $\underline{\boldsymbol{x}}\in\underline{\mathcal{X}}_{\boldsymbol{c}}$.

10. Examples

In this section we compute the expected value, standard deviation, and distribution of the waiting time T_m until all m targets appear, for some example models with uniform transition probabilities (2) between nucleotides, as described in Appendix A. Since there are many parameters of the model, we use the default parameter setting of Table 3, and then vary one or a few parameters at a time. More specifically, we start by deriving the waiting time distribution for a simple neutral model with $m=1$ gene in Section 10.1. Then in Section 10.2 we investigate how the number of genes m, the order of target appearance, back mutations, and stochastic tunneling affect the waiting time distribution, in particular for neutral models. Most numerical illustrations can be found in Appendix B, where we study in detail how the waiting time distribution is impacted by the word length W, the coarseness of the array component decomposition, the number K of binding site targets per gene, the values of the selection coefficients $s_1,\dots,s_m$ (non-neutral models), and finally overdispersion of hit variables, for binding site targets with many self repeats.

Table 3
Default parameter settings for the numerical illustrations of Section 10. The fixed state condition (17) is satisfied, since $NL\mu=10^{4+3-8}=10^{-1}\ll 1$.

Description	Parameter(s)	Default value(s)
Population size	N	10 000
Length of regulatory sequence	L	1000
Number of genes	m	1
Mutation rate	μ	10^{-8}
Back mutation probability	γ	1
Length of binding site	W	6
Selection coefficients	$(s_1,\dots,s_m)$	$(1,\dots,1)$
Nr. of binding sites per gene	K	1
Nr. of reg. sequence components	C	2
Intervals with distances to binding site	$\Delta_0,\dots,\Delta_{C-1}$	$\{0,\dots,\delta_{\max}\},\{\delta_{\max}+1,\dots,W\}$
Maximal mismatch	$\delta_{\max}$	0
Number of hit variable intervals	$\{n_{jc}\}$	{1,. . .,1}
Target appearance	TA	fixed order
Stochastic tunneling (ST)		Yes
Accounting for overdispersion		No

10.1. Default model

Since default model of Table 3 has one single ($m=1$) gene, we will drop the gene index j (=1). We refer to

$$\boldsymbol{C}_t\in\mathcal{C}=\{0,1\} \tag{79}$$

in (46) as a regulatory sequence component process with $\overline{C}=C=2$ states, of which 0 is absorbing and 1 is non-absorbing. This state space $\mathcal{C}=\mathcal{C}_{\text{red}}$ corresponds to the coarse decomposition of regulatory sequences in Section 6.1. But $\mathcal{C}=\mathcal{C}_{\text{mism}}$ also coincides with the distance-based decomposition of regulatory sequences in Section 6.2, since $m=1$ and there are only two sets Δ_0 and Δ_1 of distances to the targeted binding site.

In order to find the statistical properties of $\boldsymbol{C}_t$, we first notice that there are $L_0=L-W+1=995$ possible locations for the only ($K=1$) targeted binding site $\boldsymbol{b}_1$ of length $W=6$. Since no mismatches between a substring of length $W=6$ of the regulatory sequence and $\boldsymbol{b}_1$ are allowed ($\delta_{\max}=0$),

$$\begin{aligned}\mathcal{B}_0 &= \{\boldsymbol{b}_1\}\\ \mathcal{B}_1 &= \{\boldsymbol{b};\boldsymbol{b}\neq\boldsymbol{b}_1\},\end{aligned} \tag{80}$$

represent the sets of substrings that correspond to a reached or missed target respectively. Let H_0 and H_1 denote the number of subsequences of length $W=6$ along the consensus regulatory sequence $\boldsymbol{X}_0$ at time $t=0$ that belong to $\mathcal{B}_0$ and $\mathcal{B}_1$ respectively. From (63) and formula (A.10) of Appendix A we find that

$$\begin{aligned}E(H_0) &= L_04^{-W}=0.2429,\\ E(H_1) &= L_0\left(1-4^{-W}\right)=994.7571,\end{aligned} \tag{81}$$

respectively. Since we assume no overdispersion of H_0 and H_1 for the default model of Table 3, these two random variables are independent and Poisson distributed. It then follows from (66) and (68) that the initial distribution of $\boldsymbol{C}_0$ is (κ_0, κ_1), where

$$\begin{array}{rcl} \kappa_0 & = & 1 - e^{-E(H_0)} = 1 - e^{-L_0 4^{-W}} = 0.2157, \\ \kappa_1 & = & \left(1 - e^{-E(H_1)}\right)e^{-E(H_0)} \approx e^{-L_0 4^{-W}} = 0.7843. \end{array} \qquad (82)$$

Consequently, the probability is $\kappa_0 = 0.2157$ that the binding site $\boldsymbol{b}_1$ is fixed already at time 0. We assume that $\boldsymbol{C}_t$ is a Markov process with intensity matrix (55), which simplifies to

$$\boldsymbol{\Lambda} = \begin{pmatrix} -\lambda_{01} & \lambda_{01} \\ \lambda_{10} & -\lambda_{10} \end{pmatrix}, \qquad (83)$$

where λ_{01} and λ_{10} are the rates at which the regulatory sequence looses or acquires the targeted binding site $\boldsymbol{b}_1$ respectively. In order to find these two rates from (73), we first compute the one-step transition rates (72), back and forth between $\mathcal{B}_0$ and $\mathcal{B}_1$. It follows from (A.8) and (A.11) that

$$\begin{array}{rcl} \theta_{01}^{(1)} & = & \binom{W}{1}3/\left[3 \cdot 4^W \cdot 4^{-W}\right] = W, \\ \theta_{10}^{(1)} & = & \binom{W}{1}3/\left[3 \cdot 4^W\left(1 - 4^{-W}\right)\right] = W/\left(4^W - 1\right). \end{array} \qquad (84)$$

Using the fact that the model of Table 3 is selectively neutral $(s_1 = 1, \beta(s_1) = 1/N)$ with all back mutations allowed $(\gamma = 1)$, we find from (63), (67), (73), (81), and (84) that

$$\begin{array}{rcl} \lambda_{01} & = & \mu\theta_{01}^{(1)}E(H_0)e^{-E(H_0)}/\left(1 - e^{-E(H_0)}\right) \\ & = & \mu L_0 W 4^{-W} e^{-L_0 4^{-W}}/\left(1 - e^{-L_0 4^{-W}}\right) \\ & = & 10^{-8} \cdot 995 \cdot 6 \cdot 4^{-6} \cdot e^{-0.2429}/\left(1 - e^{-0.2429}\right) \\ & = & 5.3007 \cdot 10^{-8}, \\ \lambda_{10} & = & \mu\theta_{10}^{(1)}E(H_1)/\left(1 - e^{-E(H_1)}\right) \\ & = & \mu L_0 W 4^{-W}/\left(1 - e^{-L_0\left(1 - 4^{-W}\right)}\right) \\ & \approx & \mu L_0 W 4^{-W} \\ & = & 10^{-8} \cdot 995 \cdot 6 \cdot 4^{-6} \\ & = & 1.4575 \cdot 10^{-8}. \end{array} \qquad (85)$$

It follows from (56) that the distribution of the waiting time

$$\begin{array}{rcl} T_1 & \overset{\mathcal{L}}{\in} & \kappa_0 \cdot \delta_0 + \kappa_1 \mathrm{Exp}(\lambda_{10}) \\ & = & \left(1 - e^{-L_0 4^{-W}}\right)\delta_0 + e^{-L_0 4^{-W}}\mathrm{Exp}\left(\mu L_0 W 4^{-W}\right) \\ & = & 0.2157 \cdot \delta_0 + 0.7843 \cdot \mathrm{Exp}\left(0.1458 \cdot 10^{-7}\right) \end{array} \qquad (86)$$

is a mixture between a one point distribution at 0 and an exponential distribution, with expected value and standard deviation

$$\begin{array}{rcl} E(T_1) & = & \kappa_1 \cdot \lambda_{10}^{-1} = (\mu L_0 W)^{-1} 4^W e^{-L_0 4^{-W}} = 5.381 \cdot 10^7, \\ D(T_1) & = & \sqrt{1 - \kappa_0^2} \cdot \lambda_{10}^{-1} = 6.700 \cdot 10^7 \end{array} \qquad (87)$$

respectively. Notice in particular that (86) and (87) verify formulas (6) and (7) of Section 2.2, in the special case of a selectively neutral model. In Appendix B.1.1 we will motivate (6)-(7) in a different way, using three regulatory sequence components rather than two, as in (79).

10.2. Varying number of genes, backward mutation probability, selection coefficients, order of target appearance, and absence or presence stochastic tunneling

In this subsection we will investigate how the distribution of the waiting time T_m depends on the number of genes m, the backward mutation probability γ, the target appearance rule TA, and absence/presence of stochastic tunneling (ST). We will use a distance-based state space $\mathcal{C}_{\text{mism}}$ of the array component process $\boldsymbol{C}_t = (C_{t1}, \ldots, C_{tm})$, defined in (32) and (49). This state space has C^m elements, and we will describe how to reduce its size under $C = 2$ and $C = 3$ scenarios. We will consider general selection models $\{s_h\}_{h=0}^m$, but the multiplicative model

$$s_h = s^h, \quad h = 0, 1, \ldots, m \qquad (88)$$

will be treated in particular detail, since it simplifies some formulas for the waiting time distribution. Notice that the neutral model is a special case $(s = 1)$ of (88). All other parameters are given by Table 3.

10.2.1. Arbitrary order of target appearance, $C = 2$

In the next two subsections we will assume that local targets may appear in any order (TA = arbitrary), as defined in (35). Recall from Section 9.2 that all single mutations (73) and many double mutations (77) only affect one single component C_{tj} of $\boldsymbol{C}_t$, whereas other double mutations (78) involve two components of $\boldsymbol{C}_t$. We will ignore these latter type of double mutations. If additionally the multiplicative selection model (88) is used it follows that all $C_{t1}, \ldots, C_{tm}$ evolve as independent Markov processes with C levels each.

Assume that $C = 2$, so that $C_{tj} \in \{0, 1\}$ keeps track of whether the local target of gene j has been reached (0) or not (1). This implies in particular that all double mutations are ignored, since any double mutation (77) that only affects one gene requires at least $C = 3$ levels of C_{tj}. When $C = 2$, it is possible to replace the distance-based state space (32) of $\boldsymbol{C}_t$ of size 2^m, by the maximally reduced state space (28)-(29) of size $m + 1$. Since the Markov property of $\boldsymbol{C}_t$ is not lost by such a state space reduction, when double mutations are ignored, no information is lost. In more detail, we replace $\boldsymbol{C}_t = (C_{t1}, \ldots, C_{tm})$ with $\boldsymbol{C}_t = \sum_{j=1}^m C_{tj} = m - g(\boldsymbol{C}_t)$, so that $\boldsymbol{C}_t \in \mathcal{C}_{\text{red}} = \{0, \ldots, m\}$ monitors how may local targets that have not yet been reached, with 0 absorbing and all other states non-absorbing.

In order to find the waiting time distribution F_{T_m} in (57), we must find the marginal distribution $\boldsymbol{\kappa}_{\text{red}} = (\kappa_{\text{red}})_{c=0}^m$ and the transition matrix $\boldsymbol{\Lambda}_{\text{red}} = \left(\lambda_{\text{red},cd}\right)_{c,d=0}^m$ of the state space reduced process $\boldsymbol{C}_t \in \mathcal{C}_{\text{red}}$. Since the components of C_{tj} evolve independently, the initial distribution $\boldsymbol{C}_0 \sim \mathrm{Bin}(m, \kappa_1)$ of the array component process is binomial, i.e.

$$\kappa_{\text{red},c} = \binom{m}{c}(1 - \kappa_1)^{m-c}\kappa_1^c, \quad c = 0, \ldots, m, \qquad (89)$$

with κ_1 as in (82), the probability that one single gene has not yet reached its local target at time $t = 0$. The entries of the transition matrix $\boldsymbol{\Lambda}_{\text{red}}$ will involve the transition rates λ_{10} and λ_{01} at which one single gene acquires and looses a binding site. It follows from (73) that

$$\begin{array}{rcl} \lambda_{10} & = & \mu\beta(s)L_0 W 4^{-W}, \\ \lambda_{01} & = & \gamma\mu\beta(s^{-1})L_0 W 4^{-W} e^{-L_0 4^{-W}}/\left(1 - e^{-L_0 4^{-W}}\right). \end{array} \qquad (90)$$

Compared to (85), formula (90) is a generalization that also accounts for selection in terms of the s parameter of the multiplicative selection model (88) and back mutations in terms of γ. The elements of $\boldsymbol{\Lambda}_{\text{red}}$ are given by

$$\lambda_{\text{red},cd} = \begin{cases} c\lambda_{10}, & 1 \leqslant c \leqslant m, d = c - 1, \\ -[c\lambda_{10} + (m - c)\lambda_{01}], & 0 \leqslant c \leqslant m, d = c, \\ (m - c)\lambda_{01}, & 0 \leqslant c \leqslant m - 1, d = c + 1, \\ 0, & \text{otherwise}. \end{cases} \qquad (91)$$

Indeed, when $1 \leqslant c \leqslant m$ genes have not yet reached their targets, there is a total forward rate $c\lambda_{10}$ for any of them to a acquire a binding site, and a total backward rate $(m-c)\lambda_{01}$ that any of the other $m-c$ genes will loose their binding sites. The phase-type distribution of T_m can be derived from (56), using either the full state space (49) of size 2^m, or, when double mutations are ignored, the analogous formula for the reduced state space (29). It is helpful to analyze the latter formula, since it enables us to derive an explicit expression for the distribution F_{T_m} of the waiting time T_m until all m targets have been reached. In analogy with (57) and (59), we obtain a distribution function and expected value

$$\begin{aligned} F_{T_m}(t) &= 1 - \boldsymbol{\kappa}_{\text{red,n}} \exp\left(\boldsymbol{\Lambda}_{\text{red,n}} t\right) \mathbf{1}_m, \quad t \geqslant 0, \\ E(T_m) &= -\boldsymbol{\kappa}_{\text{red,n}} \boldsymbol{\Lambda}_{\text{red,n}}^{-1} \mathbf{1}_m \end{aligned} \tag{92}$$

of T_m, with $\boldsymbol{\kappa}_{\text{red,n}} = (\kappa_{\text{red},c})_{c=1}^{m}$ the distribution of $\boldsymbol{C}_0$ among the non-absorbing states, and $\boldsymbol{\Lambda}_{\text{red,n}} = (\lambda_{\text{red},cd})_{c,d=1}^{m}$ the submatrix of $\boldsymbol{\Lambda}_{\text{red}}$ for the non-absorbing states. The formula for the expected waiting time $E(T_m)$ can be simplified, by introducing the ratio

$$r = \frac{\lambda_{01}}{\lambda_{10}} = \frac{\gamma\beta(s^{-1})e^{-L_0 4^{-W}}}{\beta(s)\left(1 - e^{-L_0 4^{-W}}\right)} \tag{93}$$

between the rates at which a single gene looses or acquires a binding site in its enhancer/promoter region. Indeed, it follows from formulas (12.181), (12.182), (12.184), and (12.186) of Hössjer et al. (2018), that $E(T_m)$ is given by (9)-(10) of Section 2.2, where the special case of (93) for a neutral model ($s = 1, \beta(s) = \beta(s^{-1}) = 1/N$) was used. This formula for the expected waiting time is a function of the number of genes m, the ratio r in (93) and the forward rate $\lambda_{10} = \mu\beta(s)L_0 W 4^{-W}$ at which new local targets are acquired. We notice that $E(T_m)$ increases exponentially with m when back mutations are allowed ($\gamma > 0$, or equivalently, $r > 0$), whereas it increases at a logarithmic rate when no back mutations are present ($\gamma = r = 0$).

In Table 4 we calculated $E(T_m)$ as a function of m for $\gamma = 0$ and 1, for a neutral model ($s = 1$ in (88)). The explicit formula (9) only applies when stochastic tunneling (ST) is not accounted for, whereas the larger state space of size 2^m is used when ST is fully accounted for. However, it can be seen from Table 4 that ST typically has a very small impact on the expected waiting time for neutral models.

It turns out that formula (92) is also applicable for any type of selection model $\{s_h\}_{h=0}^{m}$. It is only the intensity matrix $\boldsymbol{\Lambda}_{\text{red}}$ that needs to be generalized from (91) to

$$\lambda_{\text{red},cd} = \begin{cases} c\lambda_{10}(c), & 1 \leqslant c \leqslant m, d = c-1, \\ -[c\lambda_{10}(c) + (m-c)\lambda_{01}(c)], & 0 \leqslant c \leqslant m, d = c, \\ (m-c)\lambda_{01}(c), & 0 \leqslant c \leqslant m-1, d = c+1, \\ 0, & \text{otherwise}. \end{cases} \tag{94}$$

where the rates

$$\begin{aligned} \lambda_{10}(c) &= \mu\beta(s_{m-c+1}/s_{m-c})L_0 W 4^{-W}, \\ \lambda_{01}(c) &= \gamma\mu\beta(s_{m-c-1}/s_{m-c})L_0 W 4^{-W} e^{-L_0 4^{-W}} / \left(1 - e^{-L_0 4^{-W}}\right). \end{aligned} \tag{95}$$

at which a single genes acquires or loses a targeted binding site will depend on how many genes c that have not yet reached their local targets. In Appendix B we will analyze the properties of the waiting time distribution for a target-selected model

$$s_0 = s_1 = \ldots = s_{m-1} = 1, \quad s_m = s. \tag{96}$$

In particular we will investigate how sensitive the expected waiting time is to the selection coefficient $s = s_m$ of the final target.

10.2.2. Arbitrary order of target appearance, $C = 3$

As in the previous subsection, assume an arbitrary order of target appearance (TA = arbitrary), and to start with, a multiplicative selection model (88). We will describe how the waiting time distribution is determined when the components $C_{tj} \in \{0, 1, 2\}$ of the array component process $\boldsymbol{C}_t$ has $C = 3$ levels, of which $C_{tj} = 0$ corresponds to a locally reached target of gene j. Double mutations (78) that affect two genes will be ignored, so that C_{tj} evolve as independent Markov processes for $j = 1, \ldots, m$, with a marginal distribution $(\kappa_0, \kappa_1, \kappa_2)$ obtained from (49) and a transition matrix $\boldsymbol{\Lambda} = (\lambda_{cd})_{c,d=0}^{2}$ obtained from (73) and (77). It is possible then to "half-reduce" the state space of $\boldsymbol{C}_t$ to

$$\mathcal{C}_{\text{halfred}} = \{\boldsymbol{c} = (c_1, c_2);\, 0 \leqslant c_1 + c_2 \leqslant m\}. \tag{97}$$

Whereas the unreduced state space (49) has 3^m states, the half-reduced one in (97) has $(m+1)(m+2)/2$ states. If $\boldsymbol{C}_t = (c_1, c_2)$, there are c_1, c_2, and $m - c_1 - c_2$ genes for which the corresponding regulatory sequence process C_{tj} equals 0, 1, and 2 respectively. In particular, $(c_1, c_2) = (m, 0)$ is the absorbing state, whereas all other states of $\mathcal{C}_{\text{halfred}}$ are non-absorbing. Since the components of C_{tj} evolve independently, it follows that the marginal distribution of $\boldsymbol{C}_0$ is multinomial, i.e.

$$P(\boldsymbol{C}_0 = \boldsymbol{c}) = \kappa_{\text{halfred},\boldsymbol{c}} = \frac{m!}{c_2! c_1! (m - c_1 - c_2)!} \kappa_0^{c_1} \kappa_1^{c_2} \kappa_2^{m-c_1-c_2}. \tag{98}$$

Table 4
Expected waiting times $E(T_m)$, obtained by varying m, γ, TA, and presence/absence of stochastic tunneling (ST). All other parameters are the same as for the default model of Table 3. Formulas (9)-(10) and (92) apply to models without ST, but they also provide an excellent approximation when ST is accounted for.

	$\gamma = 0$				$\gamma = 1$	
	TA fixed		TA arbitrary		TA arbitrary or fixed	
m	No ST	ST	No ST	ST	No ST	ST
1	$5.3813 \cdot 10^7$	$5.3813 \cdot 10^7$	$5.3813 \cdot 10^7$	$5.3813 \cdot 10^7$	$5.3813 \cdot 10^7$	$5.3813 \cdot 10^7$
2	$1.0763 \cdot 10^8$	$1.0756 \cdot 10^8$	$8.6522 \cdot 10^7$	$8.6502 \cdot 10^7$	$2.0548 \cdot 10^8$	$2.0518 \cdot 10^8$
3	$1.6144 \cdot 10^8$	$1.6128 \cdot 10^8$	$1.0916 \cdot 10^8$	$1.0912 \cdot 10^8$	$6.9225 \cdot 10^8$	$6.9065 \cdot 10^8$
4	$2.1523 \cdot 10^8$	$2.1498 \cdot 10^8$	$1.2628 \cdot 10^8$	$1.2622 \cdot 10^8$	$2.3985 \cdot 10^9$	$2.3915 \cdot 10^9$
5	$2.6906 \cdot 10^8$	$2.6867 \cdot 10^8$	$1.3999 \cdot 10^8$	$1.3993 \cdot 10^8$	$8.7385 \cdot 10^9$	$8.7090 \cdot 10^9$
6	$3.2288 \cdot 10^8$	$3.2235 \cdot 10^8$	$1.5143 \cdot 10^8$	$1.5135 \cdot 10^8$	$3.3245 \cdot 10^{10}$	$3.3120 \cdot 10^{10}$

The elements of the transition matrix $\boldsymbol{\Lambda}_{\text{halfred}} = \left(\lambda_{\text{halfred},\boldsymbol{cd}}\right)$ are

$$\lambda_{\text{halfred},\boldsymbol{cd}} = \begin{cases} (m - c_1 - c_2)\lambda_{21}, & d_1 = c_1, d_2 = c_2 + 1, c_1 + c_2 < m, \\ c_2\lambda_{12}, & d_1 = c_1, d_2 = c_2 - 1, c_2 > 0, \\ c_2\lambda_{10}, & d_1 = c_1 + 1, d_2 = c_2 - 1, c_2 > 0, \\ c_1\lambda_{01}, & d_1 = c_1 - 1, d_2 = c_2 + 1, c_1 > 0, \\ -\sum_{\boldsymbol{e} \neq \boldsymbol{c}} \lambda_{\text{halfred},\boldsymbol{ce}}, & d_1 = c_1, d_2 = c_2, \\ 0, & \text{otherwise}, \end{cases} \tag{99}$$

where the first four rows of (99) correspond to scenarios where the regulatory sequence component process C_{tj} of some gene changes according to $2 \to 1, 1 \to 2, 1 \to 0$, and $0 \to 1$ respectively. The distribution and expected value of the waiting time T_m are then derived from

$$\begin{aligned} F_{T_m}(t) &= 1 - \boldsymbol{\kappa}_{\text{halfred,n}} \exp\left(\boldsymbol{\Lambda}_{\text{halfred,n}} t\right) \boldsymbol{1}_n, \quad t \geqslant 0, \\ E(T_m) &= -\boldsymbol{\kappa}_{\text{halfred,n}} \boldsymbol{\Lambda}^{-1}_{\text{halfred,n}} \boldsymbol{1}_n, \end{aligned} \tag{100}$$

where $\boldsymbol{\kappa}_{\text{halfred,n}}$ and $\boldsymbol{\Lambda}_{\text{halfred,n}}$ contain the marginal distribution probabilities and the transition rates among the non-absorbing states of $\mathcal{C}_{\text{halfred}}$, whereas n = (m+1)(m+2)/2-1 is the number of non-absorbing states.

Formula (100) is also valid for a general selection model $\{s_h\}_{h=0}^m$. A transition from $\boldsymbol{c} = (c_1, c_2)$ to $\boldsymbol{d} = (d_1, d_2)$ corresponds to a regulatory array component that changes type from $g(\boldsymbol{c}) = c_1$ to $g(\boldsymbol{d}) = d_1$. The corresponding transition rate $\lambda_{\text{halfred},\boldsymbol{cd}}$ in (100) will therefore be a function of $\beta\left(s_{d_1}/s_{c_1}\right)$ in the single transition rate term on the right hand side. This generalized waiting time formula can be applied, for instance, to the target-selected model (96).

It is shown in Appendix B.1 that not much accuracy is lost by using $C = 2$ intervals per gene instead of $C = 3$ for neutral models. On the other hand, we also demonstrate in Appendix B.1 that for a target-selected model (96), $C = 3$ is often required in order to obtain an accurate approximation of the waiting time distribution when $s = s_m$ is large, whereas the coarser array decomposition with $C = 2$ can be very misleading.

10.2.3. Fixed order of target appearance, $C = 2$

When the gene specific binding sites have to appear in a fixed order (TA = fixed), the waiting time T_m will also depend on the back mutation probability γ, although not as strongly as when TA = arbitrary. In order to highlight this, the upper and lower parts of Fig. 2 in Section 2.2 depict the distribution and density functions of T_m, for $\gamma = 0$ and $\gamma = 1$ respectively, when the state space (32) of $\boldsymbol{C}_t$ has $C = 2$ components per gene. It is more difficult to find closed form expressions of $E(T_m)$ when TA = fixed (compared to TA = arbitrary), even in the absence of stochastic tunneling, due to the fact that the state space of $\boldsymbol{C}_t$ cannot be reduced to (29) when back mutations are possible, i.e. when $0 < \gamma \leqslant 1$. Indeed, when TA = fixed, the targets of all genes with order number j ($2 \leqslant j \leqslant m$) are silent as long the first $j - 1$ targets have not yet appeared. Therefore, when back mutations are possible and a non-silent binding site is lost, the reduced state space will not always keep track of the fact that one or several targets are silenced. In spite of this, it can be shown that the distribution of T_m, for a multiplicative model (88), is the same for TA = fixed and TA = arbitrary when $\gamma = 1$, since the evolution of the m regulatory sequences can be monitored independently, until all of them have reached their targets. In particular, it follows from this that $E(T_m)$ is given by (9) when TA = fixed, $\gamma = 1$ and there is no stochastic tunneling (ST). In contrast, the distribution of T_m will differ for the two TA schemes when $0 < \gamma < 1$, since back mutations are always allowed at silent genes (since they do not decrease $g(\boldsymbol{c})$) when TA = fixed, but only with probability γ when TA = arbitrary.

It turns out that the distribution of T_m has a very explicit form for a multiplicative selection model when TA = fixed, $\gamma = 0$, and ST is not accounted for. Then the array component process $\boldsymbol{C}_t$ is still Markovian after a reduction of the state space to (29). In particular, the initial distribution of $m - \boldsymbol{C}_t$ is truncated geometric, with

$$P(\boldsymbol{C}_0 = c) = \kappa_{\text{red},c} = \begin{cases} \kappa_0^m, & c = 0, \\ \kappa_0^{m-c}(1 - \kappa_0), & c = 1, \ldots, m, \end{cases}$$

with κ_0 as in (82), whereas the transition matrix $\boldsymbol{\Lambda}_{\text{red}} = \left(\lambda_{\text{red},cd}\right)_{c,d=0}^m$ has elements

$$\lambda_{\text{red},cd} = \begin{cases} 0, & c + 1 \leqslant d \leqslant m, \\ \kappa_0^{c-d-1}(1 - \kappa_0)\lambda_{10}, & 1 \leqslant d \leqslant c - 1, \\ \kappa_0^{c-1}\lambda_{10}, & d = 0, \\ -\sum_{e=0}^{c-1} \lambda_{\text{red},ce}, & d = c, \end{cases}$$

with λ_{10} as in (90). From this, and the upper part of (92), it follows that whenever a particular gene j becomes non-silent (that is, when the binding site gets fixed in the previous gene $j - 1$), with probability κ_0 it has its binding site at place already, whereas with probability κ_1 it takes an Exp(λ_{10})-distributed time for this to happen. Since target appearance happens independently for all genes, we deduce that

$$T_m \overset{\mathcal{L}}{\in} \Gamma(J, \lambda_{10})$$

is a mixture of gamma distributions, and the mixing variable $J \sim \text{Bin}(m, \kappa_1)$ is the number of genes that don't have their targets fixed when they are non-silenced. Based on this, it is possible to derive the expected value and standard deviation of the waiting time T_m, as

$$\left.\begin{aligned} E(T_m) &= m\kappa_1 \cdot \lambda_{10}^{-1}, \\ D(T_m) &= \sqrt{m(1 - \kappa_0^2)} \cdot \lambda_{10}^{-1}, \end{aligned}\right. \quad \text{if } \gamma = 0, \tag{101}$$

in agreement with formula (11) for a selectively neutral model ($s = 1$ in (88)). In particular, we notice that $E(T_m)$ grows linearly with m in (101), when no back mutations appear. From Table 4 we also see that the upper part of (101) (which assumes no ST), provides an excellent approximation of $E(T_m)$ also when ST occurs. Again, this is a typical feature of a selectively neutral model.

11. Discussion

In this paper we developed a general analytical framework for the statistical distribution of the waiting time T_m until one or several binding sites appear in the enhancer or promoter regions of m genes, in order for a new trait or phenotype to be expressed. Assuming that the population is small enough to warrant a fixed-state approximation, whereby all individuals have approximately the same regulatory sequences at all m genes, we derived a phase-type distribution approximation of the waiting time. Our model includes a number of parameters, such as the number of genes m, the population size N, the length of the regulatory regions and of the binding sites at all genes, the frequencies of the nucleobases A, C, G, T, the mutation rate per nucleotide, the selective fitness of individuals for which one or several of the binding site targets have appeared, and finally a parameter that controls for the extent to which back mutations are allowed. We also incorporated stochastic tunneling, whereby a newly mutated sequence acquires a second mutation before it spreads to the whole population, so that these two mutations effectively will resemble a double mutation, on a larger timescale. Finally, we allowed the binding site targets at the m genes to arrive in a pre-specified or arbitrary order, with possible mismatches.

For neutral models (that is, when the selective fitness is the same for all individuals, regardless of the number of binding sites that were acquired in their regulatory sequences), we found that the expected waiting time $E(T_m)$ increases logarithmically (linearly) with the number of genes m, when back mutations are not allowed and the binding site targets may appear in an arbitrary (fixed) order. On the other hand, $E(T_m)$ increases exponentially with m when back mutations are allowed, regardless of whether the binding site targets appear in a fixed or arbitrary order. We believe these are important findings. Indeed, it seems that back mutations are difficult to avoid in a model, without incorporating external information (Marks et al., 2017). We also found that stochastic tunneling has a negligible impact on the waiting time T_m for neutral models, and moderate influence for double mutation events where the first mutation is neutral and the second one is highly beneficial. On the other hand, stochastic tunneling may greatly reduce the expected waiting time when the first mutation is deleterious and the second one is not.

A number of generalizations of our work is possible. Below we divide them into two categories in Sections 11.1–11.2, depending on whether they require an extension of the model in Sections 2–4 or not.

11.1. Our imposed waiting time distribution approximations

Suppose that the model of Sections 2.1 and 3–4 provides a reasonable description of the regulatory sequence dynamics at m genes. Our analytical phase-type distribution of the waiting time T_m, until all m targeted binding sites appear, is still an approximation that relies on a number of assumptions. First, it requires the fixed state population model of Section 5, whereby all individuals share the same consensus regulatory sequence at the enhancer/promoter regions of all genes. This assumption is more accurate the smaller the population size is, the shorter the regulatory sequences are, and the smaller the mutation rate is.

Second, the rates in Section 5 at which the consensus sequences change, due to single or double mutations, are also based on asymptotic approximations. In particular, our derivation in Appendix C.1 of double mutation rates due to stochastic tunneling, is a bit heuristic. It is indeed an interesting topic of future research to find more accurate expressions for double mutation rates of multitype Moran processes.

Third, in Sections 6–7 we decomposed the set of possible consensus regulatory sequences of all genes (the regulatory array) into components. We assumed that jumps between these components form a Markov process, whose jumps are either caused by single or double mutations. The numerical results of Appendix B.1 and the theoretical analysis of Appendix C.4 indicate that a decomposition (31) with components having $C = 2$ intervals per gene works well for neutral models, where C is the number of intervals of the number of mismatches between each regulatory sequence and the corresponding closest targeted binding site. We also found that a finer decomposition of all regulatory arrays with at least $C = 3$ intervals per gene might be necessary for some models where the final target has a higher fitness. On the other hand, the even finer decomposition (41), which also includes hit variables, does not seem to increase the accuracy of the waiting time distribution a lot. In general, a coarser decomposition of regulatory arrays into larger components has the effect of shortening the waiting time. Indeed, if regulatory arrays that are very distant from the final target are mixed with those that are closer to the target, the average effect is to shorten the waiting time. For this reason, if the goal is to demonstrate that the waiting time is long, it is conservative to choose a coarser array decomposition.

We did not account for the fact that some mutations may cause the same binding site sequence to appear at several overlapping locations, except in a special case (see Appendices C.2 and C.5). Although the rates of double mutations and stochastic tunneling is typically negligible compared to the single mutation rates for neutral models, they are sometimes important to account for in models where the intermediate states have a selective disadvantage. In addition, if the transition rates of single and double mutations are of the same order, it turns out that the phase-type distribution approximation is less accurate (Durrett et al., 2009; Hössjer et al., 2018). It is also possible that triple mutations are needed for some models where several intermediate states, before the final target is reached, have a low fitness.

In spite of these approximations, we still believe that the phase-type distribution approximation gives insights about the waiting time T_m, and how it depends on various biological parameters.

11.2. Extensions of the model

The model of Sections 2–4 can be modified or extended in several ways. First, in Section 4 we assumed point mutations such that nucleotides change independently in space and time, with probabilities that are the same for all loci. This implies that the nucleotides of DNA strings, for a neutral model, are independent and identically distributed random variables. The most obvious extension is to consider mutation rates that vary between genes, and even within each regulatory sequence (mutational hotspots). A more elaborate model allows for serial dependence between nucleotides (Behrens and Vingron, 2010; Behrens et al., 2012), for instance when mutations operate on codons, i.e. triplets of nucleotides (Goldman and Yang, 1994), when the length of the regulatory array may change through insertions and deletions (Tuğrul et al., 2015), or when multiple mutations arrive together in clusters (Chan and Gordenin, 2015). If the environment is modeled explicitly, it is also possible to consider stress-induced and other types of non-random mutations (Fitzgerald and Rosenberg, 2019).

Second, it is possible to consider a population size $N = N(t)$ that changes over time t. It follows from formula (23) that the transition rates of the Markov processes in Sections 5 and 7 are independent of population size for a selectively neutral model, and in this case our results are valid also when $N(t)$ varies over time. This is not true for non-neutral models, since the fixation probability (5) depends not only on selection coefficients, but also on the population size. This implies that the transition matrix $\mathbf{\Lambda} = \mathbf{\Lambda}(t)$ of the Markov process for the regulatory array components in Section 7 will vary over time for non-neutral models. However, it is possible to generalize the density formula (58) for the waiting time T_m, to

$$f_{T_m}(t) = \boldsymbol{\kappa}_{\mathrm{n}} \exp\left(\int_0^t \mathbf{\Lambda}_{\mathrm{n}}(s) ds\right) \mathbf{\Lambda}_{\mathrm{na}}(t) \mathbf{1}_{\underline{C}},$$

when t > 0, where $\mathbf{\Lambda}_{\mathrm{n}}(t)$ and $\mathbf{\Lambda}_{\mathrm{na}}(t)$ are the submatrices of $\mathbf{\Lambda}(t)$, containing all transitions between the non-absorbing states, and the transitions from the non-absorbing to the absorbing states, respectively.

Third, while we analyzed a Moran model of a homogeneous population, it is also possible to study Moran models for populations with geographic substructure. In evolutionary graph theory, a Moran model for a population of size N is depicted as a graph, where each individual corresponds to a node with at least one 1 and at most $N - 1$ neighbors, depicted as edges between the nodes. The meaning of these edges is that an offspring of a parent can only replace a neighbor of that parent. Whereas a homogeneous population corresponds to a complete graph, where all pairs of individuals are connected, there are many other possibilities. Tkadlec et al. (2018) have shown, when the mutation rate is small, that the speed of evolution is sometimes slightly faster for a certain class of sparse graphs that correspond to geographically subdivided populations. These results are of some relevance for our work, since the fixed state assumption of Section 5 requires a small mutation rate. However, it seems that the order of magnitude of the expected waiting time *cannot* be reduced by imposing geographic substructure.

Fourth, it is also possible to extend our model to two-sex and diploid populations, where each individual has two parents and two copies of a regulatory array. Suppose there are N_1 males and N_2 females in the population, with $N = N_1 + N_2$ so small that the fixed state assumption (17) holds and the regulatory arrays of the population can be represented by a consensus array. It is reasonable then to believe that recombinations between two parental regulatory sequences can be ignored, since both of them are well approximated by the corresponding consensus regulatory sequence. We conjecture that the results of this paper still hold for such a diploid model, if N is replaced by some notion of effective population size N_e for a two-sex and diploid Moran model (Moran, 1958b; Watterson, 1964).

Fifth, we have analyzed the waiting time until a *fixed* and prespecified target is reached; m binding sites within the enhancer/promoter regions of m distinct genes. We also allowed for multiple pathways towards a pre-specified target, by allowing the local targets of all genes to appear in an arbitrary order. In particular, we found that this shortened the waiting time when some back mutations are not allowed ($\gamma < 1$). The waiting time will be shortened even more if there are many possible targets of coordinated mutations (Starr et al., 2017). In Appendix D we have quantified this by describing how to include multiple targets within our framework. In more detail, we analyze the waiting time until *some* target of m genes is reached, when there is a pool of $M \geqslant m$ genes, all of which are part of at least one target of m genes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors wish to thank two anonymous reviewers for helpful comments that improved the manuscript.

Appendix A. Marginal distribution and transition rate of the array component process for the Jukes-Cantor model

Recall that the stationary distribution and transition rate matrix of the array component process $\boldsymbol{C}_t$ in (46) are needed in order to find the phasetype distrubtion (56) of T_m, the waiting time until a global target of m genes has been reached. In Section 9 we analyzed properties of $\boldsymbol{C}_t$, by deriving an explicit formula for its stationary distribution in (68), and expressions for the transition rates due to single mutations (73) and double mutations (77) and (78). In order to further simplify these formulas we need expressions for the probability Π_{jc} in (62) that a randomly chosen word of length W belongs to set $\mathcal{B}_{jc}$ in (38), the corresponding joint probability $\Pi_{jc\eta}$ in (64) that two binding sites at distance η belong to $\mathcal{B}_{jc}$, as well as the one- and two-step transition rates $\theta_{jcd}^{(1)}$ and $\theta_{jcd}^{(2)}$, in (72) and (76), between the disjoint sets $\mathcal{B}_{jc}$ and $\mathcal{B}_{jd}$ of words of length W.

In this appendix we will compute these quantities when the transition probabilities between nucleotides are uniform, as for the Jukes-Cantor model in (2). This implies that all four nucleobases have the same marginal probability 1/4 (cf. (3)), so that the probability of any word $\boldsymbol{b}$ of length W is $\pi_{\boldsymbol{b}} = 4^{-W}$. Consequently, the probability

$$\Pi_{jc} = \frac{|\mathcal{B}_{jc}|}{4^W} \tag{A.1}$$

that a randomly chosen binding site belongs to $\mathcal{B}_{jc}$ is proportional to the number of elements $|\mathcal{B}_{jc}|$ of this set, whereas

$$\Pi_{jc\eta} = \frac{|\mathcal{B}_{jc}|_\eta}{4^{W+\eta}} \tag{A.2}$$

is proportional to the number

$$\begin{aligned} |\mathcal{B}_{jc}|_\eta &= \sum_{\boldsymbol{b}\in\mathcal{B}_{jc}} \sum_{\boldsymbol{b}'\in\mathcal{B}_{jc}} 1\left(\boldsymbol{b}_{\eta+1:W} = \boldsymbol{b}'_{1:W-\eta}\right) \\ &= \sum_{b_{\eta+1},\ldots,b_W} \sum_{b_1,\ldots,b_\eta} 1(\boldsymbol{b} \in \mathcal{B}_{jc}) \\ &\quad \cdot \sum_{b_1,\ldots,b_\eta} 1((b_{\eta+1},\ldots,b_W,b_1,\ldots,b_\eta) \in \mathcal{B}_{jc}) \end{aligned} \tag{A.3}$$

of pairs of binding sites in $\mathcal{B}_{jc}$, whose last and first $W - \eta$ letters agree (with $|\mathcal{B}_{jc}|_0 = |\mathcal{B}_{jc}|$). The equivalent formulation in the second step of (A.3) reveals that it is possible to compute $|\mathcal{B}_{jc}|_\eta$ in $O\left(4^W\right)$ operations. Since the transition probability for the Jukes-Cantor model is $p_{xy} = 1/3$ between any two distinct nucleotides $x \neq y$, given that a mutation occurs, it follows from (72) and (A.1) that

$$\theta_{jcd}^{(1)} = \frac{1}{3|\mathcal{B}_{jc}|} \sum_{\boldsymbol{b}\in\mathcal{B}_{jc}} \sum_{\boldsymbol{b}'\in\mathcal{B}_{jd}} 1(|\boldsymbol{b}' - \boldsymbol{b}| = 1) \tag{A.4}$$

is found by summing over all pairs $\boldsymbol{b}, \boldsymbol{b}'$ with $\boldsymbol{b} \in \mathcal{B}_{jc}$ and $\boldsymbol{b}' \in \mathcal{B}_{jd}$ such that $|\boldsymbol{b}' - \boldsymbol{b}| = 1$. A similar analysis for double mutations, based on (76) and (A.1), reveals that

$$\theta_{jcd}^{(2)} = \frac{1}{9|\mathcal{B}_{jc}|} \sum_{\boldsymbol{b}\in\mathcal{B}_{jc}} \sum_{\boldsymbol{b}'\in\mathcal{B}_{jd}} 1(|\boldsymbol{b} - \boldsymbol{b}| = 2) \tag{A.5}$$

involves summation over pairs $\boldsymbol{b}, \boldsymbol{b}'$ with $\boldsymbol{b} \in \mathcal{B}_{jc}$ and $\boldsymbol{b}' \in \mathcal{B}_{jd}$ such that $|\boldsymbol{b}' - \boldsymbol{b}| = 2$.

In order to find $\Pi_{jc}, \Pi_{jc\eta}, \theta_{jcd}^{(1)}$, and $\theta_{jcd}^{(2)}$, we need to compute the number of elements $|\mathcal{B}_{jc}|$ of $\mathcal{B}_{jc}$ in (A.1), the number $|\mathcal{B}_{jc}|_\eta$ of pairs of binding sites in (A.3), and the double sums that appear in (A.4) and (A.5). To this end, we will assume that the division of the state space $\mathcal{X}$ of regulatory sequences into distinct components is very similar for all promoter regions j. In more detail, we restrict ourselves to the case when the number of targeted binding sites $K_j = K$ and the number of components $C_j = C$ in formula (30) is the same for all $j = 1, \ldots, m$. We will also postulate that for each $c \in \{0, \ldots, C-1\}$, the corresponding component $\mathcal{X}_{jc}$ of regulatory sequences is based on the same set of distances to the closest binding site for all $j = 1, \ldots, m$, so that (33) simplifies to

$$\Delta_{jc} = \Delta_c = \{\delta_c, \ldots, \delta_{c+1} - 1\}, \tag{A.6}$$

for some pre-specified numbers $0 = \delta_0 < \delta_1 < \ldots < \delta_C = W + 1$. In particular, if $\delta_{C'} = \delta_{\max} + 1$ for some $1 \leqslant C' < C$, it follows that the first C' components $\mathcal{X}_{j0}, \ldots, \mathcal{X}_{j,C'-1}$ correspond to a reached target for gene j. Therefore, the type $g(\boldsymbol{c})$ of all regulatory arrays within component $\underline{\mathcal{X}}_{\boldsymbol{c}}$ simplifies from (35) to

$$g(\boldsymbol{c}) = \begin{cases} \min\{j;\, 1 \leqslant j \leqslant m, c_j \geqslant C'\} - 1, & \text{for fixed order of TA}, \\ \sum_{j=1}^{m} 1(c_j < C'), & \text{for arbitrary order of TA}. \end{cases}$$

The set of binding sites in (38), with distances (33) to the target of enhancer/promoter region j, takes the form

$$\mathcal{B}_{jc} = \left\{\boldsymbol{b} \in \mathcal{B}_{\text{all}};\, \delta_c \leqslant \min_{1 \leqslant k \leqslant K} |\boldsymbol{b} - \boldsymbol{b}_{jk}| \leqslant \delta_{c+1} - 1\right\}. \tag{A.7}$$

Since the targeted binding sites $\boldsymbol{b}_{j1},\ldots,\boldsymbol{b}_{jK}$ may vary with region, so will $\mathcal{B}_{jc}$ and $\mathcal{X}_{jc}$. But it turns out that $|\mathcal{B}_{jc}|$, $\theta^{(1)}_{jcd}$, and $\theta^{(2)}_{jcd}$ will still have a tractable form when K equals 1 or 2. In particular, we know from formulas (73), (77), and (78) that in order to find $\theta^{(1)}_{jcd}$ it suffices to consider neighboring sets $\mathcal{B}_{jc}$ and $\mathcal{B}_{jd}$ of binding sites with $|d-c|=1$, whereas for $\theta^{(2)}_{jcd}$ we also need to analyze instances when $|d-c|=2$. In addition, since (A.4) and (A.5) are symmetric in $\mathcal{B}_{jc}$ and $\mathcal{B}_{jd}$, we have that

$$\begin{aligned} |\mathcal{B}_{jc}|\theta^{(1)}_{jc,c-1} &= |\mathcal{B}_{j,c-1}|\theta^{(1)}_{j,c-1,c}, \\ |\mathcal{B}_{jc}|\theta^{(2)}_{jc,c-2} &= |\mathcal{B}_{j,c-2}|\theta^{(2)}_{j,c-2,c}. \end{aligned} \tag{A.8}$$

For this reason, we will only give expressions for $\theta^{(1)}_{jcd}$ when $d=c-1$, and for $\theta^{(2)}_{jcd}$ when $d=c-2$.

Our analyses depend crucially on the number K of targeted binding sites per region j. In the next three subsections of this appendix we will consider the cases $K=1, K=2$, and arbitrary K separately.

A.1. Each local target consists of one binding site ($K = 1$)

When $K=1$ there is one targeted binding site $\boldsymbol{b}_j=\boldsymbol{b}_{j1}$ of each regulatory region j. Formula (A.7) simplifies to

$$\mathcal{B}_{jc}=\left\{\boldsymbol{b}\in\mathcal{B}_{\text{all}};\, \delta_c\leqslant|\boldsymbol{b}-\boldsymbol{b}_j|\leqslant\delta_{c+1}-1\right\}. \tag{A.9}$$

Since the nucleotides of a word of length W are chosen independently with uniform probabilities $1/4$, its number of mismatches with the targeted binding site $\boldsymbol{b}_j$ has a binomial distribution with parameters W and $3/4$. From this it follows that the probability (A.1) that a randomly chosen word of length W belongs to $\mathcal{B}_{jc}$, can be written as

$$\Pi_{jc}=\frac{|\mathcal{B}_{jc}|}{4^W}=\frac{1}{4^W}\sum_{\delta=\delta_c}^{\delta_{c+1}-1}\binom{W}{\delta}3^{\delta}. \tag{A.10}$$

In order to obtain the one-step transition rate $\theta^{(1)}_{jc,c-1}$ between $\mathcal{B}_{jc}$ and $\mathcal{B}_{j,c-1}$, we must first find the number of pairs $\boldsymbol{b}\in\mathcal{B}_{jc}$ and $\boldsymbol{b}'\in\mathcal{B}_{j,c-1}$ with $|\boldsymbol{b}'-\boldsymbol{b}|=1$, such that a single letter change from $\boldsymbol{b}$ to $\boldsymbol{b}'$ decreases the distance to $\boldsymbol{b}_j$ from δ_c to δ_c-1. There are $\binom{W}{\delta_c}3^{\delta_c}$ ways of choosing $\boldsymbol{b}$, and for each such $\boldsymbol{b}$ there are δ_c ways of selecting $\boldsymbol{b}'$, since any of the δ_c letters of $\boldsymbol{b}$ that do not match $\boldsymbol{b}_j$ can be changed in a unique way in order to equal the corresponding letter of $\boldsymbol{b}_j$. Multiplying these two numbers, and then dividing by the normalizing factor $3|\mathcal{B}_{jc}|$ in (A.4), we find that

$$\theta^{(1)}_{jc,c-1}=\frac{\binom{W}{\delta_c}3^{\delta_c}\delta_c}{3|\mathcal{B}_{jc}|}. \tag{A.11}$$

When $d=c-2$, we need to find the number of pairs $\boldsymbol{b}\in\mathcal{B}_{jc}$ and $\boldsymbol{b}'\in\mathcal{B}_{j,c-2}$, with $|\boldsymbol{b}'-\boldsymbol{b}|=2$, such that changing two letters of $\boldsymbol{b}$ decreases its distance to $\boldsymbol{b}_j$ from δ_c to δ_c-2. This is only possible if $\delta_{c-1}=\delta_c-1$ (otherwise the second mutation will be silent). Then there are $\binom{W}{\delta_c}3^{\delta_c}$ ways of choosing $\boldsymbol{b}$, and for each such $\boldsymbol{b}$ there are $\binom{\delta_c}{2}$ ways of choosing $\boldsymbol{b}'$, by changing any two of the δ_c positions of $\boldsymbol{b}$ that do not match $\boldsymbol{b}_1$. Multiplying these two numbers, and then dividing by the normalizing factor $9|\mathcal{B}_{jc}|$ in (A.5), we find that

$$\theta^{(2)}_{jc,c-2}=\begin{cases}0, & \delta_{c-1}<\delta_c-1,\\ \binom{W}{\delta_c}3^{\delta_c}\binom{\delta_c}{2}/\left(9|\mathcal{B}_{jc}|\right), & \delta_{c-1}=\delta_c-1.\end{cases} \tag{A.12}$$

A.2. Each local target consists of two binding sites ($K = 2$)

When $K=2$ each regulatory region j has two targeted binding sites $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$. This implies that the c:th set of binding sites in (A.7), can be written as

$$\mathcal{B}_{jc}=\left\{\boldsymbol{b}\in\mathcal{B}_{\text{all}};\, \delta_c\leqslant\delta_j(\boldsymbol{b})\leqslant\delta_{c+1}-1\right\}, \tag{A.13}$$

where $\delta_j(\boldsymbol{b})=\min\left(|\boldsymbol{b}-\boldsymbol{b}_{j1}|,|\boldsymbol{b}-\boldsymbol{b}_{j2}|\right)$ is the distance from $\boldsymbol{b}$ to the closest targeted binding site. The three quantities of interest, $\Pi_{jc}, \theta^{(1)}_{jc,c-1}$, and $\theta^{(2)}_{c,c-2}$, will all depend on the number of letters $0<W_j=|\boldsymbol{b}_{j2}-\boldsymbol{b}_{j1}|\leqslant W$ where $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$ differ. The number of mismatches between $\boldsymbol{b}$ and the closest targeted binding site, can be written as

$$\begin{aligned} \delta_j(\boldsymbol{b}) &= V_3+\min\left(W_j-V_1,W_j-V_2\right)\\ &= W_j+V_3-\max(V_1,V_2) \end{aligned}$$

where V_k $(0\leqslant V_k\leqslant W_j)$ is the number of letters where $\boldsymbol{b}$ agrees with $\boldsymbol{b}_{jk}$, for $k=1,2$, among those W_j positions where $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$ differ, whereas $0\leqslant V_3\leqslant W-W_j$ is the number of letters where $\boldsymbol{b}$ differs from both of $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$, among those $W-W_j$ positions where $\boldsymbol{b}_{j1}$ equals $\boldsymbol{b}_{j2}$. This is illustrated in Fig. 6.

Notice that V_1,V_2,V_3 depend on $\boldsymbol{b}$ as well as j, although this will not be made explicit in the notation. In the calculations below, we will divide the W letters of $\boldsymbol{b}$ into five regions of size $W-W_j-V_3, V_1, V_2, V_3$, and $W_j-V_1-V_2$. These five regions refer to whether $\boldsymbol{b}$ matches both, only one or none of the targeted binding sites $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$. Since $\left(V_1,V_2,W_j-V_1-V_2\right)$ has a multinomial distribution with W_j trials and cell probabilities 1/4, 1/4, and 2/4 for a randomly chosen $\boldsymbol{b}$, whereas V_3 has a binomial distribution with number of trials $W-W_j$ and probability $3/4$, for a randomly chosen $\boldsymbol{b}$, it follows that

$$\begin{aligned} \Pi_{jc} &= \sum_{v_1,v_2,v_3\in\mathcal{V}_{jc}} P(V_1=v_1,V_2=v_2)P(V_3=v_3)\\ &= \sum_{v_1,v_2,v_3\in\mathcal{V}_{jc}} \frac{W_j!}{v_1!v_2!\left(W_j-v_1-v_2\right)!}(1/4)^{v_1+v_2}(2/4)^{W_j-v_1-v_2}\\ &\qquad\cdot\binom{W-W_j}{v_3}(1/4)^{W-W_j-v_3}(3/4)^{v_3}\\ &= 4^{-W}\sum_{v_1,v_2,v_3\in\mathcal{V}_{jc}}\frac{W_j!\left(W-W_j\right)!}{v_1!v_2!v_3!\left(W_j-v_1-v_2\right)!\left(W-W_j-v_3\right)!}2^{W_j-v_1-v_2}3^{v_3}, \end{aligned}$$

where

$$\mathcal{V}_{jc}=\left\{(v_1,v_2,v_3);\,\delta_c\leqslant W_j+v_3-\max(v_1,v_2)\leqslant\delta_{c+1}-1\right\}.$$

In order to get an explicit expression for the one-step transition rate $\theta^{(1)}_{jc,c-1}$ between $\mathcal{B}_{jc}$ and $\mathcal{B}_{j,c-1}$, we need to find the number of pairs $\boldsymbol{b}\in\mathcal{B}_{jc}$ and $\boldsymbol{b}'\in\mathcal{B}_{j,c-1}$ with $|\boldsymbol{b}'-\boldsymbol{b}|=1$, such that a single changed

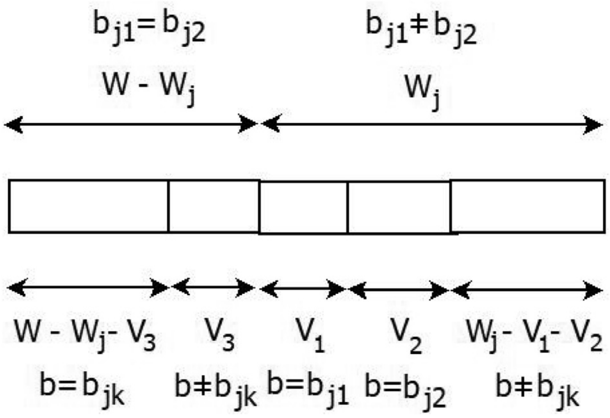


Fig. 6. Illustration of how the W letters of a word $\boldsymbol{b}$ is divided into 5 blocks depending on how they relate to the two possible targeted binding sites $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$ of length W in the regulatory region of gene j. From the upper part of the figure we see that $0<W_j\leqslant W$ and $W-W_j$ are the number of letters where $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$ differ and agree. From the lower part of the figure we notice that these two subwords are further divided into 2 or 3 smaller words, depending on how $\boldsymbol{b}$ relates to $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$.

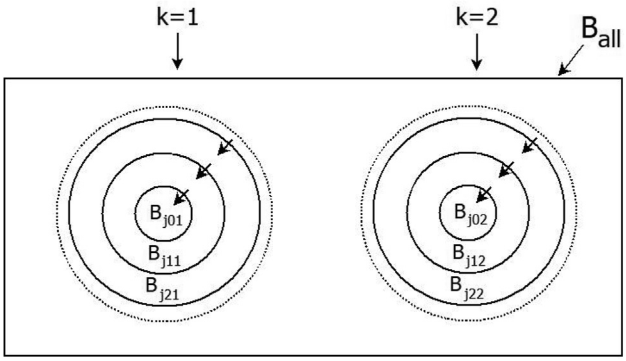


Fig. 7. Illustration of how the set $\mathcal{B}_{\text{all}}$ of words of length W clusters into $K = 2$ regions when the corresponding targeted binding sites $\boldsymbol{b}_{j1}, \ldots, \boldsymbol{b}_{jK}$ of gene j are so widely dispersed that the $C = 4$ distance classes satisfy (A.16). The regions between the solid concentric circles correspond to the sets $\mathcal{B}_{jck}$ in (A.18), whereas those words whose distance to the closest binding site is δ_{C-1}, correspond to the regions between the outermost solid circles and the dotted circles. The short diagonal arrows that cross the circles represent the transition rates $\theta_{32}^{(1)}$ (outer), $\theta_{21}^{(1)}$ (middle) and $\theta_{10}^{(1)}$ (inner).

letter from $\boldsymbol{b}$ to $\boldsymbol{b}'$ decreases $\delta_j(\boldsymbol{b})$ from δ_c to $\delta_c - 1$. For each such $\boldsymbol{b}$ we compute the numbers v_1, v_2, v_3 defined above, and find from (A.4) that

$$\begin{aligned}\theta_{jc,c-1}^{(1)} \quad = \quad & \frac{1}{3|\mathcal{B}_{jc}|} \sum_{v_1,v_2,v_3 \in \mathcal{V}'_{jc}} \frac{W_j!(W-W_j)!}{v_1!v_2!v_3!(W_j-v_1-v_2)!(W-W_j-v_3)!} 2^{W_j-v_1-v_2} 3^{v_3} \\ & \cdot (W_j + v_3 - v_1)^{1(v_1>v_2)} (2W_j + v_3 - v_1 - v_2)^{1(v_1=v_2)} \\ & \cdot (W_j + v_3 - v_2)^{1(v_1<v_2)},\end{aligned} \tag{A.14}$$

where summation is over the set

$$\mathcal{V}'_{jc} = \{(v_1, v_2, v_3);\ W_j + v_3 - \max(v_1, v_2) = \delta_c\}.$$

We can motivate (A.14) by noticing that the upper row terms of the summation on the right hand side of (A.14) is the number of words $\boldsymbol{b}$ that match the corresponding triple (v_1, v_2, v_3). Given $\boldsymbol{b}$, the number of possible changes of letters that move this word one step closer to the target will depend on (v_1, v_2, v_3). The number of such changes is $v_2 + (W_j - v_1 - v_2) + v_3 = W_j + v_3 - v_1$ when $v_1 > v_2$, by symmetry it is $W_j + v_3 - v_2$ when $v_1 < v_2$, and it is $v_1 + v_2 + v_3 + 2(W_j - v_1 - v_2) = 2W_j + v_3 - v_1 - v_2$ in the case when $v_1 = v_2$, since at all the $W_j - v_1 - v_2$ positions where $\boldsymbol{b}$ differs from $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$, it is possible to have a change that matches $\boldsymbol{b}_{j1}$ or $\boldsymbol{b}_{j2}$.

As a next step we will derive an expression for $\theta_{jc,c-2}^{(2)}$, and for this we need the number of pairs $\boldsymbol{b}, \boldsymbol{b}'$ with $|\boldsymbol{b}' - \boldsymbol{b}| = 2$ such that two mutations change $\boldsymbol{b}$ into $\boldsymbol{b}'$ and thereby decreases $\delta_j(\boldsymbol{b})$ from δ_c to $\delta_c - 2$, if $\delta_{c-1} = \delta_c - 1$. We find from (A.5) that

Table 5
Values of the expected waiting time $E(T_m)$ until the binding site targets of all $m = 1$ or $m = 2$ genes appear for a neutral model ($s_0 = \ldots = s_m = 1$), and different combinations of binding site length W, maximal mismatch $\delta_{\max}$, and a regulatory array decomposition (31) that uses distance intervals $\Delta_0, \ldots, \Delta_{C-1}$ in (A.6) for the number of mismatches (recall from the paragraph below (34) that $\delta_{\max}$ has to be a right end point of one of these intervals). All other parameters are given by Table 3. The last column displays the expected values of the minus one hit variable $H_{\text{m}1} = H_{\{\delta_{\max}+1\}}(\boldsymbol{X}_{0j})$, that is, the expected value of the number of substrings of a regulatory sequence of length W that miss the local target of gene j by one letter.

W	$\delta_{\max}$	C	$\Delta_0, \ldots, \Delta_{C-1}$	$m = 1$	$m = 2$	$E(H_{\text{m}1})$
6	0	2	$\{0\}, \{1,2,3,4,5,6\}$	$5.3813 \cdot 10^7$	$2.0518 \cdot 10^8$	4.37
		3	$\{0\}, \{1\}, \{2,3,4,5,6\}$	$5.3857 \cdot 10^7$	$2.0532 \cdot 10^8$	
		4	$\{0\}, \{1\}, \{2\}, \{3,4,5,6\}$	$5.3857 \cdot 10^7$	$2.0532 \cdot 10^8$	
		5	$\{0\}, \{1\}, \{2\}, \{3\}, \{4,5,6\}$	$5.3857 \cdot 10^7$	$2.0532 \cdot 10^8$	
	1	2	$\{0,1\}, \{2,3,4,5,6\}$	$4.5271 \cdot 10^4$	$9.0758 \cdot 10^4$	32.79
		3	$\{0,1\}, \{2\}, \{3,4,5,6\}$	$4.5271 \cdot 10^4$	$9.0758 \cdot 10^4$	
		4	$\{0,1\}, \{2\}, \{3\}, \{4,5,6\}$	$4.5271 \cdot 10^4$	$9.0758 \cdot 10^4$	
	2	2	$\{0,1,2\}, \{3,4,5,6\}$	$4.318 \cdot 10^{-11}$	$8.637 \cdot 10^{-11}$	131.18
		3	$\{0,1,2\}, \{3\}, \{4,5,6\}$	$4.318 \cdot 10^{-11}$	$8.637 \cdot 10^{-11}$	
8	0	2	$\{0\}, \{1,2,3,4,5,6,7,8\}$	$8.1257 \cdot 10^8$	$2.8186 \cdot 10^{10}$	0.36
		3	$\{0\}, \{1\}, \{2,3,4,5,6,7,8\}$	$8.3810 \cdot 10^8$	$2.8659 \cdot 10^{10}$	
		4	$\{0\}, \{1\}, \{2\}, \{3,4,5,6,7,8\}$	$8.3816 \cdot 10^8$	$2.8660 \cdot 10^{10}$	
		5	$\{0\}, \{1\}, \{2\}, \{3\}, \{4,5,6,7,8\}$	$8.3816 \cdot 10^8$	$2.8660 \cdot 10^{10}$	
	1	2	$\{0,1\}, \{2,3,4,5,6,7,8\}$	$2.6897 \cdot 10^7$	$8.2858 \cdot 10^7$	3.82
		3	$\{0,1\}, \{2\}, \{3,4,5,6,7,8\}$	$2.6962 \cdot 10^7$	$8.3035 \cdot 10^7$	
		4	$\{0,1\}, \{2\}, \{3\}, \{4,5,6,7,8\}$	$2.6962 \cdot 10^7$	$8.3035 \cdot 10^7$	
	2	2	$\{0,1,2\}, \{3,4,5,6,7,8\}$	$6.5645 \cdot 10^4$	$1.3177 \cdot 10^5$	22.91
		3	$\{0,1,2\}, \{3\}, \{4,5,6,7,8\}$	$6.5645 \cdot 10^4$	$1.3177 \cdot 10^5$	
10	0	2	$\{0\}, \{1,2,3,4,5,6,7,8,9,10\}$	$1.0571 \cdot 10^{10}$	$5.5986 \cdot 10^{12}$	0.03
		3	$\{0\}, \{1\}, \{2,3,4,5,6,7,8,9,10\}$	$1.0920 \cdot 10^{10}$	$5.6850 \cdot 10^{12}$	
		4	$\{0\}, \{1\}, \{2\}, \{3,4,5,6,7,8,9,10\}$	$1.0941 \cdot 10^{10}$	$5.6864 \cdot 10^{12}$	
		5	$\{0\}, \{1\}, \{2\}, \{3\}, \{4,5,6,7,8,9,10\}$	$1.0941 \cdot 10^{10}$	$5.6864 \cdot 10^{12}$	
	1	2	$\{0,1\}, \{2,3,4,5,6,7,8,9,10\}$	$3.8057 \cdot 10^8$	$7.1473 \cdot 10^9$	0.38
		3	$\{0,1\}, \{2\}, \{3,4,5,6,7,8,9,10\}$	$4.0143 \cdot 10^8$	$7.3738 \cdot 10^9$	
		4	$\{0,1\}, \{2\}, \{3\}, \{4,5,6,7,8,9,10\}$	$4.0157 \cdot 10^8$	$7.3746 \cdot 10^9$	
	2	2	$\{0,1,2\}, \{3,4,5,6,7,8,9,10\}$	$2.1628 \cdot 10^7$	$6.4344 \cdot 10^7$	3.06
		3	$\{0,1,2\}, \{3\}, \{4,5,6,7,8,9,10\}$	$2.1770 \cdot 10^7$	$6.4722 \cdot 10^7$	

$$\begin{aligned}\theta^{(2)}_{jc,c-2} &= 1(\delta_{c-1}=\delta_c-1)/(9|\mathcal{B}_{jc}|)\\ &\cdot \sum_{v_1,v_2,v_3\in\mathcal{V}'_{jc}} \frac{W_j!(W-W_j)!}{v_1!v_2!v_3!(W_j-v_1-v_2)!(W-W_j-v_3)!}2^{W_j-v_1-v_2}3^{v_3}\\ &\cdot \binom{W_j+v_3-v_1}{2}^{1(v_1>v_2)}\binom{W_j+v_3-v_2}{2}^{1(v_1<v_2)}\\ &\cdot\left[\binom{W_j+v_3}{2}-v_1v_2+v_3(W_j-v_1-v_2)+\binom{W_j-v_1-v_2}{2}\right]^{1(v_1=v_2)}.\end{aligned} \tag{A.15}$$

Eq. (A.15) is analogous to (A.14), and in particular the last two rows give the number of ways of changing the letters of $\boldsymbol{b}$, so that this word is moved two steps closer to the target of gene j, for different choices of (v_1,v_2,v_3). When $v_1>v_2$, we choose the two mutations of $\boldsymbol{b}$ from any of the $v_3+v_2+(W_j-v_1-v_2)=W_j+v_3-v_1$ positions where $\boldsymbol{b}$ either differs from both of $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$, or where $\boldsymbol{b}$ differs from $\boldsymbol{b}_{j1}$ only. This gives $\binom{W_j+v_3-v_1}{2}$ ways of choosing $\boldsymbol{b}'$. By symmetry it follows that there are $\binom{W_j+v_3-v_2}{2}$ ways of selecting $\boldsymbol{b}'$ when $v_1<v_2$. The case $v_1=v_2$ is more complicated, since the positions where $\boldsymbol{b}$ differs from exactly one of $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$, or positions where $\boldsymbol{b}$ differs from a common value of $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$, can be changed in one way only in order to decrease the distance to the closest targeted binding site, whereas the $W_j-v_1-v_2$ positions where $\boldsymbol{b}$ differs from both of $\boldsymbol{b}_{j1}$ and $\boldsymbol{b}_{j2}$ (when these are distinct), can be changed in two ways. From this it follows that total number of ways to choose exactly two mutations in $\boldsymbol{b}$ is $\binom{v_3}{2}+\binom{v_1}{2}+\binom{v_2}{2}+2\binom{W_j-v_1-v_2}{2}+v_3v_1+v_3v_2+2v_3(W_j-v_1-v_2)+v_1(W_j-v_1-v_2)+v_2(W_j-v_1-v_2)$, which equals the last expression within squared brackets in (A.15).

A.3. Each local target consists of any number K of binding sites

When $K>2$, the method of Appendix A.2 is computationally unfeasible; to bookkeep all the ways in which the targeted binding sites $\boldsymbol{b}_{j1},\ldots,\boldsymbol{b}_{jK}$ of gene j may overlap.

A general brute force approach of computing $\Pi_{jc}, \theta^{(1)}_{jcd}$ and $\theta^{(2)}_{jcd}$, is to enumerate all binding sites $\boldsymbol{b}\in\mathcal{B}_{jc}$ or to list all neighboring pairs $\boldsymbol{b}\in\mathcal{B}_{jc}$ and $\boldsymbol{b}'\in\mathcal{B}_{jd}$ of binding sites for which $|\boldsymbol{b}'-\boldsymbol{b}|$ equals 1 or 2. According to (A.1), (A.4), and (A.5), this requires $|\mathcal{B}_{jc}|, 3W|\mathcal{B}_{jc}|$, and $9\binom{W}{2}|\mathcal{B}_{jc}|$ operations respectively, since each $\boldsymbol{b}\in\mathcal{B}_{jc}$ has $3W$ neighboring binding sites $\boldsymbol{b}'$ at distance 1, and $9\binom{W}{2}$ binding sites $\boldsymbol{b}'$ at distance 2. Since $\sum_c|\mathcal{B}_{jc}|=4^W$, it follows that the total complexity to compute all $\Pi_{jc}, \theta^{(1)}_{jcd}$, and $\theta^{(2)}_{jcd}$ for region j, is of order $4^W, 3W4^W$, and $9W(W-1)4^{W-1/2}$ respectively. Although the computational complexity of this approach is independent of K, it is till only feasible for moderately large values of W.

It is possible though to find a rapid way of computing $\Pi_{jc}, \theta^{(1)}_{jcd}$, and $\theta^{(2)}_{jcd}$, if the targeted binding sites $\boldsymbol{b}_{j1},\ldots,\boldsymbol{b}_{jK}$ of gene j are sufficiently dispersed among all words $\mathcal{B}_{\text{all}}$ of length W. More specifically, let $\mathcal{B}(\boldsymbol{b},\delta)=\{\boldsymbol{b}'\in\mathcal{B}_{\text{all}};\,|\boldsymbol{b}'-\boldsymbol{b}|\leqslant\delta\}$ refer to the ball of radius δ around $\boldsymbol{b}$. We will assume that the K balls

$$\mathcal{B}(\boldsymbol{b}_{j1},\delta_{C-1}),\ldots,\mathcal{B}(\boldsymbol{b}_{jK},\delta_{C-1}) \text{ are disjoint,} \tag{A.16}$$

and a necessary condition for this to hold is that

$$K\sum_{\delta=0}^{\delta_{C-1}}\binom{W}{\delta}3^\delta\leqslant 4^W. \tag{A.17}$$

Recall from (A.7) that $\mathcal{B}_{jc}$ is the set of words of length W whose distance to the closest targeted binding site of gene j belongs to interval Δ_c in (A.6). As illustrated in Fig. 7, a consequence of (A.16) is that $\mathcal{B}_{jc}=\cup_{k=1}^K\mathcal{B}_{jck}$, when $0\leqslant c\leqslant C-2$, is a disjoint union of the sets

$$\mathcal{B}_{jck}=\left\{\boldsymbol{b}\in\mathcal{B}_{\text{all}};\,\delta_c\leqslant|\boldsymbol{b}-\boldsymbol{b}_{jk}|\leqslant\delta_{c+1}-1\right\}, \tag{A.18}$$

with words having a distance to $\boldsymbol{b}_{jk}$ that belongs Δ_c. Therefore, the probability that a randomly chosen word has a distance in Δ_c to the closest binding site is found (when $0\leqslant c\leqslant C-2$) by first computing the probability (A.10) that the distance to one specific binding site belongs to Δ_c, and then multiplying this number by K, i.e.

$$\Pi_{jc}=\frac{|\mathcal{B}_{jc}|}{4^W}=\begin{cases}K\sum_{\delta=\delta_c}^{\delta_{c+1}-1}\binom{W}{\delta}3^\delta/4^W, & c=0,1,\ldots,C-2,\\ 1-(\Pi_{j0}+\ldots+\Pi_{j,C-2}), & c=C-1.\end{cases} \tag{A.19}$$

Likewise, (A.16) implies that any word $\boldsymbol{b}$ of length W has distance δ_{C-1} to at most one targeted binding site of gene j. Therefore, we find from (A.11) and (A.12) that

$$\theta^{(1)}_{jc,c-1}=\frac{K\binom{W}{\delta_c}3^{\delta_c}\delta_c}{3|\mathcal{B}_{jc}|} \tag{A.20}$$

for $c=1,2,\ldots,C-1$, and

$$\theta^{(2)}_{jc,c-2}=\begin{cases}0, & \delta_{c-1}<\delta_c-1,\\ K\binom{W}{\delta_c}3^{\delta_c}\binom{\delta_c}{2}/(9|\mathcal{B}_{jc}|), & \delta_{c-1}=\delta_c,\end{cases} \tag{A.21}$$

for $c=2,3,\ldots,C-1$, whenever (A.16) holds.

Appendix B. Further numerical waiting time results

In this appendix we will continue the analysis of Section 10 by giving further numerical illustrations of the distribution of T_m, the waiting time until targeted binding sites in the regulatory regions of m genes are reached. In particular, we will continue using the default parameter setting of Table 3 and vary one or a few parameters at a time.

B.1. Varying coarseness of the array components

Recall that the finer the array component decomposition (24) is, the more accurate is the Markov approximation of the array component process $\boldsymbol{C}_t$ and the better the distribution of the waiting time T_m is approximated by the phase-type distribution formula (57). Here in Appendix B.1 we will investigate how the coarseness of the array component decomposition affects the expected value and quantiles of T_m, for some systems with $m=1$ and $m=2$ genes. This numerical study is accompanied by a more theoretical study in Appendix C.4, where explicit conditions are given under which the Markov process approximation of $\boldsymbol{C}_t$ is accurate.

B.1.1. Array components based on number of mismatches

B.1.1.1. Neutral models

Recall that the distance (14) to the local target at each gene is defined in terms of the number of mismatches between substrings of the regulatory sequence and the closest binding site of that gene. The corresponding array decomposition (31) records, for each gene, to which interval $\Delta_0,\ldots,\Delta_{C-1}$ in (A.6) this distance belongs. It follows from (32) that the regulatory array component process $\boldsymbol{C}_t$ has a state space whose size $|\mathcal{C}_{\text{mism}}|=C^m$ grows exponentially with the number of genes m. It is therefore important to keep C as small as possible, and at the same time not loose accuracy.

Table 5 displays values of $E(T_1)$ and $E(T_2)$ for various combinations of word length W and maximal mismatch $\delta_{\max}$. The model is selectively neutral model ($s_0=\ldots=s_m=1$), so that double mutations can effectively be ignored. For each combination of W and $\delta_{\max}$, the number C of distance intervals is varied. It can be seen from this table that C has a very small impact on the expected

Table 6
Values of the expected waiting time $E(T_m)$ until the binding site targets of all $m = 1$ or $m = 2$ genes appear for a non-neutral model (96) with $s_0 = \ldots = s_{m-1} = 1$ and $s_m = 10$. All other details are the same as in Table 5.

W	$\delta_{\max}$	C	$\Delta_0, \ldots, \Delta_{C-1}$	$m = 1$	$m = 2$	$E(H_{m1})$
6	0	2	$\{0\}, \{1,2,3,4,5,6\}$	$5.9792 \cdot 10^{3}$	$7.3978 \cdot 10^{6}$	4.37
		3	$\{0\}, \{1\}, \{2,3,4,5,6\}$	$3.6635 \cdot 10^{4}$	$7.4818 \cdot 10^{6}$	
		4	$\{0\}, \{1\}, \{2\}, \{3,4,5,6\}$	$3.6635 \cdot 10^{4}$	$7.4818 \cdot 10^{6}$	
		5	$\{0\}, \{1\}, \{2\}, \{3\}, \{4,5,6\}$	$3.6635 \cdot 10^{4}$	$7.4818 \cdot 10^{6}$	
	1	2	$\{0,1\}, \{2,3,4,5,6\}$	$5.0301 \cdot 10^{0}$	$1.7735 \cdot 10^{1}$	32.79
		3	$\{0,1\}, \{2\}, \{3,4,5,6\}$	$5.0301 \cdot 10^{0}$	$1.7735 \cdot 10^{1}$	
		4	$\{0,1\}, \{2\}, \{3\}, \{4,5,6\}$	$5.0301 \cdot 10^{0}$	$1.7735 \cdot 10^{1}$	
	2	2	$\{0,1,2\}, \{3,4,5,6\}$	$4.7983 \cdot 10^{-15}$	$9.5966 \cdot 10^{-15}$	131.18
		3	$\{0,1,2\}, \{3\}, \{4,5,6\}$	$4.7983 \cdot 10^{-15}$	$9.5966 \cdot 10^{-15}$	
8	0	2	$\{0\}, \{1,2,3,4,5,6,7,8\}$	$9.0285 \cdot 10^{4}$	$3.4902 \cdot 10^{8}$	0.36
		3	$\{0\}, \{1\}, \{2,3,4,5,6,7,8\}$	$1.8297 \cdot 10^{7}$	$6.4869 \cdot 10^{8}$	
		4	$\{0\}, \{1\}, \{2\}, \{3,4,5,6,7,8\}$	$1.8362 \cdot 10^{7}$	$6.4917 \cdot 10^{8}$	
		5	$\{0\}, \{1\}, \{2\}, \{3\}, \{4,5,6,7,8\}$	$1.8362 \cdot 10^{7}$	$6.4917 \cdot 10^{8}$	
	1	2	$\{0,1\}, \{2,3,4,5,6,7,8\}$	$2.9886 \cdot 10^{3}$	$2.1745 \cdot 10^{6}$	3.82
		3	$\{0,1\}, \{2\}, \{3,4,5,6,7,8\}$	$3.6609 \cdot 10^{4}$	$2.2792 \cdot 10^{6}$	
		4	$\{0,1\}, \{2\}, \{3\}, \{4,5,6,7,8\}$	$3.6609 \cdot 10^{4}$	$2.2792 \cdot 10^{6}$	
	2	2	$\{0,1,2\}, \{3,4,5,6,7,8\}$	$7.2939 \cdot 10^{0}$	$3.0741 \cdot 10^{1}$	22.91
		3	$\{0,1,2\}, \{3\}, \{4,5,6,7,8\}$	$7.2939 \cdot 10^{0}$	$3.0741 \cdot 10^{1}$	
10	0	2	$\{0\}, \{1,2,3,4,5,6,7,8,9,10\}$	$1.1746 \cdot 10^{6}$	$5.8251 \cdot 10^{9}$	0.03
		3	$\{0\}, \{1\}, \{2,3,4,5,6,7,8,9,10\}$	$2.5867 \cdot 10^{8}$	$7.3885 \cdot 10^{10}$	
		4	$\{0\}, \{1\}, \{2\}, \{3,4,5,6,7,8,9,10\}$	$2.7992 \cdot 10^{8}$	$7.5295 \cdot 10^{10}$	
		5	$\{0\}, \{1\}, \{2\}, \{3\}, \{4,5,6,7,8,9,10\}$	$2.8006 \cdot 10^{8}$	$7.5298 \cdot 10^{10}$	
	1	2	$\{0,1\}, \{2,3,4,5,6,7,8,9,10\}$	$4.2286 \cdot 10^{4}$	$1.3994 \cdot 10^{8}$	0.38
		3	$\{0,1\}, \{2\}, \{3,4,5,6,7,8,9,10\}$	$1.1117 \cdot 10^{7}$	$2.4471 \cdot 10^{8}$	
		4	$\{0,1\}, \{2\}, \{3\}, \{4,5,6,7,8,9,10\}$	$1.1261 \cdot 10^{7}$	$2.4540 \cdot 10^{8}$	
	2	2	$\{0,1,2\}, \{3,4,5,6,7,8,9,10\}$	$2.4032 \cdot 10^{3}$	$1.4646 \cdot 10^{6}$	3.06
		3	$\{0,1,2\}, \{3\}, \{4,5,6,7,8,9,10\}$	$6.1932 \cdot 10^{4}$	$1.6782 \cdot 10^{6}$	

Table 7
Values of the expected value $E(T_1)$, median $F_{T_1}^{-1}(0.5)$ and 0.99 quantile $F_{T_1}^{-1}(0.99)$ of the waiting time T_1 until the targeted binding site of $m = 1$ gene appears for a non-neutral model with $s_0 = 1$ and $s_1 = 10$. No mismatches with the targeted binding site are allowed ($\delta_{\max} = 0$), and a mismatch-based regulatory sequence decomposition (31) is used with $C = 2$ or $C = 3$ intervals $\Delta_0 = \{0\}, \ldots, \Delta_{C-1} = \{C-2\}$, and $\Delta_C = \{C-1, \ldots, W\}$. When $C = 3$, the impact of stochastic tunneling (ST) is shown. See also Table 6 for comparison.

W	C	ST	$E(T_1)$	$F_{T_1}^{-1}(0.5)$	$F_{T_1}^{-1}(0.99)$
6	2	No	$5.9792 \cdot 10^{3}$	$3.4322 \cdot 10^{3}$	$3.3255 \cdot 10^{4}$
	3	No	$5.1250 \cdot 10^{4}$	$3.4438 \cdot 10^{3}$	$6.2047 \cdot 10^{4}$
	3	Yes	$3.6635 \cdot 10^{4}$	$3.4438 \cdot 10^{3}$	$6.0190 \cdot 10^{4}$
7	2	No	$2.4623 \cdot 10^{4}$	$1.6548 \cdot 10^{4}$	$1.1890 \cdot 10^{5}$
	3	No	$3.4681 \cdot 10^{6}$	$1.9821 \cdot 10^{4}$	$4.2845 \cdot 10^{7}$
	3	Yes	$2.3559 \cdot 10^{6}$	$1.9805 \cdot 10^{4}$	$2.9053 \cdot 10^{7}$
8	2	No	$9.0285 \cdot 10^{4}$	$6.2148 \cdot 10^{4}$	$4.2074 \cdot 10^{5}$
	3	No	$2.6988 \cdot 10^{7}$	$1.2425 \cdot 10^{7}$	$1.6636 \cdot 10^{8}$
	3	Yes	$1.8297 \cdot 10^{7}$	$8.4226 \cdot 10^{6}$	$1.1277 \cdot 10^{8}$
9	2	No	$3.2501 \cdot 10^{5}$	$2.2490 \cdot 10^{5}$	$1.5012 \cdot 10^{6}$
	3	No	$1.1037 \cdot 10^{8}$	$7.2074 \cdot 10^{7}$	$5.5183 \cdot 10^{8}$
	3	Yes	$7.4797 \cdot 10^{7}$	$4.8845 \cdot 10^{7}$	$3.7398 \cdot 10^{8}$
10	2	No	$1.1746 \cdot 10^{6}$	$8.1380 \cdot 10^{5}$	$5.4130 \cdot 10^{6}$
	3	No	$3.8175 \cdot 10^{8}$	$2.6098 \cdot 10^{8}$	$1.7985 \cdot 10^{9}$
	3	Yes	$2.5867 \cdot 10^{8}$	$1.7684 \cdot 10^{8}$	$1.2187 \cdot 10^{9}$

waiting time $E(T_m)$. Although the last column of Table 5 reveals that $E(T_m)$ is somewhat more sensitive to the choice of C when the expected value of the minus 1 hit variable $H_{\text{m}1}$ is small, it is still the case that $C = 2$ intervals $\{0,\dots,\delta_{\max}\}$ and $\{\delta_{\max}+1,\dots,W\}$ is sufficient for a neutral model, where all nonzero transition rates between array components will be of the same order or magnitude (so that clumping of states has a smaller effect), and double mutations can be disregarded.

In order to illustrate analytically that $C = 2$ and $C = 3$ intervals give very similar results, we consider a model with one single gene ($m = 1$) where a perfect match ($\delta_{\max} = 0$) to the single targeted binding site ($K_1 = 1$) is required. The distribution of the waiting time T_1 was derived in (86) for $C = 2$ distance-based intervals $\Delta_0 = \{0\}$ and $\Delta_1 = \{1,\dots,W\}$. In order to analyze the $C = 3$ model, recall from Table 5 that its distance intervals are given by $\Delta_0 = \{0\}, \Delta_1 = \{1\}$, and $\Delta_2 = \{2,\dots,W\}$. Let H_c refer to the number of substrings of the regulatory sequence whose number of mismatches to the targeted binding site belongs to Δ_c. From (63) and (A.10) we find that

$$\begin{array}{rcl} E(H_0) & = & L_0 4^{-W}, \\ E(H_1) = E(H_{\text{m}1}) & = & L_0 3W 4^{-W}, \\ E(H_2) & = & L_0\left(1 - 4^{-W} - 3W4^{-W}\right). \end{array} \tag{B.1}$$

Since $C = 3$, we have that the regulatory array component process $\mathbf{C}_t \in \mathcal{C} = \{0,1,2\}$. According to (66) and (68), the distribution of $\mathbf{C}_0$ is $(\kappa_0,\kappa_1,\kappa_2)$, where

$$\begin{array}{rcl} \kappa_0 & = & 1 - e^{-E(H_0)}, \\ \kappa_1 & = & e^{-E(H_0)}\left(1 - e^{-E(H_1)}\right), \\ \kappa_2 & = & e^{-(E(H_0)+E(H_1))}\left(1 - e^{-E(H_2)}\right). \end{array}$$

In order to find the distribution of T_1 we must also find the transition matrix

$$\mathbf{\Lambda}_{\text{n}} = \begin{pmatrix} -(\lambda_{10}+\lambda_{12}) & \lambda_{12} \\ \lambda_{21} & -\lambda_{21} \end{pmatrix}$$

among the non-absorbing states $\mathcal{C}_{\text{n}} = \{1,2\}$. For simplicity we will ignore stochastic tunneling. For a selectively neutral model ($s_0 = s_1 = s_2 = 1$), it follows from (73) and some computations that the three (single mutation) transition rates of this matrix are

$$\begin{array}{rcl} \lambda_{10} & = & \mu L_0 W 4^{-W}\left(1 - e^{-L_0 3W 4^{-W}}\right)^{-1}, \\ \lambda_{12} & = & \mu L_0 3W(W-1)4^{-W} e^{-L_0 3W 4^{-W}}\left(1 - e^{-L_0 3W 4^{-W}}\right)^{-1}, \\ \lambda_{21} & = & \mu L_0 3W(W-1)4^{-W}. \end{array}$$

The distribution of T_1 can now be retrieved from (57) as

$$F_{T_1}(t) = 1 - (\kappa_1,\kappa_2)\exp(\mathbf{\Lambda}_{\text{n}} t)\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

for all $t \geqslant 0$. After some quite long calculations, it can be shown that this formula simplifies to a simple mixture

$$T_1 \overset{\mathcal{L}}{\in} \kappa_0 \cdot \delta_0 + (1-\kappa_0)\text{Exp}(\lambda) \tag{B.2}$$

of a one point distribution δ_0 at 0, and an exponential distribution with rate parameter

$$\lambda = \frac{\lambda_{10}+\lambda_{12}+\lambda_{21}}{2} - \sqrt{\left(\frac{\lambda_{10}+\lambda_{12}+\lambda_{21}}{2}\right)^2 - \lambda_{10}\lambda_{21}},$$

where $-\lambda$ is the largest eigenvalue of $\mathbf{\Lambda}_{\text{n}}$. Our objective is to find how much the expected value of T_1 is impacted by using $C = 2$ or $C = 3$ intervals, and whether the expected value of the minus one hit variable $H_1 = H_{\text{m}1}$ influences the result. Comparing formulas (87) and (B.2), and using the fact that the probability κ_0 is the same for both values of C, it follows that

$$\frac{E(T_1)_{C=3}}{E(T_2)_{C=2}} = \frac{\mu L_0 W 4^{-W}}{\lambda} \approx \begin{cases} 1, & E(H_{\text{m}1}) \gg 1, \\ (3W-2)/(3W-3), & E(H_{\text{m}1}) \ll 1. \end{cases}$$

As mentioned above, this confirms the numerical values of Table 5, where it was found that $E(T_1)$ is essentially independent of C for large values of $E(H_{\text{m}1})$, and only slightly dependent on C for small values of $E(H_{\text{m}1})$.

We also notice some other features from Table 5. First, $E(T_2)/E(T_1)$ is larger for those scenarios that have a long waiting time. This is due to the fact that the expected waiting time obeys formula (9) when backward mutations are allowed ($\gamma = 1$), so that $E(T_{m+1})/E(T_m)$ is close to r, the ratio in (93) between the rates at which binding sites are lost and acquired. And this ratio tends to get larger the longer the waiting time is. Second, if $\delta_{\max} = 2$ and $W = 6$, with very high probability, the target is reached already at time $t = 0$, and therefore the order of $E(T_m)$ for this scenario is very small, between 10^{-11} and 10^{-10}.

B.1.1.2. Non-neutral models

Table 6 illustrates how the expected waiting time $E(T_m)$ is affected by the coarseness of the array decomposition (31). This is similar to Table 5, apart from the fact that the final state in Table 6 has a high fitness ($s_0 = \ldots = s_{m-1} = 1, s_m = 10$). In this case $E(T_1)$ and $E(T_2)$ are estimated well with $C = 2$ distance intervals when the expected value of the minus 1 hit variable $H_{\text{m}1}$ in the last column is large. On the other hand, when the expected number of minus 1 hits is moderately large or small, the model with $C = 2$ intervals underestimates $E(T_1)$ and $E(T_2)$ severely. In this case at least $C = 3$ intervals are needed to obtain good approximations of these two expected waiting times. Notice also that all models in Table 5 with $C \geqslant 3$ include stochastic tunneling (ST). Since the effect of double mutations is to *reduce* the waiting time, $E(T_1)$ and $E(T_2)$ would differ *even more* between the $C = 2$ and $C = 3$ scenarios if ST was not accounted for the latter models.

In order to explain the results of Table 6, it is instructive to compare the $C = 2$ and $C = 3$ scenarios for a model with one single gene ($m = 1$) and no allowed mismatches ($\delta_{\max} = 0$) between substrings of the regulatory sequence of length $W = 6$ and the targeted binding site. We have that

$$\left.\begin{array}{rcl} \kappa_0 & = & 0.7843, \\ \kappa_1 & = & 0.2157, \\ \lambda_{10} & = & 1.3118 \cdot 10^{-4}, \end{array}\right. \quad \text{when } C = 2,$$

whereas

$$\left.\begin{array}{rcl} \kappa_0 & = & 0.2157, \\ \kappa_1 & = & 0.7744, \\ \kappa_2 & = & 0.0099, \\ \lambda_{10} & = & 1.3285 \cdot 10^{-4}, \\ \lambda_{21} & = & 2.1863 \cdot 10^{-7}, \\ \lambda_{20} & = & 1.0372 \cdot 10^{-7}, \end{array}\right. \quad \text{when } C = 3.$$

When $\mathbf{C}_t$ has $C = 3$ states, the two transition rates from state 2 (the tunneling rate λ_{20} and the single mutation rate λ_{21}) are both much smaller than the single mutation rate λ_{10} from state 1. Therefore, the waiting time $T_1|\mathbf{C}_0 = 2$ will be orders of magnitude larger than $T_1|\mathbf{C}_0 = 1$. And although the probability $\kappa_2 = 0099$ is small that the system starts with at least two mismatches ($\mathbf{C}_0 = 2$), this will still impact $E(T_1)$, since the distribution of T_1 is highly skewed to the right. Notice for instance that the median $F_{T_1}^{-1}(0.5)$ of the waiting time is quite insensitive to the number of array components C when κ_2 is small. This can be seen from Table 7 when W equals 6 or 7. In

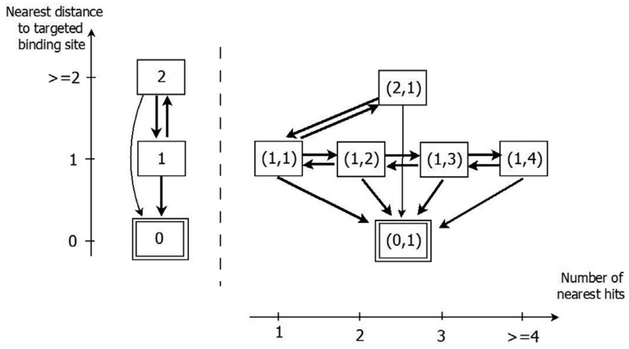


Fig. 8. Illustration of two array component decompositions for a system with $m = 1$ targeted gene. The states 0,1, and 2 of the coarser array decomposition (31) to the left correspond to $C = 3$ sets $\Delta_0 = \{0\}, \Delta_1 = \{1\}$ and $\Delta_2 = \{2, \ldots, W\}$ of distances to the nearest binding site. The states (c, n) of the finer decomposition (41) to the right includes information about distance $\delta \in \Delta_c$ to the nearest binding site, and number $H_c \in \mathcal{H}_{cn}$ of substrings of the regulatory sequence with such a distance, where $\mathcal{H}_{01} = \mathcal{H}_{21} = \{1, 2, \ldots, \}, \mathcal{H}_{11} = \{1\}, \mathcal{H}_{12} = \{2\}, \mathcal{H}_{13} = \{3\}$, and $\mathcal{H}_{14} = \{4, 5, \ldots\}$. Both array decompositions have one single absorbing state, 0 and $(0, 1)$ respectively, marked with double boxes. All single mutation transition rates $\lambda^{(1)}_{\boldsymbol{cd}}$ (thick arrows) and double mutation transition rates $\lambda^{(2)}_{\boldsymbol{cd}}$ (thin arrows) from non-absorbing states $\boldsymbol{c}$ are shown. These are averages of the actual transition rates of the consensus array process $\underline{\boldsymbol{X}}_t$. When the near hit variable H_1 is large, the Markov approximation of $\boldsymbol{C}_t$ works well for the smaller state space to the left, since there is little variation in the actual single mutation transition rates from state 1 into state 0 (and the double mutation rate from 2 to 0 can be ignored). However, when H_1 gets small, the finer decomposition to the right is needed in order to have a smaller variation of the actual transition rates within each state $(1, n)$ to the absorbing state $(0, 1)$. Depending on the model, the double mutation rate from $(2, 1)$ to $(0, 1)$ also becomes important, possibly requiring an even large state space based on $C = 4$ distance intervals. See Appendix C.4 for more details.

this table we also illustrate the impact of stochastic tunneling (ST). As mentioned above, when ST is not accounted for, the distribution of the waiting time differs even more between the $C = 2$ and $C = 3$ scenarios.

B.1.2. Array decomposition based on number of mismatches and hit variables

As a next step, we will investigate whether the distribution of the waiting time T_m changes a lot if instead of array decomposition (31), based on the number of mismatches to local targets, the finer decomposition (41) is used, that also includes hit variables. In order to illustrate this, we will consider a system with $m = 1$ gene, no allowed mismatch to the targeted binding sites ($\delta_{\max} = 0$) and no double mutations. We allow the selection coefficient s_1 for arrays that have reached the target, to vary, and consider an array decomposition based on the $C = 3$ distance intervals $\Delta_0 = \{0\}, \Delta_1 = \{1\}$, and $\Delta_2 = \{2, \ldots, W\}$. Since $m = 1$, we drop subscript j and assume there is $n_0 = n_2 = 1$ interval $\mathcal{H}_{01} = \mathcal{H}_{21} = \{1, \ldots\}$ for the number of hit variables of distance classes Δ_0 and Δ_2 respectively. To begin with, we allow for infinitely many $(n_1 = \infty)$ intervals $\mathcal{H}_{1n} = \{n\}, n = 1, 2, \ldots$ for the near hit variable H_1. The elements of the infinitely large state space of $\boldsymbol{C}_t$ are enumerated as

$$\mathcal{C}_{\text{mismhit}} = \{(0, 1), (1, 1), (1, 2), \ldots, (2, 1)\}, \tag{B.3}$$

where $(0, 1)$ is absorbing with type $g(0, 1) = 1$, and the other states $\boldsymbol{c}$ are non-absorbing with type $g(\boldsymbol{c}) = 0$. We will assume that the hit variables H_0, H_1, and H_2 corresponding to Δ_0, Δ_1, and Δ_2 are independent Poisson variables with expected values given by (B.1). Combining this equation with (69), we obtain the marginal distribution

$$\kappa_{\boldsymbol{c}} = \begin{cases} 1 - e^{-L_0 4^{-W}}, & \boldsymbol{c} = (0, 1), \\ e^{-L_0 4^{-W}} \left(3L_0 W 4^{-W}\right)^n e^{-3L_0 W 4^{-W}}, & \boldsymbol{c} = (1, n), \\ e^{-L_0 4^{-W}} e^{-3L_0 W 4^{-W}} \left(1 - e^{-L_0\left(1 - 4^{-W}(1+3W)\right)}\right), & \\ \quad \approx e^{-L_0(1+3W)4^{-W}}, & \boldsymbol{c} = (2, 1) \end{cases} \tag{B.4}$$

of $\boldsymbol{C}_t$, for a selectively neutral model ($s_1 = 1$). In order to derive the transition rates of $\boldsymbol{C}_t$, we first use (A.11) and compute the transition rates

$$\begin{aligned} \theta^{(1)}_{01} &= W, \\ \theta^{(1)}_{10} &= 1/3, \\ \theta^{(1)}_{12} &= W - 1, \\ \theta^{(1)}_{21} &= 3W(W-1)/\left(4^W - 1 - 3W\right) \end{aligned} \tag{B.5}$$

for words of length W. Inserting (B.5) into (74)-(75), we find that the transition rates of $\boldsymbol{C}_t$, due to single mutations, are

Table 8
Expected values as well as 0.05 and 0.95 quantiles of T_1, the waiting time until the targeted binding site $\boldsymbol{b}_1$ of length W for $m = 1$ gene is reached, with s_1 the selection coefficient of individuals that have acquired $\boldsymbol{b}_1$. The model allows for no mismatches ($\delta_{\max} = 0$), and it has $C = 3$ intervals $\Delta_0 = \{0\}, \Delta_1 = \{1\}$, and $\Delta_2 = \{2, \ldots, W - 1\}$ that correspond to the number mismatches between $\boldsymbol{b}_1$ and the closest substrings of the regulatory sequence. The model to the left corresponds to the mismatch-based array decomposition in (31), so that $n_0 = n_1 = n_2$. The model to the right corresponds to the finer array decomposition (41), with $n_0 = n_2 = 1$, whereas n_1 varies, with $n_1 = 10$ for $W = 6, 7, 8, n_1 = 9$ for $W = 9$ and $n_1 = 10$ for $W = 10$. This corresponds to using $\mathcal{H}_{1n} = \{n\}$ for $n = 1, \ldots, n_1 - 1$ and $\mathcal{H}_{1,n_1} = \{n_1, \ldots\}$, as illustrated in Fig. 8 for $n_1 = 4$. Finally, $H_{\text{m}1}$ is the number substrings of the regulatory sequence that only differ from $\boldsymbol{b}_1$ at one position. There is no stochastic tunneling, but the other parameters are the same as in Table 5.

		$n_0 = 1, n_1 = 1, n_2 = 1$			$n_0 = n_2 = 1, n_1$ varies			
s_1	W	$F^{-1}_{T_1}(0.05)$	$E(T_1)$	$F^{-1}_{T_1}(0.95)$	$F^{-1}_{T_1}(0.05)$	$E(T_1)$	$F^{-1}_{T_1}(0.95)$	$E(H_{\text{m}1})$
1	6	0	$5.3858 \cdot 10^7$	$1.8904 \cdot 10^8$	0	$5.6607 \cdot 10^7$	$2.0013 \cdot 10^8$	4.37
	7	0	$2.2505 \cdot 10^8$	$7.0214 \cdot 10^8$	0	$2.3315 \cdot 10^8$	$7.2864 \cdot 10^8$	1.27
	8	$3.0452 \cdot 10^7$	$8.3947 \cdot 10^8$	$2.5411 \cdot 10^9$	$3.0655 \cdot 10^7$	$8.5063 \cdot 10^8$	$2.5753 \cdot 10^9$	0.36
	9	$1.4433 \cdot 10^8$	$3.0352 \cdot 10^9$	$9.1163 \cdot 10^9$	$1.4481 \cdot 10^8$	$3.0465 \cdot 10^9$	$9.1505 \cdot 10^9$	0.10
	10	$5.5155 \cdot 10^8$	$1.0952 \cdot 10^{10}$	$3.2829 \cdot 10^{10}$	$5.5207 \cdot 10^8$	$1.0962 \cdot 10^{10}$	$3.2861 \cdot 10^{10}$	0.03
10	6	0	$5.1250 \cdot 10^4$	$2.2282 \cdot 10^4$	0	$5.3737 \cdot 10^4$	$3.3087 \cdot 10^4$	4.37
	7	0	$3.4681 \cdot 10^6$	$2.1779 \cdot 10^7$	0	$3.4840 \cdot 10^6$	$2.1839 \cdot 10^7$	1.27
	8	$3.4592 \cdot 10^3$	$2.6988 \cdot 10^7$	$1.0303 \cdot 10^8$	$3.4893 \cdot 10^3$	$2.7015 \cdot 10^7$	$1.0311 \cdot 10^8$	0.36
	9	$2.0577 \cdot 10^4$	$1.1037 \cdot 10^8$	$3.5445 \cdot 10^8$	$2.0844 \cdot 10^4$	$1.1040 \cdot 10^8$	$3.5455 \cdot 10^8$	0.10
	10	$8.7105 \cdot 10^6$	$3.8175 \cdot 10^8$	$1.1660 \cdot 10^9$	$8.7122 \cdot 10^6$	$3.8178 \cdot 10^8$	$1.1661 \cdot 10^9$	0.03

$$\lambda^{(1)}_{\boldsymbol{cd}} = \begin{cases} N\mu\beta(s_1^{-1})\theta^{(1)}_{01}\frac{P(H_0=1)P(H_1=n-1)}{P(H_0>0)} & \\ \quad = N\mu\beta(s_1^{-1})\frac{\left(L_0 3W4^{-W}\right)^n e^{-L_0(1+3W)4^{-W}}}{3(n-1)!\left(1-e^{-L_0 4^{-W}}\right)}, & \boldsymbol{c}=(0,1),\boldsymbol{d}=(1,n), \\ N\mu\beta(s_1)\theta^{(1)}_{10}n = \frac{1}{3}N\mu\beta(s_1)n, & \boldsymbol{c}=(1,n),\boldsymbol{d}=(0,1), \\ \mu\theta^{(1)}_{21}E(H_2)/P(H_2>0) & \\ \quad = \mu 3L_0W(W-1)4^{-W}, & \boldsymbol{c}=(1,n),\boldsymbol{d}=(1,n+1), \\ \mu\theta^{(1)}_{12}n = \mu n(W-1), & \boldsymbol{c}=(1,n),\boldsymbol{d}=(1,n-1), n \geqslant 2, \\ \mu\theta^{(1)}_{12} = \mu(W-1), & \boldsymbol{c}=(1,1),\boldsymbol{d}=(2,1), \\ \mu\theta^{(1)}_{21}\frac{E(H_2)}{P(H_2>0)} = \mu 3L_0W(W-1)4^{-W}, & \boldsymbol{c}=(2,1),\boldsymbol{d}=(1,1). \end{cases} \tag{B.6}$$

In practice, we need to reduce the infinitely large state space (B.3). This is achieved by choosing a truncation point n_1, and then clumping all states $(1,n_1),(1,n_1+1),\ldots$ into one new state $(1,n_1)$, as illustrated in Fig. 8 when $n_1 = 4$. It is possible to adjust the marginal distribution (B.4) and the transition rates (B.6) of $\boldsymbol{C}_t$, to account for this clumping of states. Then we apply the phase-type distribution formula (56) for T_m, which has a marginal distribution $\boldsymbol{\kappa}_{\rm n} = (\kappa_{11},\kappa_{12},\ldots,\kappa_{1n_1},\kappa_{21})$ and intensity matrix $\boldsymbol{\Lambda}_{\rm n} = \left(\lambda^{(1)}_{\boldsymbol{cd}}\right)_{\boldsymbol{c},\boldsymbol{d}\in\mathcal{C}_{\rm n}}$, where $\mathcal{C}_{\rm n}$ includes all states except the absorbing state $(0,1)$.

Table 8 gives quantiles and expected values of the waiting time T_1 for two models, where no hit variables are accounted for in the regulatory sequence decomposition of the first model ($n_1 = 1$). For the second model, the truncation point n_1 is chosen much larger than the expected value $E(H_1) = E(H_{\rm m1})$ of the number of substring hits that only miss the target at one letter. This essentially means that the second model of Table 8 fully accounts for the value of H_1, as in (B.3).

It can be seen from Table 8 that the distribution of T_1 is somewhat different for the two models when gets $E(H_1)$ large. This is due to the fact that the transition rate $\lambda^{(1)}_{(1,n),(0,1)} = N\mu n\beta(s_1)/3$, from $(1,n)$ to the absorbing state $(0,1)$, varies with n. Therefore, some accuracy is lost by clumping states $(1,n)$ such that $P(H_1 = n)$ is non-negligible. On the other hand, when $E(H_1)$ gets small, the two models of Table 8 predict essentially the same waiting time distribution, and then there is no need to employ the finer partition (B.3) of regulatory sequences. However, when $E(H_1)$ is very small, it might be necessary to split $(2,1)$ into more states, since the probability $\kappa_{(2,1)}$ of $(2,1)$ is close to 1. This is true in particular if stochastic tunneling is taken into account.

When s_1 is large it is possible to extend (8) and approximate the phase-type distribution formula (56), for the distribution of the waiting time T_1, by the mixture distribution

$$\begin{aligned} T_1 \overset{\mathcal{L}}{\in} \quad & \kappa_{(0,1)}\delta_0 + \sum_{n=1}^{\infty}\kappa_{(1,n)}\text{Exp}\left(\lambda^{(1)}_{(1,n),(0,1)}\right) \\ + \quad & \kappa_{(2,1)}\text{Exp}\left(\lambda^{(1)}_{(2,1),(1,1)}\right) * \text{Exp}\left(\lambda^{(1)}_{(1,1),(0,1)}\right), \end{aligned} \tag{B.7}$$

where $*$ refers to convolution of two distributions. The rationale for (B.7) is that for big s_1 the rates $\lambda^{(1)}_{(1,n),(0,1)}$ from the non-absorbing states $(1,n)$ to the absorbing state $(0,1)$ are much larger than all rates between the non-absorbing states. Therefore, if the system starts in $(1,n)$, all transitions to other non-absorbing states can be ignored. Instead we await the transition to $(0,1)$, which happens after an $\text{Exp}\left(\lambda^{(1)}_{(1,n),(0,1)}\right)$-distributed time. Likewise, if the population starts in $(2,1)$ there will first be a transition to $(1,1)$, and after that a second transition to $(0,1)$. The total time for this to happen is $\text{Exp}\left(\lambda^{(1)}_{(2,1),(1,1)}\right) * \text{Exp}\left(\lambda^{(1)}_{(1,1),(0,1)}\right)$-distributed. When $s_1 = 10$, formula (B.7) predicts that the expected waiting times equal $E(T_1) = 5.3338\cdot 10^4, 3.4683\cdot 10^6, 2.6929\cdot 10^7, 1.1007\cdot 10^8,$ and $3.8061\cdot 10^8$, when the word length $W = 6,7,8,9,10$ respectively. This is very close to the corresponding expected waiting times, based on (59), in the lower right part of Table 8.

Table 9
Values of the expected waiting time $E(T_1)$ until one of K possible binding sites $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k$ of length W appears, within the regulatory sequence of the enhancer/promoter region of $m = 1$ gene, with maximal mismatch $\delta_{\max}$ between a targeted binding site and some substring of the regulatory sequence. All other parameters are given by Table 3. When $K > 1$, the targeted binding sites $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_K$ either differ in 1 position or in W positions.

		$K = 1$	$K = 2$		$K = 3$	
W	$\delta_{\max}$		$\lvert\boldsymbol{b}_1 - \boldsymbol{b}_2\rvert = 1$	$\lvert\boldsymbol{b}_1 - \boldsymbol{b}_2\rvert = W$	$\lvert\boldsymbol{b}_{k_1} - \boldsymbol{b}_{k_2}\rvert = 1$	$\lvert\boldsymbol{b}_{k_1} - \boldsymbol{b}_{k_2}\rvert = W$
6	0	$5.38\cdot 10^7$	$2.23\cdot 10^7$	$2.11\cdot 10^7$	$1.24\cdot 10^7$	$1.10\cdot 10^7$
6	1	$4.53\cdot 10^4$	$7.10\cdot 10^2$	$2.24\cdot 10^2$	$1.39\cdot 10^1$	$1.48\cdot 10^0$
7	0	$2.22\cdot 10^8$	$1.10\cdot 10^8$	$1.04\cdot 10^8$	$7.23\cdot 10^7$	$6.54\cdot 10^7$
7	1	$3.44\cdot 10^6$	$6.74\cdot 10^5$	$4.53\cdot 10^5$	$1.66\cdot 10^5$	$7.95\cdot 10^4$

Table 10
This table displays the upper bound $K_{\rm lim}$, defined in (B.8), for the number K of binding sites of length W that can be dispersed at pairwise distances larger than δ_{C-1}. If K is a positive integer smaller than $K_{\rm lim}$, and the targeted binding sites of genes $j = 1,\ldots,m$ are positioned in this way, the simpler algorithm of Appendix A.3 can be used for computing $\Pi_{jc}, \theta^{(1)}_{jc,c-1}$, and $\theta^{(2)}_{jc,c-2}$, quantities needed for deriving the marginal distribution and intensity matrix of the array component process $\boldsymbol{C}_t$.

	δ_{C-1}					
W	1	2	3	4	5	6
6	215.58	26.60	5.90	2.15	1.22	1.00
7	744.73	77.65	14.17	4.11	1.80	1.15
8	$2.62\cdot 10^3$	236.59	36.63	8.79	3.11	1.58
9	$9.36\cdot 10^3$	744.43	100.06	20.44	6.03	2.50
10	$3.38\cdot 10^4$	$2.40\cdot 10^3$	285.25	50.69	12.80	4.46
12	$4.53\cdot 10^5$	$2.66\cdot 10^4$	$2.55\cdot 10^3$	359.52	70.16	18.38
15	$2.33\cdot 10^7$	$1.08\cdot 10^6$	$8.09\cdot 10^4$	$8.67\cdot 10^3$	$1.26\cdot 10^3$	238.49
20	$1.80\cdot 10^{10}$	$6.21\cdot 10^8$	$3.38\cdot 10^7$	$2.59\cdot 10^6$	$2.62\cdot 10^5$	$3.39\cdot 10^4$

B.2. Varying number of binding site targets per gene

In this subsection we investigate how $E(T_1)$, the expected waiting time for $m = 1$ gene, depends on the number K of binding site targets, and also on how closely spaced these binding sites $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_K$ are. It is not surprising that $E(T_1)$ decreases the more binding sites there are (see Table 9), but $E(T_1)$ also gets smaller the more widely spaced these targets are. This effect is more pronounced when a mismatch between the regulatory sequence and the targeted binding site is allowed ($\delta_{\max} > 0$), in particular when the length W of the binding sites is small.

When K gets large, it is however intractable to find the distribution of T_1 for any combination of binding sites $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_K$. We recall from Appendix A that when K is large, many operations are required to compute certain quantities ($\Pi_{jc}, \theta^{(1)}_{jc,c-1}$, and $\theta^{(2)}_{jc,c-2}$) needed in order to find the marginal distribution $\boldsymbol{\kappa}$ and the intensity matrix $\boldsymbol{\Lambda}$ of the array component process $\boldsymbol{C}_t$. When W is sufficiently large and/or when the left end point δ_{C-1} of the interval Δ_{C-1} in (A.6) is sufficiently small, it is possible though for the targeted binding sites $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_K$ to be so widely spread that the simple algorithm of Appendix A.3 can be used. Eq. (A.17) gives a sufficient condition for such a positioning of the K binding sites. It corresponds to an upper limit

$$K_{\text{lim}} = \frac{4^W}{\sum_{\delta=0}^{\delta_{C-1}} \binom{W}{\delta} 3^\delta} \tag{B.8}$$

of K, as a function of W and δ_{C-1}. Table 10 displays values of K_{lim} for different combinations of W and δ_{C-1}. In order for the simple algorithm of Appendix A.3 to be used, it is necessary that K is an integer smaller than K_{lim}. In particular, in Table 9 we used $C = 2$ intervals $\Delta_0 = \{0, \ldots, \delta_{\max}\}$ and $\Delta_1 = \{\delta_{\max} + 1, \ldots, W\}$, so that in this case $\delta_{C-1} = \delta_{\max} + 1$. Therefore, the simple algorithm of Appendix A.3 is exact for all entries of Table 9 where the targeted binding sites are maximally dispersed, at distance W.

B.3. Varying selection coefficients, order of target appearance, and absence or presence of stochastic tunneling

In this subsection we consider a model with $m = 2$ genes, and investigate how the expected waiting time $E(T_2)$ until the targets of both genes have been fixed, depends on the selection coefficients s_1 and s_2 for individuals with 1 and 2 targets in place, where $s_1 \leqslant 1$ is either deleterious or neutral, whereas $s_2 \geqslant 1$ is either advantageous or neutral. We also vary the order of target appearance (TA) and whether stochastic tunneling (ST) is accounted for or not.

Tables 11 and 12 list values of $E(T_2)$ for a distance-based decomposition (31) of regulatory arrays based on $C = 3$ intervals. It is evident from these two tables that $E(T_2)$ increases dramatically when s_1 decreases, for one of the target appearance schemes (TA = arbitrary), but not for the other (TA = fixed). Although $E(T_2)$ is somewhat smaller for an array decomposition with $C = 2$ intervals (not shown), the behavior its qualitatively the same as in Tables 11 and 12, and easier to explain conceptually. Fig. 9 illustrates the transition rates between the array components for a system with $C = 2$. This figure explains the different behavior of TA = fixed and TA = arbitrary, since one of the states $(1, 0)$ of the array component process $\boldsymbol{C}_t$ will have different selection coefficients under these two schemes. Indeed, if the correct binding site has appeared along the enhancer/promoter region of the second but not of the first gene, the regulatory array is of type 1 ($g(1, 0) = 1$) when TA = arbitrary, but of type 0 when TA = fixed ($g(1, 0) = 0$). Consequently, when TA = fixed there is a route $(1, 1) \to (1, 0) \to (0, 0)$ with single mutations towards the final target with no deleterious states, whereas no such path exists for TA = arbitrary.

We also notice from Table 11 that $E(T_2)$ is not a monotone decreasing function of s_1 when $s_0 = s_2 = 1$ and TA = fixed. This behavior may seem surprising at first, but it can also be explained by considering the simpler system in Fig. 9 with $C = 2$. Since the forward rate $\lambda_{(1,1),(0,1)}$ decreases dramatically when s_1 decreases, and the

Table 11
Expected waiting times $E(T_2)$ until both targets have been fixed, for a model with $m = 2$ genes, and different choices of W, the length of the targeted binding site. The selection coefficient with g targets in place is s_g, with $s_0 = s_2 = 1$, whereas s_1 varies. The order of target appearance (TA) is either fixed or arbitrary, and stochastic tunneling (ST) is either accounted for or not. There are $C = 3$ distance-based intervals $\Delta_0 = \{0\}, \Delta_1 = \{1\}$, and $\Delta_2 = \{2, \ldots, W\}$. All other parameters are given by Table 3. The corresponding values of $E(T_2)$ with $C = 2$ intervals $\Delta_0 = \{0\}$ and $\Delta_1 = \{1, \ldots, W\}$ (not shown) are very close. The values marked $*$ are numerically instable.

		$s_2 = 1$			
		TA = fixed		TA = arbitrary	
W	s_1	No ST	ST	No ST	ST
6	0.99700	$3.4340 \cdot 10^8$	$3.4287 \cdot 10^8$	$5.2173 \cdot 10^{19} *$	$6.2374 \cdot 10^{12}$
	0.99900	$3.4341 \cdot 10^8$	$3.4286 \cdot 10^8$	$3.1039 \cdot 10^{11}$	$2.7003 \cdot 10^{11}$
	0.99970	$3.0043 \cdot 10^8$	$2.9999 \cdot 10^8$	$9.1314 \cdot 10^8$	$9.1199 \cdot 10^8$
	0.99990	$2.3359 \cdot 10^8$	$2.3328 \cdot 10^8$	$2.8228 \cdot 10^8$	$2.8201 \cdot 10^8$
	0.99997	$2.1225 \cdot 10^8$	$2.1195 \cdot 10^8$	$2.2003 \cdot 10^8$	$2.1975 \cdot 10^8$
	1.00000	$2.0563 \cdot 10^8$	$2.0532 \cdot 10^8$	$2.0563 \cdot 10^8$	$2.0532 \cdot 10^8$
8	0.99700	$5.6619 \cdot 10^{10}$	$5.6447 \cdot 10^{10}$	$1.0341 \cdot 10^{22} *$	$1.0233 \cdot 10^{15}$
	0.99900	$5.6594 \cdot 10^{10}$	$5.6419 \cdot 10^{10}$	$6.2499 \cdot 10^{13}$	$5.2928 \cdot 10^{13}$
	0.99970	$4.9031 \cdot 10^{10}$	$4.8908 \cdot 10^{10}$	$1.7923 \cdot 10^{11}$	$1.7937 \cdot 10^{11}$
	0.99990	$3.6050 \cdot 10^{10}$	$3.5979 \cdot 10^{10}$	$4.8648 \cdot 10^{10}$	$4.8638 \cdot 10^{10}$
	0.99997	$3.0850 \cdot 10^{10}$	$3.0771 \cdot 10^{10}$	$3.3307 \cdot 10^{10}$	$3.3240 \cdot 10^{10}$
	1.00000	$2.8746 \cdot 10^{10}$	$2.8659 \cdot 10^{10}$	$2.8746 \cdot 10^{10}$	$2.8659 \cdot 10^{10}$
10	0.99700	$1.1406 \cdot 10^{13}$	$1.1357 \cdot 10^{13}$	$1.0995 \cdot 10^{24} *$	$1.6844 \cdot 10^{17}$
	0.99900	$1.1401 \cdot 10^{13}$	$1.1351 \cdot 10^{13}$	$1.2758 \cdot 10^{16}$	$1.0428 \cdot 10^{16}$
	0.99970	$9.8691 \cdot 10^{12}$	$9.8342 \cdot 10^{12}$	$3.6552 \cdot 10^{13}$	$3.6602 \cdot 10^{13}$
	0.99990	$7.2233 \cdot 10^{12}$	$7.2025 \cdot 10^{12}$	$9.8334 \cdot 10^{12}$	$9.8294 \cdot 10^{12}$
	0.99997	$6.1494 \cdot 10^{12}$	$6.1268 \cdot 10^{12}$	$6.6629 \cdot 10^{12}$	$6.6432 \cdot 10^{12}$
	1.00000	$5.7096 \cdot 10^{12}$	$5.6850 \cdot 10^{12}$	$5.7096 \cdot 10^{12}$	$5.6850 \cdot 10^{12}$

Table 12
Expected waiting times $E(T_2)$ until both targets have been fixed, for a model with $m = 2$ genes and different choices of W, the length of the targeted binding site. The selection coefficient with g targets in place is s_g, with $s_0 = 1, s_2 = 10$, and s_1 varying. The order of target appearance (TA) is either fixed or arbitrary, and stochastic tunneling (ST) is either accounted for or not. There are $C = 3$ distance-based intervals $\Delta_0 = \{0\}, \Delta_1 = \{1\}$, and $\Delta_2 = \{2, \ldots, W\}$. All other parameters are given by Table 3. The corresponding values of $E(T_2)$ with $C = 2$ intervals $\Delta_0 = \{0\}$ and $\Delta_1 = \{1, \ldots, W\}$ (not shown) are very close for $W = 6$, but smaller for $W = 8$ and $W = 10$. The values marked $*$ are numerically instable.

		$s_2 = 10$			
		TA = fixed		TA = arbitrary	
W	s_1	No ST	ST	No ST	ST
6	0.99700	$4.2719 \cdot 10^7$	$1.4749 \cdot 10^7$	$8.1242 \cdot 10^{18}$	$4.9028 \cdot 10^8$
	0.99900	$4.2584 \cdot 10^7$	$1.4278 \cdot 10^7$	$4.7613 \cdot 10^{10}$	$1.6212 \cdot 10^8$
	0.99970	$3.6782 \cdot 10^7$	$1.2405 \cdot 10^7$	$1.3577 \cdot 10^8$	$3.6683 \cdot 10^7$
	0.99990	$2.6902 \cdot 10^7$	$9.6193 \cdot 10^6$	$3.6574 \cdot 10^7$	$1.3487 \cdot 10^7$
	0.99997	$2.2910 \cdot 10^7$	$8.1313 \cdot 10^6$	$2.4818 \cdot 10^7$	$8.9054 \cdot 10^6$
	1.00000	$2.1278 \cdot 10^7$	$7.4818 \cdot 10^6$	$2.1278 \cdot 10^7$	$7.4818 \cdot 10^6$
8	0.99700	$1.7823 \cdot 10^9$	$1.2497 \cdot 10^9$	$5.1915 \cdot 10^{20}*$	$1.1287 \cdot 10^{11}$
	0.99900	$1.7811 \cdot 10^9$	$1.2364 \cdot 10^9$	$2.8329 \cdot 10^{12}$	$3.7162 \cdot 10^{10}$
	0.99970	$1.5932 \cdot 10^9$	$1.1051 \cdot 10^9$	$7.1602 \cdot 10^9$	$4.2936 \cdot 10^9$
	0.99990	$1.1893 \cdot 10^9$	$8.4253 \cdot 10^8$	$1.7227 \cdot 10^9$	$1.2160 \cdot 10^9$
	0.99997	$9.9627 \cdot 10^8$	$7.0770 \cdot 10^8$	$1.0978 \cdot 10^9$	$7.7926 \cdot 10^8$
	1.00000	$9.1243 \cdot 10^8$	$6.4869 \cdot 10^8$	$9.1243 \cdot 10^8$	$6.4869 \cdot 10^8$
10	0.99700	$2.0657 \cdot 10^{11}$	$1.4496 \cdot 10^{11}$	$7.2839 \cdot 10^{22}*$	$1.8728 \cdot 10^{13}$
	0.99900	$2.0652 \cdot 10^{11}$	$1.4389 \cdot 10^{11}$	$3.9820 \cdot 10^{14}$	$6.1516 \cdot 10^{12}$
	0.99970	$1.8701 \cdot 10^{11}$	$1.3046 \cdot 10^{11}$	$9.5867 \cdot 10^{11}$	$6.0533 \cdot 10^{11}$
	0.99990	$1.4029 \cdot 10^{11}$	$9.9463 \cdot 10^{10}$	$2.1445 \cdot 10^{11}$	$1.5306 \cdot 10^{11}$
	0.99997	$1.1566 \cdot 10^{11}$	$8.1934 \cdot 10^{10}$	$1.2959 \cdot 10^{11}$	$9.1994 \cdot 10^{10}$
	1.00000	$1.0466 \cdot 10^{11}$	$7.3885 \cdot 10^{10}$	$1.0446 \cdot 10^{11}$	$7.3885 \cdot 10^{10}$

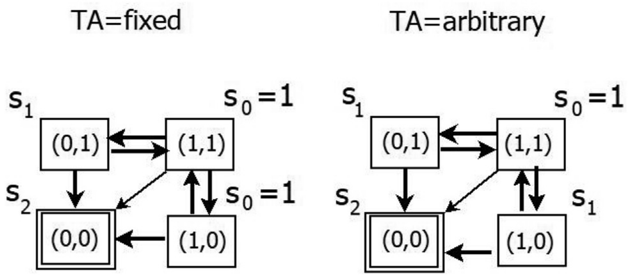


Fig. 9. A list of states (49) for the array component process $\boldsymbol{C}_t$, and selection coefficients, for a system with $m = 2$ genes, when the regulatory sequences of each gene are divided into $C = 2$ components, as in Section 6.2. The order of target appearance (TA) is either fixed or arbitrary, and the absorbing state $(0,0)$ is marked with double boxes. All single (double) mutation transitions from non-absorbing states are marked with thick (thin) arrows.

backward rate $\lambda_{(0,1),(1,1)}$ gets larger the smaller s_1 is. Both of these two transitions make it harder to reach the target when s_1 gets smaller, but there is also another forward rate $\lambda_{(0,1),(0,0)}$ which gets larger the smaller s_1 is. The expected waiting time $E(T_2)$ is therefore a complicated function of s_1, since it depends on s_1 through all these three rates, and the impact of the third rate is not negligible. Indeed, we find from (68) and (82) that there is a non-negligible probability $0.7843 \cdot 0.2157 = 0.169$ that the array components process starts at $\boldsymbol{C}_0 = (0,1)$, and from this state there is direct path $(0,1) \to (0,0)$ with rate $\lambda_{(0,1),(0,0)}$ towards the final target.

It is also evident from Tables 11 and 12 that ST makes $E(T_2)$ a lot smaller when $s_1 < 1$ and TA = arbitrary. For the other target appearance scheme (TA = fixed), ST has a small effect on $E(T_2)$ when $s_1 < 1$ and $s_2 = 1$, and a somewhat larger impact on $E(T_2)$ when $s_1 < 1$ and s_2 is large.

B.4. Accounting for overdispersion of number of binding site hits

In this subsection we will analyze the expected waiting time $E(T_1)$ until $K = 1$ targeted binding site $\boldsymbol{b}_1$ of length $W = 6$ or $W = 7$ is reached within one single regulatory sequence $(m = 1)$. Recall from (79) that the regulatory sequence component process $\boldsymbol{C}_t$ has two states 0 and 1, of which 0 is absorbing and 1 is non-absorbing. Moreover, since $\delta_{\max} = 0$, the two sets $\mathcal{B}_0 = \{\boldsymbol{b}_1\}$ and $\mathcal{B}_1 = \{\boldsymbol{b}; \boldsymbol{b} \neq \boldsymbol{b}_1\}$ in (80) represent the set of words of length W along the regulatory sequence that correspond to a reached or missed target. We will drop the assumption that the number H_0 and H_1 of subsequences of the consensus regulatory sequence $\boldsymbol{X}_0$ at time 0 that belong to $\mathcal{B}_0$ and $\mathcal{B}_1$, are Poisson distributed, and allow for overdispersion, as described in formulas (65)-(67) and (A.2)-(A.3).

It is well known that the expected waiting time $E(T_1)$ will depend on how many self repeats the targeted binding site $\boldsymbol{b}_1$ has (Durrett and Schmidt, 2007; Behrens and Vingron, 2010). A sequence with identical letters, like $\boldsymbol{b}_1 = (A, A, A, A, A, A)$ has a maximal number of self repeats, with a cycle length of $W_0 = 1$. This implies that it is possible for any two overlapping subsequences of $\boldsymbol{X}_t$ of length $W = 6$ (corresponding to a lag of $\eta = 1, \ldots, 5$) to hit $\boldsymbol{b}_1$. On the other hand, the sequence $\boldsymbol{b}_1 = (A, C, C, C, C, C)$ does not allow for any overlaps at all, so that any two subsequences of $\boldsymbol{X}_t$ of length $W = 6$ that hit $\boldsymbol{b}_1$ must be disjoint (corresponding to a lag of $\eta \geqslant 6$). The more self repeats $\boldsymbol{b}_1$ has, the more overdispersion there is, that is, the larger is the variance of H_0 compared to its expected value.

In Table 13 we computed a number of quantities when varying the targeted binding site $\boldsymbol{b}_1$ of length $W = 6$ or $W = 7$. We notice from the fourth and fifth columns of this table that for most choices of $\boldsymbol{b}_1$ there is only a small amount over over- or underdispersion, with values of $\mathrm{Var}(H_0)/E(H_0)$ very close to 1. The only exception occurs when all letters of $\boldsymbol{b}_1$ are identical, corresponding to a cycle length of $W_0 = 1$ for the repeats. In order to get an explicit upper bound of $\mathrm{Var}(H_0)/E(H_0)$, we will use the fact that $\mathcal{B}_0$ contains one single word, which in view of (A.2)-(A.3) implies that for any choice of $\boldsymbol{b}_1$ we have that $|\mathcal{B}_0|_\eta \leqslant 1$ and $\Pi_{0\eta} \leqslant 1/4^{W+\eta}$ for $\eta = 1, \ldots, W-1$, with equality if η is divisible by W_0. Then we use (65) and (81) to deduce that

Table 13
Values of the expected waiting time $E(T_1)$ until one single binding site $\boldsymbol{b}_1$ of length $W = 6$ or $W = 7$ has been reached within the regulatory sequence of $m = 1$ gene. When $\boldsymbol{b}_1$ has self repeats, the cycle length W_0 of these repeats is given. All other parameters of the model are given in Table 3. Apart from $E(T_1)$, we also display κ_0, the probability that this binding site is present at time $t = 0$ already, as well as the transition rate λ_{10} from the non-absorbing state to the absorbing state of the regulatory sequence component process $\boldsymbol{C}_t$. The variable H_0 defines the number of times a word of length W within the regulatory sequence matches $\boldsymbol{b}_1$. O/U refers to whether overdispersion O ($\mathrm{Var}(H_0) > E(H_0)$) or underdispersion U ($\mathrm{Var}(H_0) < E(H_0)$) occurs, or whether this is not accounted for (NAF). In the latter two cases, it is assumed that $E(H_0) = \mathrm{Var}(H_0)$ when calculating κ_0, λ_{10}, and $E(T_1)$.

W	$\boldsymbol{b}_1$	W_0	$E(H_0)$	$\mathrm{Var}(H_0)$	O/U	κ_0	λ_{10}	$E(T_1)$
6	(A,A,A,A,A,A)	1	0.2429	0.4038	O	0.1701	$1.458 \cdot 10^{-8}$	$5.694 \cdot 10^{7}$
	(A,C,A,C,A,C)	2	0.2429	0.2745	O	0.2042	$1.458 \cdot 10^{-8}$	$5.460 \cdot 10^{7}$
	(A,C,G,A,C,G)	3	0.2429	0.2498	O	0.2130	$1.458 \cdot 10^{-8}$	$5.400 \cdot 10^{7}$
	(A,C,G,T,A,C)	4	0.2429	0.2442	O	0.2152	$1.458 \cdot 10^{-8}$	$5.385 \cdot 10^{7}$
	(A,C,G,T,C,A)	5	0.2429	0.2427	U	0.2157	$1.458 \cdot 10^{-8}$	$5.381 \cdot 10^{7}$
	(A,C,G,T,C,C)	–	0.2429	0.2423	U	0.2157	$1.458 \cdot 10^{-8}$	$5.381 \cdot 10^{7}$
	any sequence	–	0.2429	0.2429	NAF	0.2157	$1.458 \cdot 10^{-8}$	$5.381 \cdot 10^{7}$
7	(A,A,A,A,A,A,A)	1	0.0607	0.1010	O	0.0454	$4.247 \cdot 10^{-9}$	$2.248 \cdot 10^{8}$
	(A,C,A,C,A,C,A)	2	0.0607	0.0687	O	0.0554	$4.247 \cdot 10^{-9}$	$2.224 \cdot 10^{8}$
	(A,C,G,A,C,G,A)	3	0.0607	0.0625	O	0.0580	$4.247 \cdot 10^{-9}$	$2.218 \cdot 10^{8}$
	(A,C,G,T,A,C,G)	4	0.0607	0.0611	O	0.0587	$4.247 \cdot 10^{-9}$	$2.217 \cdot 10^{8}$
	(A,C,G,T,C,A,C)	5	0.0607	0.0607	O	0.0588	$4.247 \cdot 10^{-9}$	$2.216 \cdot 10^{8}$
	(A,C,G,T,C,C,A)	6	0.0607	0.0607	U	0.0589	$4.247 \cdot 10^{-9}$	$2.216 \cdot 10^{8}$
	(A,C,G,T,C,C,C)	–	0.0607	0.0606	U	0.0589	$4.247 \cdot 10^{-9}$	$2.216 \cdot 10^{8}$
	any sequence	–	0.0607	0.0607	NAF	0.0589	$4.247 \cdot 10^{-9}$	$2.216 \cdot 10^{8}$

$$\begin{aligned}
\mathrm{Var}(H_0) &\leqslant L_0\left(\Pi_0 + 2\sum_{\eta=1}^{W-1}\Pi_{0\eta}\right) \\
&= L_0\left(4^{-W} + 2\sum_{k=1}^{[(W-1)/W_0]} 4^{-(W+kW_0)}\right) \\
&= E(H_0)\left(1 + 2\sum_{k=1}^{[(W-1)/W_0]} 4^{-kW_0}\right) \\
&\leqslant E(H_0)\left[1 + 2\cdot 4^{-W_0}/\left(1 - 4^{-W_0}\right)\right].
\end{aligned} \tag{B.9}$$

From the seventh column of Table 13 we find that κ_0, the probability that the regulatory sequence contains $\boldsymbol{b}_1$ at time $t = 0$, varies marginally between various choices of binding site $\boldsymbol{b}_1$, except when $\boldsymbol{b}_1$ has identical letters ($W_0 = 1$). We will use (B.9) in order give an upper bound for how much κ_0 is reduced due to overdispersion. Indeed, insertion of (B.9) into the negative binomial probability distribution formula in (67), gives

$$a = \frac{E(H_0)^2}{\mathrm{Var}(H_0) - E(H_0)} \geqslant \frac{4^{W_0} - 1}{2} \cdot E(H_0),$$

and a lower bound

$$\kappa_0 \approx P(H_0 > 0) \approx 1 - \left(1 + \frac{E(H_0)}{a}\right)^{-a} \geqslant 1 - \exp[-O(W_0)\cdot E(H_0)]$$

for κ_0, with

$$O(W_0) = \frac{4^{W_0} - 1}{2}\log\left[1 + \frac{2}{4^{W_0} - 1}\right] = \begin{cases} 0.7662, & W_0 = 1, \\ 0.9387, & W_0 = 2, \\ 0.9845, & W_0 = 3. \end{cases}$$

A comparison with the corresponding probability $\kappa_{1,\mathrm{Poisson}}$ when there is no overdispersion, reveals that

$$\frac{\kappa_0}{\kappa_{0,\mathrm{Poisson}}} \approx \frac{1 - \exp[-O(W_0)\cdot E(H_0)]}{1 - \exp[-E(H_0)]} \geqslant O(W_0).$$

Consequently κ_0 drops by at most 23.4%, 6.1 %, and 1.55 % due to overdispersion when W_0 equals 1,2, and 3 respectively, and even less for larger values of W_0. Turning to the eighth column of Table 13 we find that the transition rate λ_{10} from the non-absorbing to the absorbing state of $\boldsymbol{C}_t$ is virtually independent of the targeted binding site $\boldsymbol{b}_1$. This can be explained by the fact that the only term of λ_{10} in (73) that involves $\boldsymbol{b}_1$, is $P(H_1 > 0) \approx 1$, regardless of the amount of overdispersion (recall from (81) that $E(H_1) = 994.76$). From this, (73), and (81) it follows that

$$\lambda_{10} \approx N\mu\beta\left(\frac{s_1}{s_0}\right)\cdot L_0\left(1 - 4^{-W}\right)\theta_{10}^{(1)} = N\mu\beta\left(\frac{s_1}{s_0}\right)\cdot L_0 W 4^{-W}, \tag{B.10}$$

for any choice of $\boldsymbol{b}_1$, and a model with selection coefficients s_0 and s_1 for the non-absorbing and absorbing states of $\boldsymbol{C}_t$. In particular, this rate agrees with (4) and it is evaluated in (85) when $W = 6$ and $s_0 = s_1$.

The expected waiting time $E(T_1) = (1 - \kappa_0)/\lambda_{10}$, in the last column of Table 13, is a simple function of κ_0 and λ_{10}, as displayed in (87). Since the targeted binding site $\boldsymbol{b}_1$ has only a limited impact on κ_0 and no impact on the formula for λ_{10} in (B.10), it seems that $E(T_1)$ is quite insensitive to variations in $\boldsymbol{b}_1$. However, in Appendix C.2 we mention below (C.10) that it is possible to modify the transition rate (73) between the components of $\boldsymbol{C}_t$, in such a way that it adjusts for some mutations being counted more than once; those that cause $\boldsymbol{b}_1$ to appear at several overlapping locations along the regulatory sequence. In Appendix C.5 we prove that such an adjustment of (B.10) leads to

$$\lambda_{10} \approx N\mu L_0\beta\left(\frac{s_1}{s_0}\right)\left[L_0 W 4^{-W} - L_{W_0}(W - W_0)4^{-(W+W_0)}\right], \tag{B.11}$$

with $L_{W_0} = L - W - W_0 + 1$. This means that adjustment for overlapping binding sites decreases λ_{10} by a fraction 4^{-W_0} at most, for instance 25% when $W_0 = 1$ and 6.25% when $W_0 = 2$. In Table 14 we have applied (B.11), and computed λ_{10} as well as the expected waiting time $E(T_1)$, for all binding sites $\boldsymbol{b}_1$ of Table 13. We notice from this table that with full overlap adjustment, $E(T_1)$ increases by a factor of 34 % and 29% respectively, when $W = 6$ and $W = 7$, for binding sites with $W_0 = 1$ compared to those that have no self repeats. The corresponding percentages of increase of $E(T_1)$, for $W_0 = 2$, are 5.9 % and 5.0 %. Overall, it can be seen that overlap of binding sites has quite a small effect on $E(T_1)$, when no mismatches are allowed ($\delta_{\max} = 0$). It is possible though that the overlap adjustment is somewhat more important when $\delta_{\max} > 0$.

Appendix C. Derivations or motivations of some mathematical results

C.1. Transition rates of consensus array process $\underline{\boldsymbol{X}}_t$.

In this appendix we will motivate formulas (21) and (22) of Section 5. Recall that these formulas give expressions for the transition rate $\sigma_{\underline{\boldsymbol{xy}}}$ of the fixed state array component process $\underline{\boldsymbol{X}}_t$, between two states $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ that differ in one or two positions.

Starting with the first type of transition in (21), when $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 1$, suppose that the current consensus array of the population is $\underline{\boldsymbol{X}}_t = \underline{\boldsymbol{x}}$. The number of mutations in an offspring in step iii) of the Moran model is approximately Poisson distributed with a mean $mL\mu$ that is small according to (17), since $m \ll N$, so that at most one mutation occurs whenever an offspring is formed. Consequently, since the regulatory array of the parent is $\underline{\boldsymbol{x}}$, if a mutation occurs, then the array $\underline{\boldsymbol{y}}$ of the offspring satisfies $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 1$, with $x_{jl} \neq y_{jl}$ at precisely one position (j, l). The time until the next individual dies is exponentially distributed with rate N, and when a death occurs, the probability is $\mu p_{x_{jl}y_{jl}}$ for the offspring to have a mutation $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{y}}$ at position (j, l). The rate at which a mutation $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{y}}$ appears in *some* individual is therefore $N\mu p_{x_{jl}y_{jl}}$. Then, given that a mutation $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{y}}$ has occurred in one individual of a population where all other individuals have array $\underline{\boldsymbol{x}}$, the newly mutated array $\underline{\boldsymbol{y}}$ will spread and get fixed in the population with a probability that depends on its selection coefficient $s_{h(\underline{\boldsymbol{y}})}$ relative to the selection coefficient $s_{h(\underline{\boldsymbol{x}})}$ of the other $N-1$ individuals in the population. Because of (17), we can ignore the possibility of other new mutations while $\underline{\boldsymbol{y}}$ gets fixed, and also the time it takes for $\underline{\boldsymbol{y}}$ to spread, if it does. Therefore, a mutation $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{y}}$ will immediately be accepted with a probability that equals the product of the fixation probability $\beta\left(s_{h(\underline{\boldsymbol{y}})}/s_{h(\underline{\boldsymbol{y}})}\right)$, and the acceptance probability $\gamma^{1\left(h(\underline{\boldsymbol{y}})<h(\underline{\boldsymbol{x}})\right)}$. Putting things together, we find that the time until a mutation $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{y}}$ occurs, gets fixed and is accepted, can be viewed as the first event of a Poisson process with rate $N\mu p_{x_{jl}y_{jl}}$ and a thinning probability $\beta\left(s_{h(\underline{\boldsymbol{y}})}/s_{h(\underline{\boldsymbol{y}})}\right)\gamma^{1\left(h(\underline{\boldsymbol{y}})<h(\underline{\boldsymbol{x}})\right)}$. But this is another Poisson process with a rate (21) that equals the product of these two terms. Therefore, (21) gives the transition rate between the two fixed states $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$.

Next we motivate formula (22) for the transition rate $\sigma_{\underline{\boldsymbol{xy}}}$ between two arrays $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ with $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 2$, such that $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ differ at positions (j_1, l_1) and (j_2, l_2). Since $\underline{\boldsymbol{X}}_t$ is a multitype Moran process, where different consensus arrays $\underline{\boldsymbol{y}}$ compete to get fixed, an exact expression for $\sigma_{\underline{\boldsymbol{xy}}}$ would require solving a complicated system of equations. This is beyond the scope of this paper, and we will rather make a heuristic argument in two steps. First we find the rate at which *some* double mutation appears an gets fixed, and then we approximate the probability that this double mutation was indeed $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{y}}$.

Recall from Section 5 that there are two possible paths $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{v}} \to \underline{\boldsymbol{y}}$ and $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{w}} \to \underline{\boldsymbol{y}}$, where $\underline{\boldsymbol{v}}$ and $\underline{\boldsymbol{w}}$ differ from $\underline{\boldsymbol{x}}$ at positions (j_1, l_1) and (j_2, l_2) respectively. The total rate from $\underline{\boldsymbol{x}}$ to $\underline{\boldsymbol{y}}$ is given by the sum

$$\sigma_{\underline{\boldsymbol{xy}}} = \sigma_{\underline{\boldsymbol{xvy}}} + \sigma_{\underline{\boldsymbol{xwy}}} \tag{C.1}$$

of the rates along these two paths. In the rest of this appendix we will approximate the two terms on the right hand side of (C.1). Because of symmetry, we will mainly consider the first path $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{v}} \to \underline{\boldsymbol{y}}$. For this path we must first have a mutation $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{v}}$ at rate $N\mu p_{x_{j_1l_1},y_{j_1l_1}}$, and after this change the population has $N-1$ individuals with array $\underline{\boldsymbol{x}}$ and one individual with array $\underline{\boldsymbol{v}}$. Then, before the newly mutated array $\underline{\boldsymbol{v}}$ either dies out or increases to a high frequency, with probability $q_{\underline{\boldsymbol{xv}}}$, it first has another offspring $\underline{\boldsymbol{u}}$ that differs from $\underline{\boldsymbol{v}}$ in some position $(j, l) \neq (j_1, l_1)$ and then gets fixed. This offspring $\underline{\boldsymbol{u}}$ equals $\underline{\boldsymbol{y}}$ (i.e. $(j, l) = (j_2, l_2)$ and $u_{j_2l_2} = y_{j_2l_2}$) with probability $\alpha_{\underline{\boldsymbol{xvy}}}$. Finally, $\underline{\boldsymbol{y}}$ is accepted with probability $\gamma^{1\left(h(\underline{\boldsymbol{y}})<h(\underline{\boldsymbol{x}})\right)}$ if it gets fixed. Putting things together, the rate at which $\underline{\boldsymbol{y}}$ appears through the first path $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{v}} \to \underline{\boldsymbol{y}}$, is

$$\sigma_{\underline{\boldsymbol{xvy}}} = N\mu p_{x_{j_1l_1}y_{j_1l_1}} \cdot q_{\underline{\boldsymbol{xv}}} \cdot \alpha_{\underline{\boldsymbol{xvy}}} \cdot \gamma^{1\left(h(\underline{\boldsymbol{y}})<h(\underline{\boldsymbol{x}})\right)}.$$

In order to proceed, we need to find expressions for $q_{\underline{\boldsymbol{xv}}}$ and $\alpha_{\underline{\boldsymbol{xvy}}}$. For this purpose, we introduce $s_{\underline{\boldsymbol{xv}}} = s_{h(\underline{\boldsymbol{v}})}/s_{h(\underline{\boldsymbol{x}})}$, the ratio between the selection coefficients of $\underline{\boldsymbol{v}}$ and $\underline{\boldsymbol{x}}$, the total rate $(mL-1)\mu$ at which $\underline{\boldsymbol{v}}$ mutates to $\underline{\boldsymbol{u}}$ at some position different from (j_1, l_1), and the probability $\psi_{\underline{\boldsymbol{xv}}}$ that a randomly chosen $\underline{\boldsymbol{u}}$ gets fixed. Eq. (9) of Iwasa et al. (2004) provides an approximation

$$q_{\underline{\boldsymbol{xv}}} = \frac{s_{\underline{\boldsymbol{xv}}} - 1 + \sqrt{\left(s_{\underline{\boldsymbol{xv}}} - 1\right)^2 + 2\left(1 + s_{\underline{\boldsymbol{xv}}}\right)s_{\underline{\boldsymbol{xv}}} \cdot (mL-1)\mu \cdot \psi_{\underline{\boldsymbol{xv}}}}}{1 + s_{\underline{\boldsymbol{xv}}}}$$

of the probability that $\underline{\boldsymbol{u}}$ appears as an offspring of $\underline{\boldsymbol{v}}$ and gets fixed. Assume there is no competition between different possible mutations $\underline{\boldsymbol{v}} \to \underline{\boldsymbol{u}}$ to spread in the population and get fixed. In order to find $\psi_{\underline{\boldsymbol{xv}}}$ we therefore condition on which mutation $\underline{\boldsymbol{v}} \to \underline{\boldsymbol{u}}$ that occurs, and then assume that this mutation spreads and gets fixed with a probability $\beta\left(s_{h(\underline{\boldsymbol{u}})}/s_{h(\underline{\boldsymbol{x}})}\right)$ that ignores the contribution from possible other mutations $\underline{\boldsymbol{v}} \to \underline{\boldsymbol{u}}'$. Summing over all possible mutations $\underline{\boldsymbol{v}} \to \underline{\boldsymbol{u}}$, which occur with probabilities $\mu p_{x_{jl}u_{jl}}$ in step iii) of the Moran model in Section 4, and normalizing with the total probability $\mu(mL-1)$ at which *some* mutation occurs, we find that

$$\psi_{\underline{\boldsymbol{xv}}} = \frac{1}{mL-1}\sum_{(j,l)\neq(j_1,l_1)}\sum_{u_{jl}\neq x_{jl}} p_{x_{jl}u_{jl}}\beta\left(\frac{s_{h(\underline{\boldsymbol{u}})}}{s_{h(\underline{\boldsymbol{x}})}}\right), \tag{C.2}$$

and consequently

$$\alpha_{\underline{\boldsymbol{xvy}}} = \frac{p_{x_{j_2l_2}y_{j_2l_2}}\beta\left(\frac{s_{h(\underline{\boldsymbol{y}})}}{s_{h(\underline{\boldsymbol{x}})}}\right)}{(mL-1)\psi_{\underline{\boldsymbol{xv}}}}.$$

Combining the last four displayed equations, we find that

$$\sigma_{\underline{\boldsymbol{xvy}}} = N\mu(mL-1)^{-1}p_{x_{j_1l_1}y_{j_1l_1}}p_{x_{j_2l_2}y_{j_2l_2}}r_{\underline{\boldsymbol{xv}}}\beta\left(\frac{s_{h(\underline{\boldsymbol{y}})}}{s_{h(\underline{\boldsymbol{x}})}}\right)\gamma^{1\left(h(\underline{\boldsymbol{y}})<h(\underline{\boldsymbol{x}})\right)}, \tag{C.3}$$

with

$$\begin{aligned} r_{\underline{\boldsymbol{xv}}} &= \frac{q_{\underline{\boldsymbol{xv}}}}{\psi_{\underline{\boldsymbol{xv}}}} \\ &= \frac{s_{\underline{\boldsymbol{xv}}} - 1 + \sqrt{\left(s_{\underline{\boldsymbol{xv}}} - 1\right)^2 + 2\left(1 + s_{\underline{\boldsymbol{xv}}}\right)s_{\underline{\boldsymbol{xv}}} \cdot (mL-1)\mu \cdot \psi_{\underline{\boldsymbol{xv}}}}}{\left(1 + s_{\underline{\boldsymbol{xv}}}\right)\psi_{\underline{\boldsymbol{xv}}}}. \end{aligned} \tag{C.4}$$

A similar argument gives the transition rate along the other path $\underline{\boldsymbol{x}} \to \underline{\boldsymbol{w}} \to \underline{\boldsymbol{y}}$, with $r_{\underline{\boldsymbol{xw}}}$ instead of $r_{\underline{\boldsymbol{xv}}}$ in formula (C.3). Inserting these two rates into (C.1), we finally obtain (22).

Formula (C.2) is not very explicit, and therefore we need a simple estimate of $\psi_{\underline{\boldsymbol{xv}}}$ to insert into (C.4), in order to get a more explicit expression for $r_{\underline{\boldsymbol{xv}}}$. Since $r_{\underline{\boldsymbol{xv}}}$ is a decreasing function of $\psi_{\underline{\boldsymbol{xv}}}$, an approximate lower of $\psi_{\underline{\boldsymbol{xv}}}$ will provide an approximate upper bound of $r_{\underline{\boldsymbol{xv}}}$ and $\sigma_{\underline{\boldsymbol{xvy}}}$. Since most possible mutations $\underline{\boldsymbol{v}} \to \underline{\boldsymbol{u}}$ will not change the selection coefficient $(s_{h(\underline{\boldsymbol{u}})} = s_{h(\underline{\boldsymbol{v}})})$, we will use

$$\psi_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}} \geqslant \left(1 - \frac{p_{x_{j_2 l_2} y_{j_2 l_2}}}{mL-1}\right)\beta\left(\frac{s_{h(\underline{\boldsymbol{v}})}}{s_{h(\underline{\boldsymbol{x}})}}\right) + \frac{p_{x_{j_2 l_2} y_{j_2 l_2}}}{mL-1}\beta\left(\frac{s_{h(\underline{\boldsymbol{y}})}}{s_{h(\underline{\boldsymbol{x}})}}\right). \tag{C.5}$$

An analogous approximate lower bound of $\psi_{\underline{\boldsymbol{x}\boldsymbol{w}}}$ is inserted into the expression for $r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}}$, and this provides an approximate upper bound for $\sigma_{\underline{\boldsymbol{x}\boldsymbol{w}}\underline{\boldsymbol{y}}}$. Altogether, this gives an approximate upper bound of $\sigma_{\underline{\boldsymbol{x}}\underline{\boldsymbol{y}}}$.

Formula (C.4) simplifies considerably when $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{v}}$ have the same fitness, i.e. when $s_{h(\underline{\boldsymbol{x}})} = s_{h(\underline{\boldsymbol{v}})}$ and $s_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}} = 1$. We then have that

$$r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}} = \sqrt{\frac{(mL-1)\mu}{\psi_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}}}}. \tag{C.6}$$

If, in addition, the model is selectively neutral, so that $s_{h(\underline{\boldsymbol{x}})} = s_{h(\underline{\boldsymbol{v}})} = s_{h(\underline{\boldsymbol{x}})}$, we deduce $\psi_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}} = N^{-1}$ from (C.2), and consequently $r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}} = \sqrt{N(mL-1)\mu}$, as mentioned below (23).

C.2. Single locus transition rates for the array component process $\boldsymbol{C}_t$.

In this appendix we will motivate formula (73) for the single mutation transition rate $\lambda^{(1)}_{\boldsymbol{cd}}$ of the array component process $\boldsymbol{C}_t$, between two array components $\boldsymbol{c} = (c_1, \ldots, c_m)$ and $\boldsymbol{d} = (d_1, \ldots, d_m)$, defined as in Section 6.2. We have that

$$\lambda^{(1)}_{\boldsymbol{cd}} = \kappa_{\boldsymbol{c}}^{-1} \sum_{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}} \pi_{\underline{\boldsymbol{x}}} \sum_{\substack{\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}} \\ |\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 1}} \sigma_{\underline{\boldsymbol{x}}\underline{\boldsymbol{y}}}, \tag{C.7}$$

where the sum contains those terms of (52) that are due to single mutations, i.e. $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 1$, whereas $\kappa_{\boldsymbol{c}} = P(\boldsymbol{C}_t = \boldsymbol{c})$ is the marginal distribution of $\boldsymbol{C}_t$. Since $\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}$ and $\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}}$ we may assume that $\boldsymbol{c}$ and $\boldsymbol{d}$ only differ at one component in such a way that $|\boldsymbol{d} - \boldsymbol{c}| = 1$ and $d_j = c_j \pm 1$. Using (20), (21), and (26), we find that

$$\begin{aligned} \lambda^{(1)}_{\boldsymbol{cd}} &= N\mu\beta\left(\tfrac{s_{g(\boldsymbol{d})}}{s_{g(\boldsymbol{c})}}\right)\gamma^{1(g(\boldsymbol{d})<g(\boldsymbol{c}))} \\ &\cdot \kappa_{\boldsymbol{c}}^{-1} \prod_{j';j' \neq j} \pi\left(\mathcal{X}_{j'c_{j'}}\right) \sum_{\boldsymbol{x}_j \in \mathcal{X}_{jc_j}} \pi_{\boldsymbol{x}_j} \sum_{\substack{\boldsymbol{y}_j \in \mathcal{X}_{jd_j} \\ |\boldsymbol{y}_j - \boldsymbol{x}_j| = 1}} p_{x_{jl}y_{jl}}, \end{aligned} \tag{C.8}$$

with $l = l(\boldsymbol{x}_j, \boldsymbol{y}_j)$ the unique locus at which $\boldsymbol{x}_j$ and $\boldsymbol{y}_j$ differ. Because of the formula for $\pi_{\underline{\boldsymbol{x}}} = \prod_j \pi_{\boldsymbol{x}_j}$ in (20), it was possible to sum out the contribution from all $j' \neq j$.

In order to simplify (C.8) further, we will find an explicit approximation of its inner sum. For the moment we will simplify notation and put $c = c_j$ and $d = d_j$, so that the inner sum of (C.8) is the sum of the probabilities that a single mutation changes a regulatory sequence $\boldsymbol{x}_j \in \mathcal{X}_{jc}$ of region j (cf. (30)) to another regulatory sequence $\boldsymbol{y}_j$, for all possible $\boldsymbol{y}_j \in \mathcal{X}_{jd}$, where $\mathcal{X}_{jc}$ and $\mathcal{X}_{jd}$ are neighboring components of regulatory sequences ($d = c \pm 1$). In order for a single mutation to cause a transition from $\boldsymbol{x}_j$ to $\boldsymbol{y}_j \in \mathcal{X}_{jd}$, it is necessary that $\boldsymbol{x}_j$ contains at least one binding site $\boldsymbol{b} \in \mathcal{B}_{jc}$, from which a transition into some other binding site in $\mathcal{B}_{jd}$ is possible. It will be helpful to introduce

$$p^{(1)}_{\boldsymbol{b}\mathcal{B}} = \sum_{\boldsymbol{b}' \in \mathcal{B}; |\boldsymbol{b}' - \boldsymbol{b}| = 1} p_{b_w b'_w}, \tag{C.9}$$

the sum of the probabilities that a single mutation changes binding site $\boldsymbol{b} = (b_1, \ldots, b_W)$ in position w to another binding site $\boldsymbol{b}' = (b'_1, \ldots, b'_W) \in \mathcal{B}$. Based on (C.9) and the definition of the hit variables in (37) we approximate the inner sum of (C.8) by

$$\sum_{\substack{\boldsymbol{y}_j \in \mathcal{X}_{jd} \\ |\boldsymbol{y}_j - \boldsymbol{x}_j| = 1}} p_{x_{jl}y_{jl}} \approx \begin{cases} \sum_{\boldsymbol{b} \in \mathcal{B}_{jc}} H_{\{\boldsymbol{b}\}}(\boldsymbol{x}_j) p^{(1)}_{\boldsymbol{b}\mathcal{B}_{jd}}, & d = c-1, \\ 1\left(H_{\mathcal{B}_{jc}}(\boldsymbol{x}_j) = 1\right) \sum_{\boldsymbol{b} \in \mathcal{B}_{jc}} H_{\{\boldsymbol{b}\}}(\boldsymbol{x}_j) p^{(1)}_{\boldsymbol{b}\mathcal{B}_{jd}}, & d = c+1, \end{cases} \tag{C.10}$$

with l the unique locus where $\boldsymbol{x}_j$ and $\boldsymbol{y}_j$ differ. In the lower part of (C.10) we used the fact that when $d = c+1$, in order for a mutation from $\boldsymbol{b} \in \mathcal{B}_{jc}$ into $\mathcal{B}_{jd}$ to cause a change of the regulatory sequence component from $\mathcal{X}_{jc}$ to $\mathcal{X}_{jd}$, it is necessary that $\boldsymbol{x}_j$ contains no other substring of length W that belongs to $\mathcal{B}_{jc}$ (i.e. $H_{\mathcal{B}_{jc}}(\boldsymbol{x}_j) = 1$), since otherwise the distance (14) from $\boldsymbol{x}_j$ to the target of region j will not change.

The approximation in (C.10) relies on the fact that each mutation $x_{jl} \to y_{jl}$, with $x_{jl} \in b\boldsymbol{x}_j$ and $y_{jl} \in b\boldsymbol{y}_j$, only causes one of all L_0 substrings of $\boldsymbol{x}_j$ of length W to switch from $\mathcal{B}_{jc}$ to $\mathcal{B}_{jd}$. In order for this approximation to be accurate, it is essentially required that the substrings of $\boldsymbol{x}_j$ of length W that belong to $\mathcal{B}_{jc}$ and can be mutated into $\mathcal{B}_{jd}$, are not tightly clustered along the regulatory sequence of gene j. On the other hand, if some mutations cause several substrings of $\boldsymbol{x}_j$ to change from $\mathcal{B}_{jc}$ to $\mathcal{B}_{jd}$, they will be counted more than once on the right hand side of (C.10). It is then possible to modify the right hand side of (C.10) by means of a union-intersection argument (Behrens and Vingron, 2010), which involves a sum over n-tuples of overlapping substrings of $\boldsymbol{x}_j$ of length W that belong to $\mathcal{B}_{jc}$, for $n = 2, \ldots, W-1$. For each such n-tuple of overlapping substrings, there is an inner sum over mutations, which in contrast (C.9) is *not* taken over W loci, but rather at those $0 < W' < W$ loci l where the n substrings overlap *and* the n mutated substrings all belong to $\mathcal{B}_{jd}$. Such a union-intersection principle is employed for a special case in Appendix B.4 and Appendix C.5.[6]

In the general case, we will however assume that (C.10) holds, since it is simple and conservative. That is, it will give an upper bound on the transition rates $\lambda^{(1)}_{\boldsymbol{cd}}$ in (71), and hence a lower bound on the expected waiting time $E(T_m)$ in (59), until all m binding site targets have appeared. Given that (C.10) is assumed, we will use it in order to approximate the double sum of (C.8). In this process, we will make use of the formulas

$$\begin{aligned} E(H_{j\boldsymbol{b}}|H_{jc} > 0) &\approx \pi_{\boldsymbol{b}} E(H_{jc}) / [\Pi_{jc} P(H_{jc} > 0)], \\ E[H_{j\boldsymbol{b}} 1(H_{jc} = 1)|H_{jc} > 0] &\approx \pi_{\boldsymbol{b}} P(H_{jc} = 1) / [\Pi_{jc} P(H_{jc} > 0)] \end{aligned} \tag{C.11}$$

for the hit variables $H_{j\boldsymbol{b}} = H_{\{\boldsymbol{b}\}}(\boldsymbol{X}_{0j})$ and $H_{jc} = H_{\mathcal{B}_{jc}}(\boldsymbol{X}_{0j})$, with $\boldsymbol{b} \in \mathcal{B}_{jc}, \pi_{\boldsymbol{b}} = \prod_{w=1}^{W} \pi_{b_w}$, and $\Pi_{jc} = \sum_{\boldsymbol{b} \in \mathcal{B}_{jc}} \pi_{\boldsymbol{b}}$. Formula (C.11) is exact whenever H_{jc} is Poisson distributed, and approximately so when its distribution is negative binomial, with a small amount of overdispersion. Inserting (C.10) into the double sum of (C.8), we find that

$$\begin{aligned} &\sum_{\boldsymbol{x}_j \in \mathcal{X}_{jc}} \pi_{\boldsymbol{x}_j} \sum_{\substack{\boldsymbol{y}_j \in \mathcal{X}_{jd} \\ |\boldsymbol{y}_j - \boldsymbol{x}_j| = 1}} p_{x_{jl}y_{jl}} \\ &= \pi(\mathcal{X}_{jc}) E\left[\sum_{\substack{\boldsymbol{y}_j \in \mathcal{X}_{jd} \\ |\boldsymbol{y}_j - \boldsymbol{X}_{0j}| = 1}} p_{X_{0jl}y_{jl}} | \boldsymbol{X}_{0j} \in \mathcal{X}_{jc}\right] \\ &\approx \pi(\mathcal{X}_{jc}) E\left[1((H_{jc}) = 1)^{1(d=c+1)} \sum_{\boldsymbol{b} \in \mathcal{B}_{jc}} H_{j\boldsymbol{b}} p^{(1)}_{\boldsymbol{b}\mathcal{B}_{jd}} | \boldsymbol{X}_{0j} \in \mathcal{X}_{jc}\right] \\ &= \pi(\mathcal{X}_{jc}) E\left[1(H_{jc} = 1)^{1(d=c+1)} \sum_{\boldsymbol{b} \in \mathcal{B}_{jc}} H_{j\boldsymbol{b}} p^{(1)}_{\boldsymbol{b}\mathcal{B}_{jd}} | H_{jc} > 0\right] \\ &= \pi(\mathcal{X}_{jc}) \sum_{\boldsymbol{b} \in \mathcal{B}_{jc}} p^{(1)}_{\boldsymbol{b}\mathcal{B}_{jd}} E\left[1(H_{jc} = 1)^{1(d=c+1)} H_{j\boldsymbol{b}} | H_{jc} > 0\right] \\ &\approx \pi(\mathcal{X}_{jc}) P(H_{jc} = 1)^{1(d=c+1)} E(H_{jc})^{1(d=c-1)} \sum_{\boldsymbol{b} \in \mathcal{B}_{jc}} p^{(1)}_{\boldsymbol{b}\mathcal{B}_{jd}} \pi_{\boldsymbol{b}} / [\Pi_{jc} P(H_{jc} > 0)] \\ &= \pi(\mathcal{X}_{jc}) P(H_{jc} = 1)^{1(d=c+1)} E(H_{jc})^{1(d=c-1)} \theta^{(1)}_{jcd} / P(H_{jc} > 0) \end{aligned} \tag{C.12}$$

[6] In more detail, formula (B.11) of Appendix B.4 gives an explicit union-intersection generalization of (C.10), in the special case when $K_1 = 1$ single binding site is targeted without mismatches at $m = 1$ gene.

whenever $d = c \pm 1$, where in the second step of (C.12) we used (C.10), and in the third step we employed (61) and the assumed independence of all hit variables $\{H_{jc}\}_{c=0}^{C_j-1}$ of gene j. In the second last step of (C.12) we made use of (C.11), and finally, in the last step of (C.12) we invoked definition (72) of the transition rate $\theta_{jcd}^{(1)}$ between the two sets $\mathcal{B}_{jc}$ and $\mathcal{B}_{jd}$.

In order to apply (C.12) for computing the single mutation transition rate $\lambda_{\boldsymbol{cd}}^{(1)}$ between two array components indexed by $\boldsymbol{c}$ and $\boldsymbol{d}$, we assume (as mentioned above) that these vectors differ in one position j ($|\boldsymbol{d} - \boldsymbol{c}| = 1$ and $d_j = c_j \pm 1$). Inserting (C.12) into (C.8) and (C.7), and making use of the upper part of (51), we finally arrive at formula (73) for $\lambda_{\boldsymbol{cd}}^{(1)}$.

The more general formula for $\lambda_{\boldsymbol{cd}}^{(1)}$ in (74)-(75), is based on the regulatory array components of Section 6.3. It can be proved in the same way as (73).

C.3. Two locus transition rates for the array component process $\boldsymbol{C}_t$.

In this appendix we will motivate formulas (77) and (78) for the double mutation transition rate $\lambda_{\boldsymbol{cd}}^{(2)}$ of the array component process $\boldsymbol{C}_t$, between two array components $\boldsymbol{c} = (c_1, \ldots, c_m)$ and $\boldsymbol{d} = (d_1, \ldots, d_m)$, defined as in Section 6.2, when $|\boldsymbol{d} - \boldsymbol{c}| = 2$, and if $\boldsymbol{c}$ and $\boldsymbol{d}$ differ at one and two components respectively. In analogy with (C.7) we will write

$$\lambda_{\boldsymbol{cd}}^{(2)} = \kappa_{\boldsymbol{c}}^{-1} \sum_{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}} \pi_{\underline{\boldsymbol{x}}} \sum_{\substack{\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}} \\ |\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 2}} \sigma_{\underline{\boldsymbol{xy}}} \tag{C.13}$$

where $\kappa_{\boldsymbol{c}} = P(\boldsymbol{C}_t = \boldsymbol{c})$ is the marginal distribution of $\boldsymbol{C}_t$, whereas the sum contains those terms of (52) which are due to double mutations, i.e. $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 2$. Similarly as in the derivation of (C.8), we use (22) and (26), and find that

$$\begin{aligned}
\lambda_{\boldsymbol{cd}}^{(2)} &= N\mu(mL-1)^{-1}\beta\left(\frac{s_{g(\boldsymbol{d})}}{s_{g(\boldsymbol{c})}}\right)\gamma^{1(g(\boldsymbol{d})<g(\boldsymbol{c}))} \\
&\cdot \kappa_{\boldsymbol{c}}^{-1} \sum_{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}} \pi_{\underline{\boldsymbol{x}}} \sum_{\substack{\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}} \\ |\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 2}} p_{x_{j_1 l_1} y_{j_1 l_1}} p_{x_{j_2 l_2} y_{j_2 l_2}} \left(r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}} + r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{w}}}\right),
\end{aligned} \tag{C.14}$$

where (j_1, l_1) and (j_2, l_2) refer to the two positions where $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ differ, whereas $\underline{\boldsymbol{v}} = \underline{\boldsymbol{v}}\left(\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}}\right)$ and $\underline{\boldsymbol{w}} = \underline{\boldsymbol{w}}\left(\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}}\right)$ are the two regulatory arrays between $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$, as defined above (22).

Consider first the case when $\boldsymbol{c}$ and $\boldsymbol{d}$ differ at one single component $j = j_1 = j_2$, with $d_j = c_j \pm 2$. Let $\boldsymbol{x}_j \in \mathcal{X}_{jc_j}$ be a vector that belongs to the c_j:th component (30) of regulatory sequences of region j. This vector switches to $\boldsymbol{y}_j \in \mathcal{X}_{jd_j}$ through two mutations at loci l_1 and l_2. Since l_1 and l_2 must be located within the same binding site ($|l_2 - l_2| < W$), in analogy with (C.9) we introduce the transition probability

$$p_{\boldsymbol{b}\mathcal{B}}^{(2)} = \sum_{\boldsymbol{b}' \in \mathcal{B}; |\boldsymbol{b}' - \boldsymbol{b}| = 2} p_{b_v b'_v} p_{b_w b'_w} \tag{C.15}$$

that two mutation in the binding site $\boldsymbol{b} = (b_1, \ldots, b_W)$ at positions v and w changes it to some other binding site $\boldsymbol{b}' = (b'_1, \ldots, b'_W) \in \mathcal{B}$. Then as in (C.10) we find that

$$\sum_{\substack{\boldsymbol{y}_j \in \mathcal{X}_{jd_j} \\ |\boldsymbol{y}_j - \boldsymbol{x}_j| = 2}} p_{x_{jl_1} y_{jl_1}} p_{x_{jl_2} y_{jl_2}} \approx \begin{cases} \sum_{\boldsymbol{b} \in \mathcal{B}_{jc_j}} H_{\{\boldsymbol{b}\}}\left(\boldsymbol{x}_j\right) p_{\boldsymbol{b}\mathcal{B}_{jd_j}}^{(2)}, & d_j = c_j - 2, \\ 1\left(H_{\mathcal{B}_{jc_j}}\left(\boldsymbol{x}_j\right) = 1\right) 1\left(H_{\mathcal{B}_{j,c_j+1}}\left(\boldsymbol{x}_j\right) = 0\right) & \\ \quad \cdot \sum_{\boldsymbol{b} \in \mathcal{B}_{jc_j}} H_{\{\boldsymbol{b}\}}\left(\boldsymbol{x}_j\right) p_{\boldsymbol{b}\mathcal{B}_{jd_j}}^{(2)}, & d_j = c_j + 2. \end{cases} \tag{C.16}$$

From this it follows that

$$\begin{aligned}
&\sum_{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}} \pi_{\underline{\boldsymbol{x}}} \sum_{\substack{\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}} \\ |\boldsymbol{y}_j - \boldsymbol{x}_j| = 2}} p_{x_{jl_1} y_{jl_1}} p_{x_{jl_2} y_{jl_2}} \approx \prod_{j'=1}^{m} \pi\left(\mathcal{X}_{j'c_{j'}}\right) \cdot \theta_{jc_j d_j}^{(2)} \\
&\cdot \left[P\left(H_{jc_j} = 1\right) P\left(H_{j,c_j+1} = 0\right)\right]^{1\left(d_j = c_j + 2\right)} E\left(H_{jc_j}\right)^{1\left(d_j = c_j - 2\right)} / P\left(H_{jc_j} > 0\right),
\end{aligned} \tag{C.17}$$

by a similar calculation as in (C.12), which involves the identity

$$\begin{aligned}
&E\left[H_{j\boldsymbol{b}} 1\left(H_{jc} = 1\right) 1\left(H_{j,c+1} = 0\right) | H_{jc} > 0\right] \\
&\approx \pi_{\boldsymbol{b}} / \Pi_{jc} \cdot P\left(H_{jc} = 1\right) P\left(H_{j,c+1} = 0\right) / P\left(H_{jc} > 0\right),
\end{aligned}$$

whenever $\boldsymbol{b} \in \mathcal{B}_{jc}$.On the right hand side of (C.17) we also invoked $\theta_{jcd}^{(2)}$, the two-step transition rate between the two sets $\mathcal{B}_{jc}$ and $\mathcal{B}_{jd}$ of binding sites, as defined in (76).Inserting (C.17) into (C.14), and then employing the upper part of (51) and (C.13), we arrive at (77).In this process, we used (26) and (C.4) to deduce that approximately, $r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}} = r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{w}}} = r_{\boldsymbol{cd}'}$, independently of the value of $\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}}$ in the inner sum of (C.14).Let us motivate that $r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}}$ is essentially independent of $\underline{\boldsymbol{y}}$.Indeed, although $\underline{\boldsymbol{v}} = \underline{\boldsymbol{v}}\left(\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}}\right)$ is a function of $\underline{\boldsymbol{y}}$, we still have that $\underline{\boldsymbol{v}} \in \underline{\mathcal{X}}_{\boldsymbol{d}'}$ for all $\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}}$, and consequently $s_{h\left(\underline{\boldsymbol{v}}\right)} = s_{g\left(\boldsymbol{d}'\right)}$ for all $\boldsymbol{v}$ in the definition of $r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}}$ in (C.4). The only part of $r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}}$ that slightly depends on $\underline{\boldsymbol{y}}$ is $\psi_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}}$ in (C.2), but this will only have a small impact on $r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}}$. (For instance, for the Jukes-Cantor model (2), the lower bound of $\psi_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}}$ in (C.5) only depends on $\underline{\boldsymbol{y}}$ through $s_{h\left(\underline{\boldsymbol{y}}\right)}$, which is constant over $\underline{\mathcal{X}}_{\boldsymbol{d}}$.) An analogous argument reveals that $r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{w}}}$ is essentially independent of $\underline{\boldsymbol{y}}$. Therefore, in (C.14) we moved the term $r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}} + r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{w}}} = 2r_{\boldsymbol{cd}'}$ outside this sum.

Next we consider the case when the two mutations happen at different regulatory sequences j_1 and j_2, so that $d_{j_1} = c_{j_1} \pm 1$ and $d_{j_2} = c_{j_2} \pm 1$. In order to motivate (78), assume that $\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}$ and $\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}}$ are two arrays with $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 2$ that differ at positions (j_1, l_1) and (j_2, l_2). In analogy with (C.10) and (C.16), we make use of the approximation

$$\begin{aligned}
\sum_{\substack{\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}} \\ |\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}| = 2}} p_{x_{j_1 l_1} y_{j_1 l_1}} p_{x_{j_2 l_2} y_{j_2 l_2}} &= \sum_{\substack{\boldsymbol{y}_{j_1} \in \mathcal{X}_{j_1 d_{j_1}} \\ |\boldsymbol{y}_{j_1} - \boldsymbol{x}_{j_1}| = 1}} p_{x_{j_1 l_1} y_{j_1 l_1}} \cdot \sum_{\substack{\boldsymbol{y}_{j_2} \in \mathcal{X}_{j_2 d_{j_2}} \\ |\boldsymbol{y}_{j_2} - \boldsymbol{x}_{j_2}| = 1}} p_{x_{j_2 l_2} y_{j_2 l_2}} \\
&\approx 1\left(H_{\mathcal{B}_{j_1 c_{j_1}}}\left(\boldsymbol{x}_{j_1}\right) = 1\right)^{1\left(d_{j_1} = c_{j_1} + 1\right)} \sum_{\boldsymbol{b} \in \mathcal{B}_{j_1 c_{j_1}}} H_{\{\boldsymbol{b}\}}\left(\boldsymbol{x}_{j_1}\right) p_{\boldsymbol{b}\mathcal{B}_{j_1 d_{j_1}}}^{(1)} \\
&\cdot 1\left(H_{\mathcal{B}_{j_2 c_{j_2}}}\left(\boldsymbol{x}_{j_2}\right) = 1\right)^{1\left(d_{j_2} = c_{j_2} + 1\right)} \sum_{\boldsymbol{b} \in \mathcal{B}_{j_2 c_{j_2}}} H_{\{\boldsymbol{b}\}}\left(\boldsymbol{x}_{j_2}\right) p_{\boldsymbol{b}\mathcal{B}_{j_2 d_{j_2}}}^{(1)},
\end{aligned} \tag{C.18}$$

which in conjunction with (C.13) and (C.14), and a similar calculation as in (C.12), leads to (78). In this process, we used that the array component $\underline{\boldsymbol{x}}$ tunnels through either of $\underline{\boldsymbol{v}}$ or $\underline{\boldsymbol{w}}$ on its way to $\underline{\boldsymbol{y}}$, where $\underline{\boldsymbol{v}}$ and $\underline{\boldsymbol{w}}$ are defined below (21). Since approximately $r_{\boldsymbol{cd}_1} = r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}}$ and $r_{\boldsymbol{cd}_2} = r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{w}}}$, independently of $\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}}$, it follows that term $r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{v}}} + r_{\underline{\boldsymbol{x}}\underline{\boldsymbol{w}}} = r_{\boldsymbol{cd}_1} + r_{\boldsymbol{cd}_2}$ can be moved outside the inner sum of (C.14).

C.4. Markov process assumption for array component process $\boldsymbol{C}_t$

In this appendix we will give conditions under which the Markov assumption of the array component process $\boldsymbol{C}_t$ gives an accurate approximation of the distribution of the waiting time T_m until all targeted binding sites have been fixed. Recall from (47) that $\boldsymbol{C}_t$ lumps the consensus array process $\underline{\boldsymbol{X}}_t$ into $\overline{C}$ different regulatory array components $\underline{\mathcal{X}}_0, \ldots, \underline{\mathcal{X}}_{\overline{C}-1}$ in (24). This lumping introduces some bias of the waiting time distribution. In order to understand what causes this bias, the following two observations are helpful:

1. The less variation there is among the states in $\underline{\mathcal{X}}_{\boldsymbol{c}}$, for a transition of $\underline{\boldsymbol{X}}_t$ into $\underline{\mathcal{X}}_{\boldsymbol{d}}$, for all pairs of array components, the smaller is the bias caused by the lumping of states into these array components.
2. Even though there is large variation among in i), among the states in $\underline{\mathcal{X}}_{\boldsymbol{c}}$, for transitions *out of* $\mathcal{X}_{\boldsymbol{c}}$ into $\mathcal{X}_{\boldsymbol{d}}$, this can be compensated for by having a high degree of mixing *within* $\mathcal{X}_{\boldsymbol{c}}$, i.e. large transition rates between states in $\mathcal{X}_{\boldsymbol{c}}$ for which the transition rates into $\mathcal{X}_{\boldsymbol{d}}$ are very different.

We will mainly address i), but at the end of this appendix we briefly discuss ii) as well. In order to understand why i) causes a bias of the waiting time distribution, recall that although $\underline{\boldsymbol{X}}_t$ is a Markov process, $\boldsymbol{C}_t$ is typically *not* Markovian, since a transition of $\boldsymbol{C}_t$ from $\boldsymbol{c}$ to $\boldsymbol{d}$ corresponds to a jump of $\underline{\boldsymbol{X}}_t$ from *some* element $\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}$ into $\underline{\mathcal{X}}_{\boldsymbol{d}}$, and the total rate

$$\sigma_{\underline{\boldsymbol{x}}\underline{\mathcal{X}}_{\boldsymbol{d}}} = \sum_{\underline{\boldsymbol{y}} \in \underline{\mathcal{X}}_{\boldsymbol{d}}} \sigma_{\underline{\boldsymbol{x}}\underline{\boldsymbol{y}}} \tag{C.19}$$

at which $\underline{\boldsymbol{X}}_t$ changes from $\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}$ to some element in $\underline{\mathcal{X}}_{\boldsymbol{d}}$ will typically vary with $\underline{\boldsymbol{x}}$. If this is the case, we will not be able to discern the rate at which $\boldsymbol{C}_t$ switches from $\boldsymbol{c}$ to $\boldsymbol{d}$, by recording $\boldsymbol{C}_t = \boldsymbol{c}$ alone, since we must also know $\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}$ in order to know the true rate $\sigma_{\underline{\boldsymbol{x}}\underline{\mathcal{X}}_{\boldsymbol{d}}}$ at which a transition of $\underline{\boldsymbol{X}}_t$ from $\underline{\boldsymbol{x}}$ into $\underline{\mathcal{X}}_{\boldsymbol{d}}$ occurs. When the rates in (C.19) vary with $\underline{\boldsymbol{x}} \in \mathcal{X}_{\boldsymbol{c}}$, the past history of $\boldsymbol{C}_t$ might provides some additional information about $\sigma_{\underline{\boldsymbol{x}}\underline{\mathcal{X}}_{\boldsymbol{d}}}$. But this is to say that $\boldsymbol{C}_t$ has lost its memoryless property, and therefore it is no longer a Markov process. Notice however that the *assumed* transition rate of $\boldsymbol{C}_t$ between $\boldsymbol{c}$ and $\boldsymbol{d}$, in (52), is a weighted average

$$\lambda_{\boldsymbol{cd}} = \sum_{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}} \pi_{\underline{\boldsymbol{x}}} \sigma_{\underline{\boldsymbol{x}}\underline{\mathcal{X}}_{\boldsymbol{d}}} / \sum_{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}} \pi_{\underline{\boldsymbol{x}}} \tag{C.20}$$

of the *actual* rates in (C.19), for all $\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}$. In order to quantify how much these actual rates vary around the assumed average in (C.20), we introduce the generalized coefficient of variation

$$\mathrm{CV}_{\boldsymbol{cd}}^{[\alpha]} = \lambda_{\boldsymbol{cd}}^{-1} \left[\sum_{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}} \pi_{\underline{\boldsymbol{x}}} |\sigma_{\underline{\boldsymbol{x}}\underline{\mathcal{X}}_{\boldsymbol{d}}} - \lambda_{\boldsymbol{cd}}|^{\alpha} / \sum_{\underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}} \pi_{\underline{\boldsymbol{x}}} \right]^{1/\alpha}, \tag{C.21}$$

of order $1 < \alpha < \infty$, with $\alpha = 2$ the ordinary coefficient of variation. In view of the above discussion, $\boldsymbol{C}_t$ is a Markov process if and only if $\mathrm{CV}_{\boldsymbol{cd}}^{[\alpha]} = 0$ for all pairs of states $\boldsymbol{c} \neq \boldsymbol{d}$. Now $\mathrm{CV}_{\boldsymbol{cd}}^{[\alpha]}$ is a non-decreasing function of α. The $\alpha = \infty$ limit corresponds to

$$\mathrm{CV}_{\boldsymbol{cd}}^{[\infty]} = \max\left\{ |\frac{\sigma_{\underline{\boldsymbol{x}}\underline{\mathcal{X}}_{\boldsymbol{d}}}}{\lambda_{\boldsymbol{cd}}} - 1|; \underline{\boldsymbol{x}} \in \underline{\mathcal{X}}_{\boldsymbol{c}}, \pi_{\underline{\boldsymbol{x}}} > \varepsilon \right\}$$

with $\varepsilon = 0$. However, we will also allow for $\varepsilon > 0$ in order to get a practically more useful measure. Regardless of which version of the coefficient of variation we use, it might be too restrictive though to require that $\mathrm{CV}_{\boldsymbol{cd}}$ is small for all pairs $\boldsymbol{c}, \boldsymbol{d}$, since the lumping of states is less of a problem for some regulatory arrays than for others. Intuitively, a lumping of states is likely to generate more bias of the waiting time distribution for those regions of $\underline{\mathcal{X}}$ that correspond to regulatory arrays for which some local target is almost reached. In the rest of the appendix we will illustrate this for a system with $m = 1$ gene.

When $m = 1$, the Markov process approximation of $\boldsymbol{C}_t$ and the accompanying phase-type distribution approximation of F_{T_1} in (57) will still be accurate if $\mathrm{CV}_{\boldsymbol{cd}}^{[\alpha]}$ is small for all pairs of states with $\boldsymbol{c} \in \mathcal{C}_{\mathrm{n}}$ non-absorbing and $\boldsymbol{d} \in \mathcal{C}_{\mathrm{a}}$ absorbing, such that $\kappa_{\boldsymbol{c}} = P(\boldsymbol{C}_t = \boldsymbol{c})$ and $\lambda_{\boldsymbol{cd}}$ are not too small. For this reason it is of interest to compute $\mathrm{CV}_{\boldsymbol{cd}}^{[\alpha]}$ between non-absorbing and absorbing pairs of states. It turns out that very explicit expressions of $\mathrm{CV}_{\boldsymbol{cd}}^{[\alpha]}$ can be derived for the Jukes-Cantor model (2). We will consider the regulatory sequence decomposition in (40), which involves the number of mismatches to targeted binding sites as well as values of the associated hit variables.

In more detail, when $\boldsymbol{c} = (c, n)$, component $\mathcal{X}_{\boldsymbol{c}}$ consists of all regulatory sequences $\boldsymbol{x}$ whose substrings of length W have a smallest number of mismatches with a targeted binding site that belongs to the set $\Delta_c = \{\delta_c, \ldots, \delta_{c+1} - 1\}$, whereas the number $H_{\mathcal{B}_c}(\boldsymbol{x})$ of these substrings with number of mismatches in this set, belongs to $\mathcal{H}_{cn}$. Recall from (A.7) that $\mathcal{B}_c$ is the set of words of length W whose number of mismatches with the closest binding site belongs to Δ_c. Analogously, we write $\mathcal{B}_{\mathrm{mc}} \subset \mathcal{B}_c$ for the set of words whose number of mismatches to the closest binding site is exactly δ_c, corresponding to a minus 1 hit in relation to the set Δ_{c-1}. We will give expressions for $\mathrm{CV}_{\boldsymbol{cd}}^{[\alpha]}$ when $\boldsymbol{c} = (c, n)$ and $\boldsymbol{d} = (c - 1, 1)$. Since $\pi_{\boldsymbol{x}} = 4^{-L}$ is constant for the Jukes-Cantor model, and $g(\boldsymbol{d}) \geqslant g(\boldsymbol{c})$, it follows from (72)-(75) and (C.8)-(C.10) that

$$\begin{aligned} \sigma_{\boldsymbol{x}\mathcal{X}_{\boldsymbol{d}}} &= N\mu\beta\left(\frac{s_{g(\boldsymbol{d})}}{s_{g(\boldsymbol{c})}}\right)4^{-L}\sum_{\boldsymbol{b}\in\mathcal{B}_c} H_{\{\boldsymbol{b}\}}(\boldsymbol{x}) p_{\boldsymbol{b}\mathcal{B}_{c-1}}^{(1)} \\ &= N\mu\beta\left(\frac{s_{g(\boldsymbol{d})}}{s_{g(\boldsymbol{c})}}\right)4^{-L}\sum_{\boldsymbol{b}\in\mathcal{B}_{\mathrm{mc}}} H_{\{\boldsymbol{b}\}}(\boldsymbol{x}) p_{\boldsymbol{b}\mathcal{B}_{c-1}}^{(1)} \\ &= N\mu\beta\left(\frac{s_{g(\boldsymbol{d})}}{s_{g(\boldsymbol{c})}}\right)4^{-L} p_{\boldsymbol{b}\mathcal{B}_{c-1}}^{(1)} \cdot H_{B_{\mathrm{mc}}}(\boldsymbol{x}) \\ &= N\mu\beta\left(\frac{s_{g(\boldsymbol{d})}}{s_{g(\boldsymbol{c})}}\right)4^{-L}(|\mathcal{B}_c|/|\mathcal{B}_{\mathrm{mc}}|)\theta_{c,c-1}^{(1)} \cdot H_{B_{\mathrm{mc}}}(\boldsymbol{x}) \\ &\propto H_{B_{\mathrm{mc}}}(\boldsymbol{x}) \end{aligned} \tag{C.22}$$

for all $\boldsymbol{x} \in \mathcal{X}_{\boldsymbol{c}}$, where in the second step we used the fact that a transition from $\boldsymbol{b} \in \mathcal{B}_c$ to $\mathcal{B}_{c-1}$ is only possible when $\boldsymbol{b} \in \mathcal{B}_{\mathrm{mc}}$ is a minus 1 hit, and in the third step that $p_{\boldsymbol{b}\mathcal{B}_{c-1}}^{(1)}$ is the same for all $\boldsymbol{b} \in \mathcal{B}_{\mathrm{mc}}$, for the Jukes-Cantor model. Formula (C.22) implies that the transition rate from $\boldsymbol{x} \in \mathcal{X}_{\boldsymbol{c}}$ into $\mathcal{X}_{\boldsymbol{d}}$ is proportional to the number $H_{B_{\mathrm{mc}}}(\boldsymbol{x})$ of substrings of $\boldsymbol{x}$ of length W that have a minus one hit. Inserting (C.22) into (C.20)-(C.21) we notice that the proportionality constant in (C.22) cancels out, so that

$$\mathrm{CV}^{[\alpha]} = \mu_{\boldsymbol{c}}^{-1} \left[\sum_{\boldsymbol{x}\in\mathcal{X}_{\boldsymbol{c}}} \pi_{\boldsymbol{x}} |H_{B_{\mathrm{mc}}}(\boldsymbol{x}) - \mu_{\boldsymbol{c}}|^{\alpha} / \sum_{\boldsymbol{x}\in\mathcal{X}_{\boldsymbol{c}}} \pi_{\boldsymbol{x}} \right]^{1/\alpha}, \tag{C.23}$$

where $\mu_{\boldsymbol{c}} = \sum_{\boldsymbol{x}\in\mathcal{X}_{\boldsymbol{c}}} \pi_{\boldsymbol{x}} H_{B_{\mathrm{mc}}}(\boldsymbol{x}) / \sum_{\boldsymbol{x}\in\mathcal{X}_{\boldsymbol{c}}} \pi_{\boldsymbol{x}}$. Let $H_c = H_{B_c}(\boldsymbol{X})$ and $H_{\mathrm{mc}} = H_{B_{\mathrm{mc}}}(\boldsymbol{X})$ refer to hit variables of a randomly chosen regulatory sequence $\boldsymbol{X}$. Then we may rewrite (C.23) as

$$\mathrm{CV}_{\boldsymbol{cd}}^{[\alpha]} = \frac{E\left[|H_{\mathrm{mc}} - \mu|^{\alpha} | H_c \in \mathcal{H}_{cn}\right]^{1/\alpha}}{\mu_{\boldsymbol{c}}}, \tag{C.24}$$

where $\mu_{\boldsymbol{c}} = E(H_{\mathrm{mc}} | H_c \in \mathcal{H}_{cn})$. Consequently, the coefficient of variation in (C.24) quantifies how much the near hit variable H_{mc} varies within the regulatory sequence component $\mathcal{X}_{\boldsymbol{c}}$. Notice that $\mathrm{CV}_{\boldsymbol{cd}}^{[\alpha]}$ is independent of the population size, the mutation rate, and the selection coefficients $s_{g(\boldsymbol{c})}$ and $s_{g(\boldsymbol{d})}$. In particular, we have that $\mathrm{CV}_{\boldsymbol{cd}}^{[\alpha]} = 0$ when $\mathcal{X}_{\boldsymbol{c}}$ is as small as possible, in the sense that $\delta_{c+1} = \delta_c + 1$ and $|\mathcal{H}_{cn}| = 1$. It is very convenient to compute (C.24) for the Poisson model (66), with

$$\begin{aligned} H_c &\overset{\mathcal{L}}{\in} \mathrm{Po}\left(L_0 |\mathcal{B}_c| 4^{-W}\right), \\ H_{\mathrm{mc}} | H_c &\overset{\mathcal{L}}{\in} \mathrm{Bin}(H_c, |\mathcal{B}_{\mathrm{mc}}|/|\mathcal{B}_c|). \end{aligned} \tag{C.25}$$

In particular, it follows from that $\mathrm{CV}_{\boldsymbol{cd}}^{[2]} \approx 1/\sqrt{E(H_{\mathrm{mc}})}$ when $P(H_c \in \mathcal{H}_{cn})$ is close to 1.

We will end this appendix by justifying some results of Appendix B.1, where we investigated how sensitive the waiting time distribution is to the coarseness of the regulatory array decomposition

Table 14

Values of the expected waiting time $E(T_1)$ until one single binding site $\boldsymbol{b}_1$ of length $W = 6$ or $W = 7$ occurs, and the transition rate λ_{10} between the non-absorbing and absorbing states 1 and 0 of the regulatory sequence component process $\boldsymbol{C}_t$. The setup is the same as in Table 13, except that λ_{10} has been adjusted for overlapping binding sites, as shown in (B.11). The formula for $E(T_1)$ is taken from (87), with λ_{10} as in this table and κ_0 as in Table 13.

W	$\boldsymbol{b}_1$	W_0	λ_{10}	$E(T_1)$
6	(A,A,A,A,A,A)	1	$1.154 \cdot 10^{-8}$	$7.191 \cdot 10^{7}$
	(A,C,A,C,A,C)	2	$1.397 \cdot 10^{-8}$	$5.697 \cdot 10^{7}$
	(A,C,G,A,C,G)	3	$1.446 \cdot 10^{-8}$	$5.442 \cdot 10^{7}$
	(A,C,G,T,A,C)	4	$1.456 \cdot 10^{-8}$	$5.392 \cdot 10^{7}$
	(A,C,G,T,C,A)	5	$1.457 \cdot 10^{-8}$	$5.382 \cdot 10^{7}$
	(A,C,G,T,C,C)	–	$1.458 \cdot 10^{-8}$	$5.381 \cdot 10^{7}$
7	(A,A,A,A,A,A,A)	1	$3.338 \cdot 10^{-9}$	$2.860 \cdot 10^{8}$
	(A,C,A,C,A,C,A)	2	$4.058 \cdot 10^{-9}$	$2.328 \cdot 10^{8}$
	(A,C,G,A,C,G,A)	3	$4.209 \cdot 10^{-9}$	$2.238 \cdot 10^{8}$
	(A,C,G,T,A,C,G)	4	$4.240 \cdot 10^{-9}$	$2.220 \cdot 10^{8}$
	(A,C,G,T,C,A,C)	5	$4.246 \cdot 10^{-9}$	$2.217 \cdot 10^{8}$
	(A,C,G,T,C,C,A)	6	$4.247 \cdot 10^{-9}$	$2.216 \cdot 10^{8}$
	(A,C,G,T,C,C,C)	–	$4.247 \cdot 10^{-9}$	$2.216 \cdot 10^{8}$

of $\underline{\mathcal{X}}$. In particular, Table 5 indicates that a mismatch-based decomposition (31) with $C = 2$ intervals per gene gives a sufficient accuracy for neutral models, whereas Tables 6 and 7 make it clear that $C = 3$ intervals per gene are sometimes needed for non-neutral models, where the absorbing state has a selective advantage. Table 8 indicates that the finer decomposition (39), that also takes the values of hit variables into account, improves accuracy of the waiting time distribution quite marginally.

Table 15 below is an attempt to justify this theoretically for a system with $m = 1$ gene, by computing the coefficient of variation $\text{CV}_{\boldsymbol{cd}}^{[\alpha]}$ for transitions from the non-absorbing state $\boldsymbol{c} = (1,n)$ to the absorbing state $\boldsymbol{d} = (0,1)$. Recall from (C.24) that the coefficient of variation is independent of the selection coefficients of the model. Overall, it can be seen that $\text{CV}_{\boldsymbol{cd}}^{[2]}$ and $\text{CV}_{\boldsymbol{cd}}^{[4]}$ decrease more dramatically from $C = 2$ to $C = 3$ for those scenarios (31) in Tables 6 and 7 where the waiting time distributions between $C = 2$ and $C = 3$ differ the most, whereas the values of $\text{CV}_{\boldsymbol{cd}}^{[\infty]}$ is somewhat more unpredictable. The behavior $\text{CV}_{\boldsymbol{cd}}^{[2]}$ and $\text{CV}_{\boldsymbol{cd}}^{[4]}$ indicate that it was important to use the decomposition (31) of $\mathcal{X}$ with $C = 3$ for the non-neutral models where the absorbing state has a high fitness, since the transitions rates varied too much within $\mathcal{X}_{(1,n)}$ for the coarser decomposition (31) with $C = 2$. On the other hand, in spite of this it is sufficient to use the coarser decomposition (31) of $\mathcal{X}$ with $C = 2$ for neutral models, since the large variation of transition rates from $\boldsymbol{x} \in \mathcal{X}_{(1,n)}$ into $\mathcal{X}_{(0,1)}$ is compensated for by a high degree of mixing between the states in $\mathcal{X}_{(1,n)}$, in agreement with ii). On the other hand, although all $C = 3$ scenarios in Table 15 that make use of the finer decomposition (39) of states have $\text{CV}_{\boldsymbol{cd}}^{[\alpha]} = 0$, this does not improve the accuracy of the waiting time distribution a lot, since there is already a high degree of mixing within $\mathcal{X}_{(1,n)}$ for the coarser scenario (31) with $C = 3$ that does not make use of hit variables.

Finally, we notice from Table 15 that the conditional probability $P_{\text{m1}} = P(H_{\text{m1}} = 0 | H_1 > 0)$ of no minus-one hits gives additional valuable insight as to why a regulatory sequence decomposition (31) with $C = 2$ might give a poor approximation of F_{T_1} for the non-neutral models where the absorbing state has a high selection coefficient. Whenever P_{m1} is large, the coarser decomposition (31) with $C = 2$ does not account for the fact that those states within

Table 15

Values of the coefficient of variation $\text{CV}_{\boldsymbol{cd}}$, which quantifies how much the transition rate (C.19) of the regulatory sequence process $\boldsymbol{X}_t$ from $\boldsymbol{x}$ into $\mathcal{X}_{\boldsymbol{d}}$, varies among all regulatory sequences $\boldsymbol{x}$ that belong to component $\mathcal{X}_{\boldsymbol{c}}$. The system has $m = 1$ gene, the regulatory sequence has length $L = 1000$ and its targeted binding site(s) length W. There are $C \geqslant 2$ distance classes $\Delta_0, \ldots, \Delta_{C-1}$, with $\Delta_0 = \{0, \ldots, \delta_{\max}\}$ and $\Delta_1 = \{\delta_{\max} + 1, \ldots, \delta_2\}$. The regulatory sequence component $\mathcal{X}_{\boldsymbol{c}}$ corresponds to (40), that is, $\boldsymbol{x} \in \mathcal{X}_{\boldsymbol{c}}$ with $\boldsymbol{c} = (c,n)$ when the distance between a substring of $\boldsymbol{x}$ and the closest targeted binding site(s) belongs to Δ_c and the number H_c of substrings of $\boldsymbol{x}$ with such a distance to the targeted binding site (s), belongs to $\mathcal{H}_{cn}$. In the table $\boldsymbol{d} = (0,1)$ is absorbing with $\mathcal{H}_{01} = \{1\}$, whereas $c = (1,n)$ is such that $\mathcal{H}_{1n}$ varies. The rightmost three columns display $E(H_1), E(H_{\text{m1}})$, and $P_{\text{m1}} = P(H_{\text{m1}} = 0 | H_1 \geqslant 1)$, where H_1 and H_{m1} refer to the number of substrings of a randomly chosen regulatory sequence, whose distance to the targeted binding belongs to Δ_1 and $\{\delta_{\max} + 1\}$ respectively, with distributions as in (C.25). A horisontal bar represents a value that is not uniquely defined, since it depends on α.

				$\text{CV}_{\boldsymbol{cd}}^{[2]}$	$\text{CV}_{\boldsymbol{cd}}^{[4]}$	$\text{CV}_{\boldsymbol{cd}}^{[\infty]}$ $\varepsilon = 10^{-6}$	$\text{CV}_{\boldsymbol{cd}}^{[\alpha]}$			
					$\mathcal{H}_{1n}$					
W	$\delta_{\max}$	C	Δ_1		$\{1,2,\ldots\}$		$\{n\}$	$E(H_{\text{m1}})$	$E(H_1)$	P_{m1}
6	0	2	$\{1,\ldots,6\}$	0.478	0.641	2.202	–	4.37	994.76	0.013
		$\geqslant 3$	$\{1\}$	0.462	0.620	2.839	0.000	4.37	4.37	0.000
	1	2	$\{2,\ldots,6\}$	0.175	0.230	0.677	–	32.79	990.38	$5.72 \cdot 10^{-15}$
		$\geqslant 3$	$\{2\}$	0.175	0.230	0.891	0.000	32.79	32.79	0.000
8	0	2	$\{1,\ldots,8\}$	1.658	2.568	10.000	-	0.36	992.98	0.695
		$\geqslant 3$	$\{1\}$	0.378	0.664	4.030	0.000	0.36	0.36	0.000
	1	2	$\{2,\ldots,8\}$	0.512	0.688	2.405	–	3.82	992.62	0.022
		$\geqslant 3$	$\{2\}$	0.484	0.653	3.098	0.000	3.82	3.82	0.000
10	0	2	$\{1,\ldots,10\}$	5.939	14.771	69.540	–	0.03	991.00	0.972
		$\geqslant 3$	$\{1\}$	0.118	0.347	1.958	0.000	0.03	0.03	0.000
	1	2	$\{2,\ldots,10\}$	1.616	2.488	9.450	–	0.38	990.97	0.682
		$\geqslant 3$	$\{2\}$	0.386	0.672	3.985	0.000	0.38	0.38	0.000

$\mathcal{X}_{(1,n)}$ with no minus-one hits, are much more distant from the target for the non-neutral models than those that have a least one minus-one hit.

C.5. Adjusting transition rates for overlapping targeted binding sites.

In this appendix we will consider the special case of $m = 1$ regulatory sequence, along which $K = 1$ binding site $\boldsymbol{b}_1$ of length W is targeted without mismatch ($\delta_{\max} = 0$). We will prove the explicit expression (B.11) for the transition rate λ_{10} from the non-absorbing state 1 of the regulatory sequence component process $\boldsymbol{C}_t$ in (79), to the absorbing state 0. Whereas (B.10) assumes that the local target is reached at isolated locations along the regulatory sequence $\boldsymbol{X}_t$ (see Appendix C.2), formula (B.11) takes overlaps of the targeted binding site $\boldsymbol{b}_1$ into account. As in Behrens and Vingron (2010), we will use a union-intersection principle in order to handle such overlaps.

Overlapping targets can only appear when $\boldsymbol{b}_1$ has self repeats with cycle length W_0 for some $1 \leqslant W_0 \leqslant W - 1$. When $\boldsymbol{b}_1$ has no self repeats, we may formally choose W_0 to equal W. The two states of $\boldsymbol{C}_t$ correspond to sets $\mathcal{X}_0$ and $\mathcal{X}_1$ of regulatory sequences, where $\mathcal{X}_0$ contains all sequences $\boldsymbol{y}$ for which at least one subsequence of length W equals $\boldsymbol{b}_1$. The other set $\mathcal{X}_1$ contains all remaining regulatory sequences. Therefore $\boldsymbol{C}_t = 0$ corresponds to a selection coefficient s_1, whereas $\boldsymbol{C}_t = 1$ corresponds to a selection coefficient s_0. Since $\boldsymbol{C}_t$ has only two states, we need not consider double mutations. It therefore follows from (51), (71), and (73) that

$$\lambda_{10} = \frac{1}{\pi(\mathcal{X}_1)} \cdot N\mu\beta\left(\frac{s_1}{s_0}\right)S, \tag{C.26}$$

where

$$S = \sum_{\boldsymbol{x}\in\mathcal{X}_1} \pi_{\boldsymbol{x}} \sum_{\substack{\boldsymbol{y}\in\mathcal{X}_0 \\ |\boldsymbol{y}-\boldsymbol{x}|=1}} p_{x_{l_0} y_{l_0}} \tag{C.27}$$

corresponds to the double sum in (C.12) (when $j = 1, c_j = 1$ and $d_j = 0$), and $l_0 = l_0(\boldsymbol{x},\boldsymbol{y})$ is the unique locus where $\boldsymbol{x}$ and $\boldsymbol{y}$ differ. In order to prove (B.11), we will provide an exact expression for S that adjusts for overlapping copies of $\boldsymbol{b}_1$. To this end, we introduce

$$n_{\max} = 1 + \left\lfloor \frac{W-1}{W_0} \right\rfloor,$$

the maximum number of copies of $\boldsymbol{b}_1$ that may all overlap with each other, along some portion of $\boldsymbol{y} \in \mathcal{X}_0$. For any $1 \leqslant n \leqslant n_{\max}$, there are $\binom{n_{\max}}{n}$ possible vectors $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$ of length n that indicate the lags $0 = \eta_1 < \eta_2 < \ldots < \eta_n \leqslant (n_{\max} - 1)W_0$ of n overlapping copies of $\boldsymbol{b}_1$ along $\boldsymbol{y}$, relative to the leftmost locus of the leftmost copy of $\boldsymbol{b}_1$. Recalling that $\boldsymbol{b}_1$ has self repeat frequency W_0, it follows that each η_k must be divisible by W_0. Since the first component $\eta_1 = 0$ is fixed, the set $\mathcal{L} = \mathcal{L}(W_0, W)$ of such lag vectors $\boldsymbol{\eta}$ has size $2^{n_{\max}-1}$. By a union-intersection argument, we have that

$$S = \sum_{\boldsymbol{\eta}\in\mathcal{L}} (-1)^{|\boldsymbol{\eta}|} S_{\boldsymbol{\eta}}, \tag{C.28}$$

where $|\boldsymbol{\eta}| = n$ is the number of elements of $\boldsymbol{\eta}$, and

$$S_{\boldsymbol{\eta}} = \sum_{\boldsymbol{x}\in\mathcal{X}_1} \pi_{\boldsymbol{x}} S_{\boldsymbol{\eta}\boldsymbol{x}} \tag{C.29}$$

is the contribution to (C.27) from terms where overlapping copies of $\boldsymbol{b}_1$ occurs along $\boldsymbol{y} \in \mathcal{X}_0$ at locations with lags $\boldsymbol{\eta}$ compared to some leftmost locus l, whereas $S_{\boldsymbol{\eta}\boldsymbol{x}}$ is the corresponding contribution to the inner sum of (C.27), when $\boldsymbol{x} \in \mathcal{X}_1$ is the regulatory sequence before a mutation occurs. In order to find an explicit expression for $S_{\boldsymbol{\eta}}$, we need to give conditions under which it is possible for $\boldsymbol{x}$ to transition into $\boldsymbol{y} \in \mathcal{X}_0$ through one single mutation. For this to happen, the subsequences $\boldsymbol{x}_{l+\eta_k:l+\eta_k+W-1}$ of length W have to switch to $\boldsymbol{y}_{l+\eta_k:l+\eta_k+W-1} = \boldsymbol{b}_1$ for $k = 1, \ldots, |\boldsymbol{\eta}|$, and therefore the mutation must occur at a location l_0 that satisfies

$$l + \eta_n \leqslant l_0 \leqslant l + W - 1. \tag{C.30}$$

For any $n' \geqslant 0$ we let $\boldsymbol{b}_1(n')$ refer to the sequence of length $W + (n'-1)W_0$ that contains n' copies of $\boldsymbol{b}_1$, so that in particular $\boldsymbol{b}_1(1) = \boldsymbol{b}_1$. Recalling that $n = |\boldsymbol{\eta}|$, let $n' = \eta_n/W_0 + 1$ be the number of W_0-repeats that $\boldsymbol{\eta}$ spans, and define

$$\mathcal{B}_{1\boldsymbol{\eta}} \;=\; \{\boldsymbol{b} = (b_1, \ldots, b_{W+\eta_n}); \boldsymbol{b} = \boldsymbol{b}_1(n') \text{ except at } b_w, \text{ where } \eta_n + 1 \leqslant w \leqslant W\}, \tag{C.31}$$

which, in view of (C.30), is the set words of length $W + \eta_n$ that may transition into $\boldsymbol{b}_1(n')$ through one single mutation. Since $\boldsymbol{x}$ must have at least one subsequence of length $W + \eta_n$ from $\mathcal{B}_{1\boldsymbol{\eta}}$, and since a mutation in $\boldsymbol{x}$ changes a letter to the corresponding targeted letter in $\boldsymbol{y}$ with probability 1/3, for the Jukes-Cantor model (2), it follows that

$$S_{\boldsymbol{\eta}\boldsymbol{x}} = \frac{1}{3} \sum_{\boldsymbol{b}\in\mathcal{B}_{1\boldsymbol{\eta}}} H_{\{\boldsymbol{b}\}}(\boldsymbol{x}), \tag{C.32}$$

where $0 \leqslant H_{\{\boldsymbol{b}\}}(\boldsymbol{x}) \leqslant L_0 - W + 1 - \eta_n = L_{\eta_n}$ is the number of copies of $\boldsymbol{b}$ along $\boldsymbol{x}$. Insertion of (C.32) into (C.29), and arguing as in (C.12), we find that

$$\begin{aligned} S_{\boldsymbol{\eta}} &= \tfrac{1}{3} \sum_{\boldsymbol{x}\in\mathcal{X}_1} \pi_{\boldsymbol{x}} \sum_{\boldsymbol{b}\in\mathcal{B}_{1\boldsymbol{\eta}}} H_{\{\boldsymbol{b}\}}(\boldsymbol{x}) \\ &= \tfrac{1}{3} \pi(\mathcal{X}_1) \sum_{\boldsymbol{b}\in\mathcal{B}_{1\boldsymbol{\eta}}} E\big[H_{\{\boldsymbol{b}\}}(\boldsymbol{X}_0) | H_{\mathcal{B}_1}(\boldsymbol{X}_0) > 0\big] \\ &\approx \tfrac{1}{3} \pi(\mathcal{X}_1) \sum_{\boldsymbol{b}\in\mathcal{B}_{1\boldsymbol{\eta}}} E\big[H_{\{\boldsymbol{b}\}}(\boldsymbol{X}_0)\big] \\ &= \tfrac{1}{3} \pi(\mathcal{X}_1) \sum_{\boldsymbol{b}\in\mathcal{B}_{1\boldsymbol{\eta}}} L_{\eta_n} 4^{-(W+\eta_n)} \\ &= \pi(\mathcal{X}_1) L_{\eta_n} (W - \eta_n) 4^{-(W+\eta_n)}, \end{aligned} \tag{C.33}$$

where in the second step we introduced $\mathcal{B}_1$ of (80), the set of all $4^W - 1$ words of length W except $\boldsymbol{b}_1$, and in the third step we used the approximation $P(H_{\mathcal{B}_1}(\boldsymbol{X}_0) > 0) \approx 1$ for the probability of $\boldsymbol{X}_0$ having at least one subsequence different from $\boldsymbol{b}_1$. This probability is exactly 1 when $W_0 > 1$, whereas it equals $1 - 4^{-L_0}$ when $W_0 = 1$. In the fourth step of (C.33) we used the fact that all letters are independent with uniform frequencies (3) for the Jukes-Cantor model, and therefore the probability is $4^{-(W+\eta_n)}$ for a sequence of length $W + \eta_n$ to occur at a specific location. In addition, there are L_{η_n} ways of choosing the leftmost locus of such a sequence. Finally, in the last step of (C.33) we used that there are $W - \eta_n$ ways of choosing a position to mutate in (C.31), and for each such position there are three possible letters that might mutate into the targeted letter of $\boldsymbol{b}_1(n)$. Insertion of (C.33) into (C.28) yields

$$S = \pi(\mathcal{X}_1) \sum_{k=0}^{n_{\max}-1} L_{kW_0} (W - kW_0) 4^{-(W+kW_0)} R_k, \tag{C.34}$$

where

$$R_k = \sum_{\substack{\boldsymbol{\eta}\in\mathcal{L} \\ \eta_n = kW_0}} (-1)^{n-1} = \begin{cases} 1, & k = 0, \\ -1, & k = 1, \\ 0, & k = 2, \ldots, n_{\max} - 1. \end{cases} \tag{C.35}$$

By substituting (C.35) into (C.34) and (C.26), we finalize the proof of (B.11).

Appendix D. Waiting time distributions with multiple targets

In this appendix we will analyze the waiting time when several targets are possible, as described in the last paragraph of Section 11. More specifically, suppose there is a pool of $M \geqslant m$ genes, all of which are part of at least one target of m genes. Let $\underline{\boldsymbol{X}}_t$ be the consensus regulatory array of dimension $M \times L$, consisting of regulatory regions of length L from all these M genes. As before, we decompose the set of regulatory arrays into a smaller number of components, and let $\boldsymbol{C}_t$ be the consensus array component at time t to which $\underline{\boldsymbol{X}}_t$ belongs. Let $\mathcal{J}_{Mm}$ refer to all subsets $J \subset \{1, \ldots, M\}$ of size $|J| = m$ that correspond to a collection of m genes whose changed expression constitute a global target. For each $J \in \mathcal{J}_{Mm}$, write $g(\boldsymbol{C}_t; J)$ for the number of local targets within J that have been reached at time t. Then

$$G(\boldsymbol{C}_t; \mathcal{J}_{Mm}) = \max\{g(\boldsymbol{C}_t; J); J \in \mathcal{J}_{Mm}\}$$

is the maximum number of local targets that have been reached at time t, among all subsets of m genes in $\mathcal{J}_{Mm}$. The sought for waiting time

$$T_{Mm} = T_{Mm}(\mathcal{J}_{Mm}) = \min\{t \geqslant 0; G(\boldsymbol{C}_t; \mathcal{J}_{Mm}) \geqslant m\} \tag{D.1}$$

generalizes the definition in (53) from $M = m$ to $M \geqslant m$. If we assume that $\boldsymbol{C}_t$ is a continuous time Markov process, then T_{Mm} will have a phase-type distribution, as in (57). In general, the larger $\mathcal{J}_{Mm}$ is, the shorter is the waiting time. In order to illustrate this, we will consider two extremes scenarios for fixed M and m.

The first option is to make $\mathcal{J}_{Mm}$ as small as possible, given a constraint that each gene belongs to at least one $J \in \mathcal{J}_{Mm}$. Assuming that M/m is an integer, this corresponds to a scenario where each gene is part of one global target only, so that $\mathcal{J}_{Mm}$ is a collection of M/m disjoint subsets of size m. In particular, if the waiting time distribution is the same $(= F_{T_m})$ for all these M/m subsets of genes, it follows that

$$F_{T_{Mm}}(t) = 1 - (1 - F_{T_m}(t))^{M/m}.$$

In particular, if F_{T_m} is exponential, we have that

$$E(T_{Mm}) = \frac{m}{M} \cdot E(T_m).$$

The second option is to make $\mathcal{J}_{Mm}$ as large as possible, the collection of all $\binom{M}{m}$ subsets of M of size m. In order to quantify how this impacts the expected waiting time, assume to begin with a neutral selection model, K targeted binding sites per gene and an arbitrary order of target appearance among the m genes of any $J \in \mathcal{J}_{Mm}$. For neutral models it suffices to use a coarse array decomposition (31) for which there are $C = 2$ intervals $\Delta_0 = \{0, \ldots, \delta_{\max}\}$ and $\Delta_1 = \{\delta_{\max} + 1, \ldots, W\}$ per gene (see Appendix B.1 for a detailed motivation). As in Section 10.2.1, it is possible then to reduce the state space of $\boldsymbol{C}_t$ from size $C^M = 2^M$ to size $M + 1$, where $\boldsymbol{C}_t \in \{0, \ldots, M\}$ is a birth–death process that records how many genes, among the pool of M genes, that have not yet reached their local targets. In particular, T_{Mm} is the time point when $\boldsymbol{C}_t$ reaches one of the absorbing states $\{0, \ldots, M - m\}$. The birth–death nature of $\boldsymbol{C}_t$ makes it possible to derive an explicit formula for $E(T_{Mm})$. More specifically, using the same method of proof as on pages 303–305 of Hössjer et al. (2018), it can be shown that formula (10) generalizes from $m = M$ to $m < M$ as

$$\begin{aligned} E(T_{Mm}) \quad = \quad & \lambda_{10}^{-1} \sum_{c=M-m+1}^{M} \binom{M}{c} (1 - \kappa_1)^{M-c} \kappa_1^c \\ & \cdot \sum_{h=M-c}^{m-1} \sum_{k=0}^{h} \frac{\binom{M-1}{k}}{(M-k)\binom{M-1}{h}} \cdot r^{h-k}, \end{aligned} \tag{D.2}$$

where λ_{10} is the forward rate at which a binding site is acquired at each gene, $r = \lambda_{01}/\lambda_{10}$ is the ratio of the rates at which a binding site at each gene is lost and acquired, whereas $\kappa_1 = P(H_0 = 0)$ is the probability that each regulatory sequence has not yet acquired its local target at time point $t = 0$, with $H_0 \sim \mathrm{Po}\left(L_0 K 4^{-W} \sum_{\delta=0}^{\delta_{\max}} \binom{W}{\delta} 3^{\delta}\right)$ the number of substrings of a regulatory sequence at time $t = 0$ that match one of its K possible local targets, when these are widely dispersed, cf. (A.19).

Table 16
Expected waiting time $E(T_{Mm})$ until the first m of M possible targets have appeared, when any subset of m genes (among all $\binom{M}{m}$ possible subsets) constitutes a valid target of m = 5 coordinated mutations. The regulatory sequence of each gene is $L = 1000$ nucleotides long, there are $K = 3$ possible (and widely spread) binding sites per gene of length $W = 10$, and no mismatches ($\delta_{\max} = 0$) are allowed. The order of target appearance is arbitrary, and $C = 3$ distance-based intervals per gene are used for the array component process $\boldsymbol{C}_t$, with a state space reduction (97) that ignores double mutations at different genes. Two selection models are considered; a neutral model $s_h \equiv 1$ and a "valley model" with a selective disadvantage of the intermediate steps ($s_0 = s_m = s_{m+1} = \ldots = s_M = 1, s_1 = \ldots = s_{m-1} = 0.9999$). Scenarios with our without back mutations ($\gamma = 1$ and $\gamma = 0$) are shown for both selection models.

	Neutral model		Valley model	
M	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
5	$8.32 \cdot 10^9$	$1.11 \cdot 10^{19}$	$7.52 \cdot 10^9$	$1.90 \cdot 10^{19}$
10	$2.34 \cdot 10^9$	$4.48 \cdot 10^{16}$	$2.37 \cdot 10^9$	$7.61 \cdot 10^{16}$
15	$1.41 \cdot 10^9$	$3.83 \cdot 10^{15}$	$1.45 \cdot 10^9$	$6.46 \cdot 10^{15}$
20	$1.01 \cdot 10^9$	$7.55 \cdot 10^{14}$	$1.05 \cdot 10^9$	$1.26 \cdot 10^{15}$
25	$7.85 \cdot 10^8$	$2.24 \cdot 10^{14}$	$8.18 \cdot 10^8$	$3.73 \cdot 10^{14}$

In order to quantify how much multiple targets decreases the expected waiting time, we will illustrate the second option numerically. As it turns out, if the intermediate steps, before the final target is reached, have lower fitness compared to the original wildtype allele, and/or if back mutations are allowed, then the waiting time until the target is reached will often be very large, with or without adjustment for multiple targets. This is illustrated in Table 16, where a refined version of the expected waiting time (D.2) (based on $C = 3$ intervals per gene rather than $C = 2$) is computed for a system with $m = 5$, and $M = 5, 10, 15, 20, 25$. It can be seen, for this particular example, that although multiple targets decrease the expected waiting time a lot, it is not sufficient to bring it down to small numbers. In addition, one may argue that for biological reasons, (D.2) is unnecessarily restrictive. Indeed, it is often the case that only a few multiple targets (and only a few pathways to each one of them) are possible (Weinreich et al., 2006).

References

Asmussen, S., Nerman, O., Olsson, M., 1996. Fitting phase-type distributions via the EM algorithm. Scand. J. Stat. 23, 419–441.

Bateman, R.M., Crane, P.R., DiMichele, W.A., Kenrick, P.R., Rowe, N.P., Speck, T., Stein, W.E., 1998. Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation. Annu. Rev. Ecol. Syst. 29, 263–292.

Bechly, G., Meyer, S.C., 2017. The fossil record and universal common ancestry. Chapter 10 of Theistic Evolution: A Scientific, Philosophical, and Theological Critique, Moreland, J.P., Meyer, S.C., Shaw, C., Gauger, A.K. and Grudem, W. (Eds)., Crossway Publ., Wheaton, IL, pp. 331–361. .

Behrens, S., Nicaud, C., Nicodéme, P., 2012. An automaton approach for waiting times in DNA evolution. J. Comput. Biol. 19 (5), 550–562.

Behrens, S., Vingron, M., 2010. Studying evolution of promoter sequences: a waiting time problem. J. Comput. Biol. 17 (12), 1591–1606.

Bell, E.A., Boehnke, P., Harrison, T.M., Mao, W.L., 2015. Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. PNAS 112 (47), 14518–14521.

Berg, O., von Hippel, P., 1987. Selection of DNA binding sites by regulatory proteins. J. Mol. Biol. 193, 723–750.

Betancourt, A., 2007. When the going gets tough, beneficial mutations get going. Heredity 99, 359–360.

Carter, A.J.R., Wagner, G.P., 2002. Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model. Proc. R. Soc. London 269, 953–960.

Chan, K., Gordenin, D.A., 2015. Clusters of multiple mutations: Incidence and molecular mechanisms. Annu. Rev. Genet. 49, 243–267.

Chatterjee, K., Pavlogiannis, A., Adlam, B., Nowak, M.A., 2014. The time scale of evolutionary innovation. PLOS Computat. Biol. 10, (9) d1003818.

Chuong, E.B., 2013. Retroviruses facilitate the rapid evolution of the mammalian placenta. Bioessays 35 (10), 853–861.

Churchill, M., Martinez-Caceres, M., de Muizon, C., Mnieckowski, J., Geisler, J.H., 2016. The origin of high-frequency hearing in whales. Curr. Biol. 26 (16), 2144–2149.

Crow, J.F., Kimura, M., 1970. An introduction to population genetics theory. The Blackburn Press, Caldwell.

Doyle, J.A., 2012. Molecular and fossil evidence on the origin of angiosperms. Annu. Rev. Earth Planet. Sci. 40, 301–326.

Durrett, R., 2008. Probability Models for DNA Sequence Evolution. Springer, New York.

Durrett, R., Schmidt, D., 2007. Waiting for regulatory sequences to appear. Ann. Appl. Probab. 17 (1), 1–32.

Durrett, R., Schmidt, D., 2008. Waiting for two mutations: with applications to regulatory sequence evolution and the limits of Darwinian evolution. Genetics 180, 1501–1509.

Durrett, R., Schmidt, D., Schweinsberg, J., 2009. A waiting time problem arising from the study of multi-stage carinogenesis. Ann. Appl. Prob. 19 (2), 676–718.

Erwin, D.H., Valentine, J.W. (Eds.), 2013. The Cambrian Explosion: The Construction of Animal Biodiversity. Roberts and Co., Greenwood Village, CO. .

Ewens, W.J., 2004. Mathematical Population Genetics. I. Theoretical Introduction. Springer, New York.

Fields, D., He, Y., Al-Uzri, A., Stormo, G., 1997. Quantitative specificity of mnt repression. J. Mol. Biol. 271, 178–194.

Fitzgerald, D.M., Rosenberg, S.M., 2019. What is mutation? A chapter in the series: How microbes "jeopardize" the modern synthesis. PLoS Genet. 15, (4) e1007995.

Fraser, G.J., Cerny, R., Soukup, V., Bronner-Fraser, M., Streelman, J.T., 2010. The odontode explosion: the origin of tooth-like structures in vertebrates. Bioessays 32 (9), 808–817.

Gillespie, J.H., 1984. Molecular evolution over the mutational landscape. Evolution 38 (5), 1116–1129.

Goldman, N., Yang, Z., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11 (5), 725–736.

Hecht, J., 2013. Photosynthesis began on earth 3.8 billion years ago. New Scientist 217 (2905), 9.

Hössjer, O., Bechly, G., Gauger, A., 2018. Phase-type distribution approximations of the waiting time until coordinated mutations get fixed in a population. Chapter 12 in Stochastic Processes and Algebraic Structures - From Theory Towards Applications. Volume 1: Stochastic processes and Applications, S. Silvestrov, A. Malyarenko, and M. Rančić (eds.), Springer Proceedings in Mathematics and Statistics, pp. 245–313. .

Iwasa, Y., Michor, F., Nowak, M., 2004. Stochastic tunnels in evolutionary dynamics. Genetics 166, 1571–1579.

Jukes, T.H., Cantor, C., 1969. Evolution of protein molecules. In: Munro, M.N. (Ed.), Mammalian Protein Metabolism. Academic Press, New York, pp. 245–279.

Kauffman, S.A., Levin, S., 1987. Towards a general theory of adaptive walks on rugged landscapes. J. Theor. Biol. 128, 11–45. .

Kemeny, J.G., Snell, J.L., 1976. Finite Markov chains. Springer, New York.

Kimura, M., 1979. Model of effective neutral mutations in which selective constraints is incorporated. Proc. Natl. Acad. Sci. 76, 3440–3444.

Komarova, N.L., Sengupta, A., Nowak, M., 2003. Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. J. Theor. Biol. 223, 433–450.

Labandeira, C.C., 2011. Evidence for an earliest late carboniferous divergence time and the early larval ecology and diversification of major holometabola lineages. Entomologica Americana 117 (1), 9–21.

Lanave, C., Preparata, G., Saccone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20, 86–93.

MacArthur, S., Brockfield, J.F.Y., 2004. Expected rates and modes of evolution of enhancer sequences. Mol. Biol. Evol. 21 (6), 1064–1073.

Marks II, R.J., Dembski, W.A., Ewert, W., 2017. Introduction to Evolutionary Informatics. World Scientific.

Moran, P.A.P., 1958a. Random processes in genetics. Proc. Cambridge Philos. Soc. 54, 60–71.

Moran, P.A.P., 1958b. A general theory of the distribution of gene frequencies I. Overlapping generations. Proc. Roy. Soc. B 149, 102–112.

Neubauer, S., Hublin, J.-J., Gunz, P., 2018. The evolution of modern human brain shape. Sci. Adv. 4 (1), eaao5961.

Neuts, M.F., 1981. Matrix-Geometric Solutions in Stochastic Models: an Algorithmic Approach. John Hopkins University Press, Baltimore, MD.

Nicodéme, P., 2012. Revisiting waiting times in DNA evolution. arXiv:1205.6420v1. .

Orr, H.A., 2010. The population genetics of beneficial mutations. Phil. Trans. R. Soc. B 365, 1195–1201.

Park, T., Fitzgerald, E.M.G., Evans, A.R., 2016. Ultrasonic hearing and echolocation in the earliest toothed whales. Biol. Lett. 12, 20160060.

Paterson, J.R., García-Bellido, D.C., Lee, M.S.Y., Brock, G.A., Jago, J.B., Edgecombe, G.D., 2011. Acute vision in the giant Cambrian predator Anomalocaris and the origin of compound eyes. Nature 480, 237–240.

Phillips, R., Kondev, J., Theriot, J., Garcia, H.G., 2013. Physical Biology of the Cell. Garland Science, New York.

Roberts, R.M., Green, J.A., Schulz, L.C., 2016. The evolution of the placenta. Reproduction 152 (5), R179–R189.

Sanford, J., Baumgardner, J., Brewer, W., Gibson, P., Remine, W., 2007. Mendel's accountant: a biologically realistic forward-time population genetics program. Scalable Computing: Practice and Experience 8 (2), 147–165.

Sanford, J., Brewer, W., Smith, F., Baumgardner, J., 2015. The waiting time problem in a model hominin population. Theor. Biol. Med. Modell. 12, 18.

Sauquet, H. et al., 2017. The ancestral flower of angiosperms and its early diversification. Nat. Commun. 8, 16047.

Scally, A., Durbin, R., 2012. Revising the human mutation rate: implications for understanding human evolution. Nat. Rev. Genet. 13 (10), 745–753.

Shen, B., Dong, L., Xiao, S., Kowalewski, M., 2008. The avalon explosion: evolution of Ediacara Morphospace. Science 319 (5859), 81–84.

Simmons, N.B., Seymour, K.L., Habersetzer, J., Gunnell, G.F., 2008. Primitive early eocene bat from Wyoming and the evolution of flight and echolocation. Nature 451, 818–821.

Specht, C.D., Bartlett, M.E., 2009. Flower evolution: the origin and subsequent diversification of the angiosperm flower. Annu. Rev. Ecol. Evol. Syst. 40, 217–243.

Starr, T.N., Picton, L.K., Thornton, J.W., 2017. Alternative evolutionary histories in the sequence space of an ancient protein. Nature 549, 409–417.

Stone, J.R., Wray, G.A., 2001. Rapid evolution of cis-regulatory sequences via local point mutations. Mol. Biol. Evol. 18, 1764–1770.

Tenesa, A. et al., 2007. Recent human effective population size estimated from linkage disequilibrium. Genome Res. 17 (4), 520–526.

Tkadlec, J., Pavlogiannis, A., Chatterjee, K., Nowak, M.A., 2018. Fixation probability and fixation time in structured populations. arXiv:1810.02687v1 [q-bio.PE]. .

Tuğrul, M., Paixão, T., Barton, N.H., Tkačik, G., 2015. Dynamics of transcription factor analysis. PLOS Genet. 11, (11) e1005639.

Voje, K.L., Starrfeldt, J., Liow, L.H., 2018. Model adequacy and microevolutionary explanations for stasis in the fossils record. Am. Nat. 191 (4), 509–523.

Watterson, G.A., 1964. The application of diffusion theory to two population genetic models of Moran. J. Appl. Prob. 1, 233–246.

Weinreich, D.M., Delaney, N.F., DePristo, M.A., Hartl, D.L., 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312 (5770), 111–114.

Wu, P., Lai, Y.-C., Widelitz, R., Chuong, C.-M., 2018. Comprehensive molecular and cellular studies suggest avian scutate scales are secondarily derived from feathers, and more distant from reptilian scales. Sci. Rep. 8, 16766.

Yang, Z., Jiang, B., McNamara, M.E., Kearns, S.L., Pittman, M., Kaye, T.G., Orr, P.J., Xu, X., Benton, M., 2019. Pterosaur integumentary structures with complex feather-like branching. Nat. Ecol. Evol. 3, 24–30.

Yona, A.H., Alm, E.J., Gore, J., 2017. Random sequences rapidly evolve into de novo promoters. bioRxiv.org.https://doi.org/10.1101/111880.