



HAL
open science

Evolutionary escape from local fitness peaks through inversion mutations

Leonardo Trujillo, Paul Banse, Guillaume Beslon

► **To cite this version:**

Leonardo Trujillo, Paul Banse, Guillaume Beslon. Evolutionary escape from local fitness peaks through inversion mutations. 2021, pp.1-1. hal-03426022

HAL Id: hal-03426022

<https://inria.hal.science/hal-03426022v1>

Submitted on 11 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evolutionary escape from local fitness peaks through inversion mutations

Leonardo Trujillo, Paul Banse & Guillaume Beslon



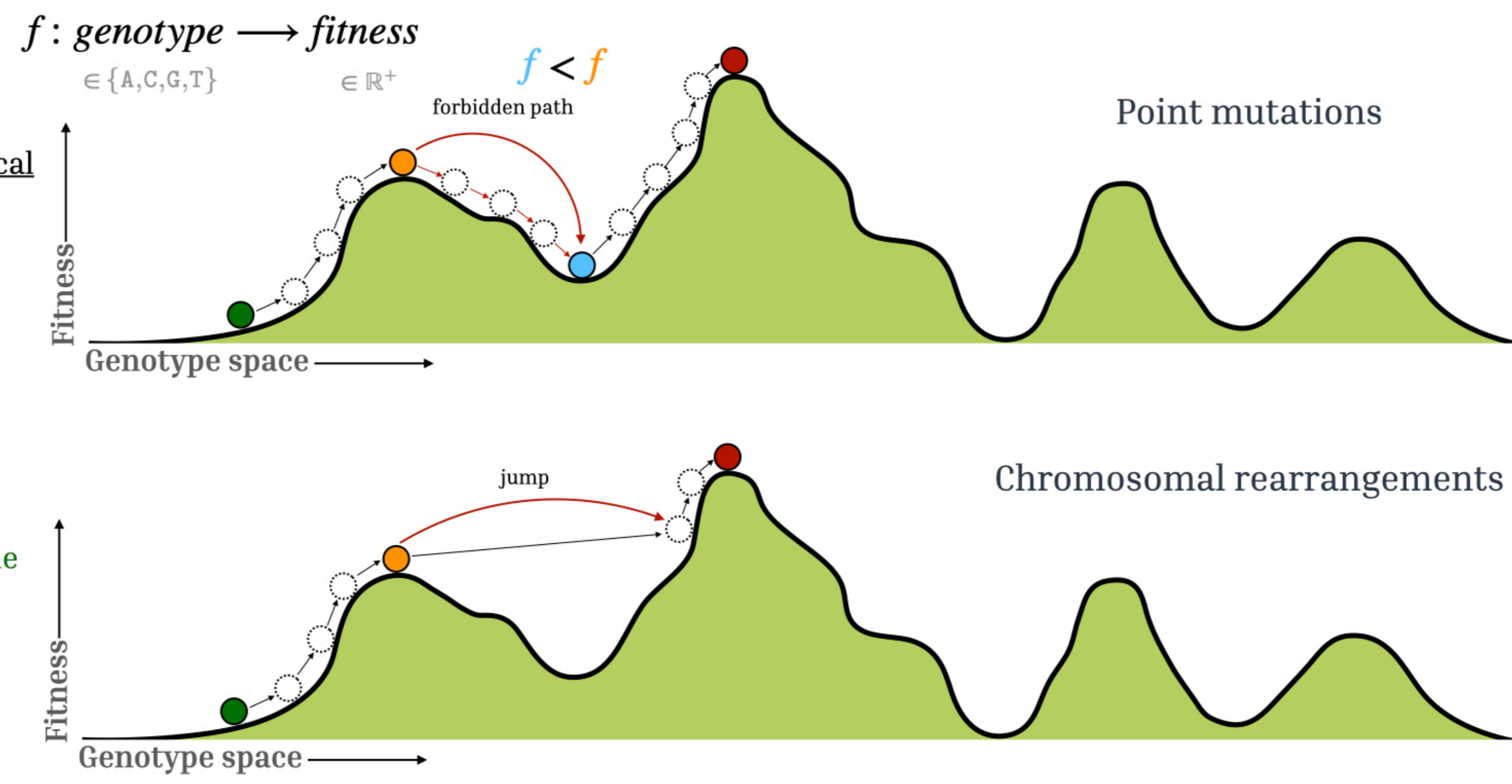
Inria Beagle Team - Artificial Evolution and Computational Biology, Université de Lyon, Inria, ECL, INSA Lyon, Université Claude Bernard Lyon 1, Univ. Lumière Lyon 2, CNRS, LIRIS UMR 5205, F-69622, France
 {leonardo.trujillo; paul.banse; guillaume.beslon}@inria.fr

Model of molecular evolution

Adaptive molecular evolution is often conceptualised as **adaptive walks on rugged fitness landscapes** that are driven by changes in the genome (mutations).

The problem

Once a local fitness peak is reached there are no more selectively accessible pathways by the mechanism of single nucleotides (point) mutations. Then, the evolutionary dynamics get stuck in a suboptimal peak, hindering any possibility to reach other higher peaks.



Proposition

We introduce a simple computational model to simulate the escaping process from a local fitness peak combining point and inversion mutations.

Algorithm 1: Mutate (x, i, j, N)
input: $x \in \{0, 1\}^N, [i, j] \in \{1, \dots, N\}$
 1: $l \leftarrow i$
 2: **repeat**
 3: $x_{temp} \leftarrow (1 - x_i)$
 4: $x_i \leftarrow (1 - x_j)$
 5: $x_j \leftarrow x_{temp}$
 6: $i \leftarrow i + 1 \bmod N$
 7: $j \leftarrow j - 1 \bmod N$
 8: **until** $j = l$
 9: $y \leftarrow x$
 10: **return** $y \in \{0, 1\}^N$

Inversion mutations operation

5' - ATCGATTGAGCTCTAGCC - 3' (A,C,G,T) → {0,1} 011001100011110010
 3' - TACGATTCAGCTCTAGCC - 5' 100110011100001101
 Inspired by molecular biology, these genotypes are **double-strained** binary vectors of length N

$$x := (x_1, x_2, \dots, x_N, \dots, x_N, \dots, x_1)$$

where the **complementary** vector is defined such that $\bar{x}_i = 1 - x_i$

01100110001110010 → 011000011100110010
 100110011100001101 → 1001111000110011101

We model the **chromosomal-like inversion** as a two step procedure:

$$\hat{I} = \hat{C} \circ \hat{P}$$

1. **Conjugation operation**

$$\hat{C} : (x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_N) \rightarrow (\bar{x}_i, \bar{x}_{i+1}, \dots, \bar{x}_j, x_j, \dots, x_{i-1}, x_i, \dots, x_1)$$

2. **Permutation operation:**

$$\hat{P} : (x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_N) \rightarrow (x_j, x_{j+1}, \dots, x_{i-1}, x_i, \dots, x_1)$$

(point-mutations when: $i=j$)

01100110001110010 $\xrightarrow{\text{switch}}$ 011001100110010
 100110011100001101 $\xrightarrow{\text{switch}}$ 100110010100001101

The model

We simulate artificial molecular evolution of digital organisms (replicators), each of which contains a single piece of DNA. We engineer a computational method to cartoon the double-stranded structure of DNA, and simulate inversion-like mutations: a permutation of the complementary strand, which are then exchanged with the main strand within the segment where the inversion occurs. We model digital genotypes made up of a binary alphabet {0, 1}. In analogy with DNA primary sequence, the bits represent (artificial binary) nucleotides. We also assume that genomes are arranged in circular strings with constant number of base-pairs. In an abstract sense, our model does not consider non-coding regions, mimicking the molecular evolution of some viruses and mitochondrial DNA with compact genomes and closed double-stranded DNA circles. It is important to emphasise that our computational model simulates intragenic-like mutations. We do not include recombination, so we are modelling asexual replication.

Algorithm 2: AdaptiveWalk

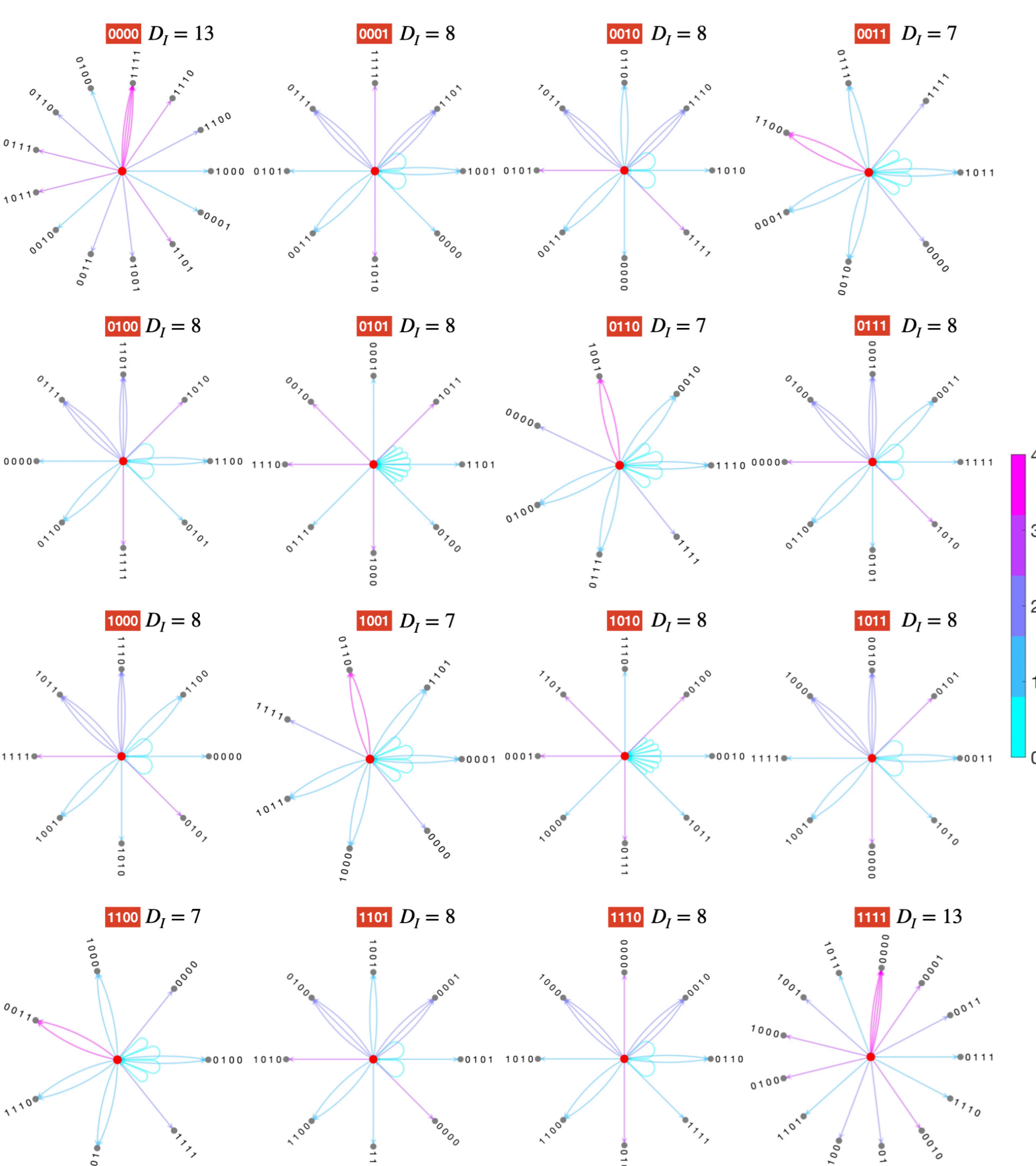
Algorithm 2: AdaptiveWalk (x, p, N, τ_{max})
input: $x \in \{0, 1\}^N, \tau_{max}, p \in (0, 1)$
 1: $t \leftarrow 0$
 2: $f_0 \leftarrow f(x)$
 3: **repeat**
 4: **if** $p = 1$ **then**
 5: $i, j \leftarrow \text{PRNG}$
 6: **else**
 7: $i \leftarrow \text{PRNG}$
 8: $j = i$
 9: **end if**
 10: $y \leftarrow \text{Mutate}(x, i, j, N)$
 11: $f' \leftarrow f(y)$
 12: **if** $f_t < f'$ **then**
 13: $x \leftarrow y$
 14: $f_{t+1} \leftarrow f'$
 15: **else**
 16: $f_{t+1} \leftarrow f_t$
 17: **end if**
 18: $t \leftarrow t + 1$
 19: **until** $t = \tau_{max}$
output: $y \in \{0, 1\}^N, f(y) \in [0, 1]$

Enumeration of accessible mutants

Examples of the number of neighbouring sequences D per genotype of size N (subscripts numbers are the occurrence frequencies of D).

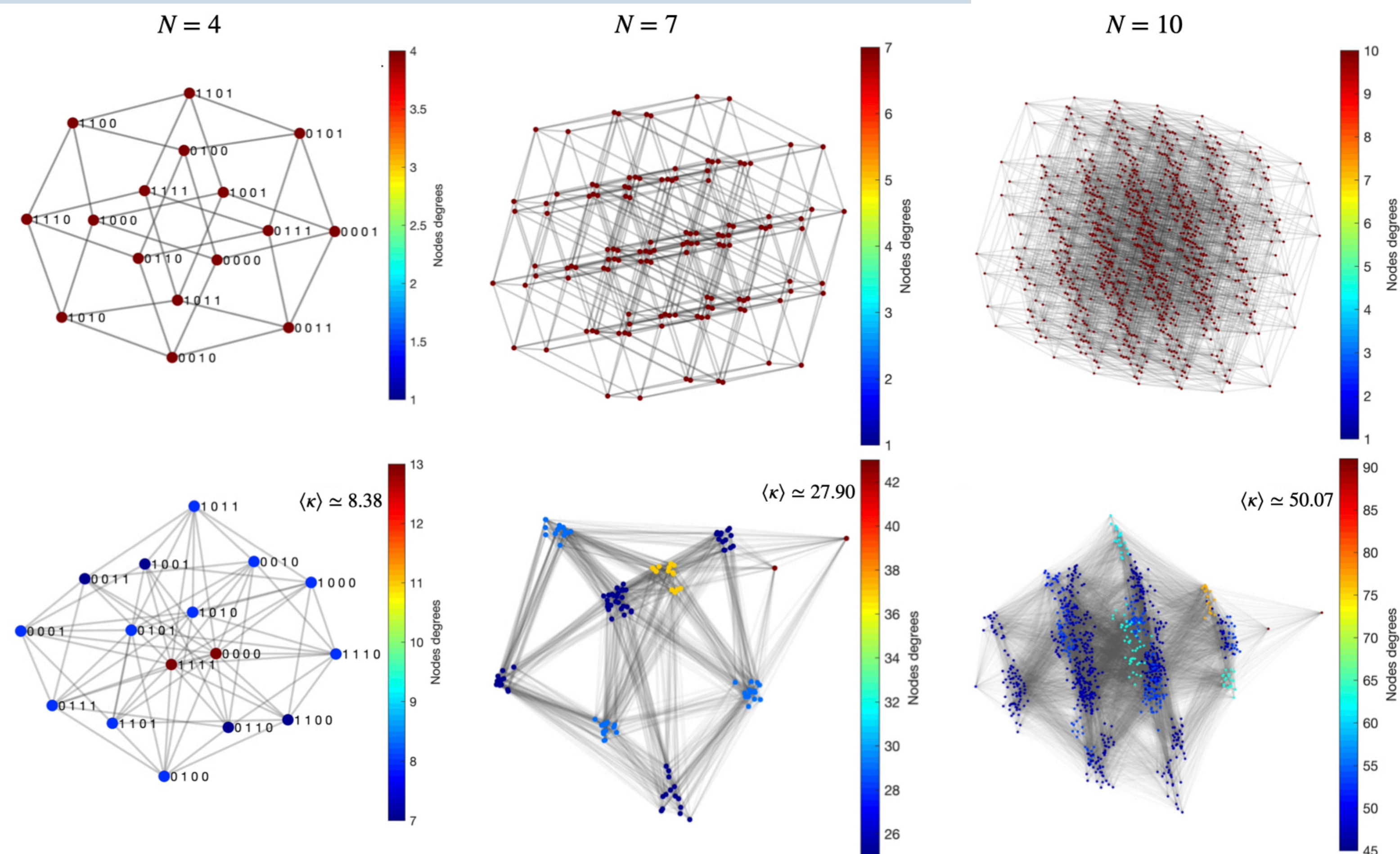
N	D_i	D_P
2	2 ₂ , 3 ₂	2
3	5 ₆ , 7 ₂	3
4	7 ₄ , 8 ₁₀ , 13 ₂	4
5	13 ₂₀ , 17 ₁₀ , 21 ₂	5
6	16 ₃₀ , 17 ₁₈ , 18 ₂ , 22 ₁₂ , 31 ₂	6
7	25 ₇₀ , 29 ₄₂ , 37 ₁₄ , 43 ₂	7
8	28 ₁₆ , 29 ₃₂ , 30 ₁₂ , 32 ₂ , 34 ₈ , 36 ₄₈ , 46 ₁₆ , 57 ₂	8
9	39 ₆ , 40 ₁₈ , 41 ₂₄ , 45 ₁₆₂ , 52 ₃₆ , 53 ₃₆ , 64 ₁₈ , 73 ₂	9
10	45 ₁₀₀ , 46 ₁₅₀ , 47 ₂₀ , 50 ₂ , 52 ₄₀ , 53 ₂₀₀ , 62 ₄₀ , 63 ₅₀ , 77 ₂₀ , 91 ₂	10

Example of accessible mutants (genotypes with N=4)



Accessible-mutants graphs

Representative examples for $N = 4, 7$ and 10 . Colours indicates the nodes degree. The reported values correspond to the mean degree $\langle k \rangle$ and the maximum. The upper graphs show the point mutation case, verifying that the accessible-mutants graphs are isomorphic to the Hamming graphs $H(4, 2)$; $(7, 2)$ and $H(10, 2)$, i.e. the hypercubes $Q(N)$ (the genotype space). The lower graphs correspond to the inversion mutations cases, where we can appreciate that the accessible-mutants graphs are not isomorphic to their respective genotype (hypercube) spaces.



NK rugged landscape

As a model for rugged fitness landscapes, we adopt the well known Kauffman NK model (N is the length of the genome and K is the "epistatic" coupling between nucleotides)

The fitness per bit $f_i : \{0, 1\}^{K+1} \rightarrow [0, 1]$
 Example: $N = 16, K = 2$

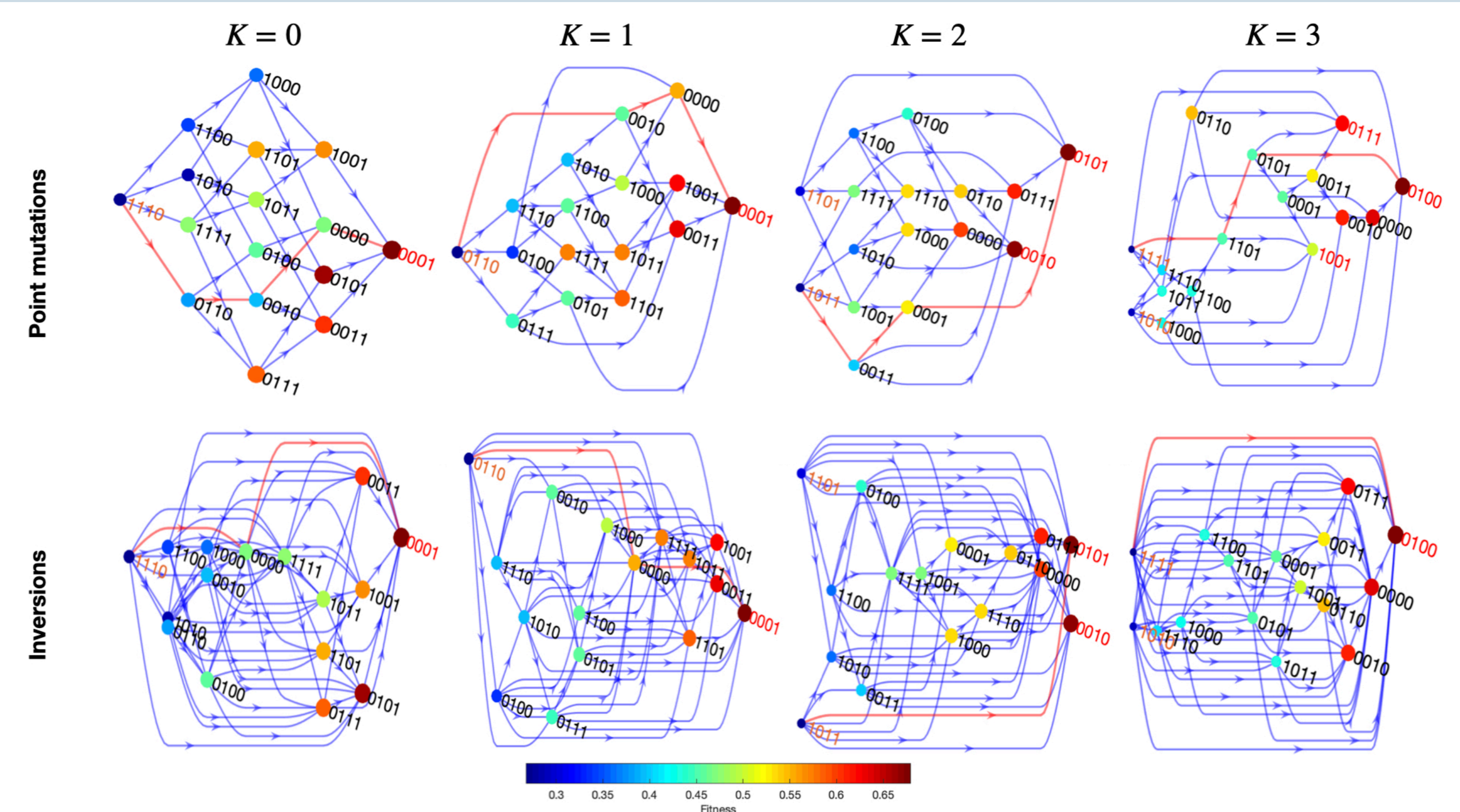
$$f = \frac{1}{N} \sum_{i=1}^N f_i(x_1, x_2, \dots, x_{K+1})$$

Adaptive walks

We assume that on a population scale the organisms are isogenic (i.e. population only contains copies of one genotype) and, therefore, can be simulated as the evolution of one genotype. We consider the simplest case of random adaptive walks, where available genomes with higher fitness are chosen at random. So, the biological significance of our model must be framed in the of strong selection weak mutations limit.

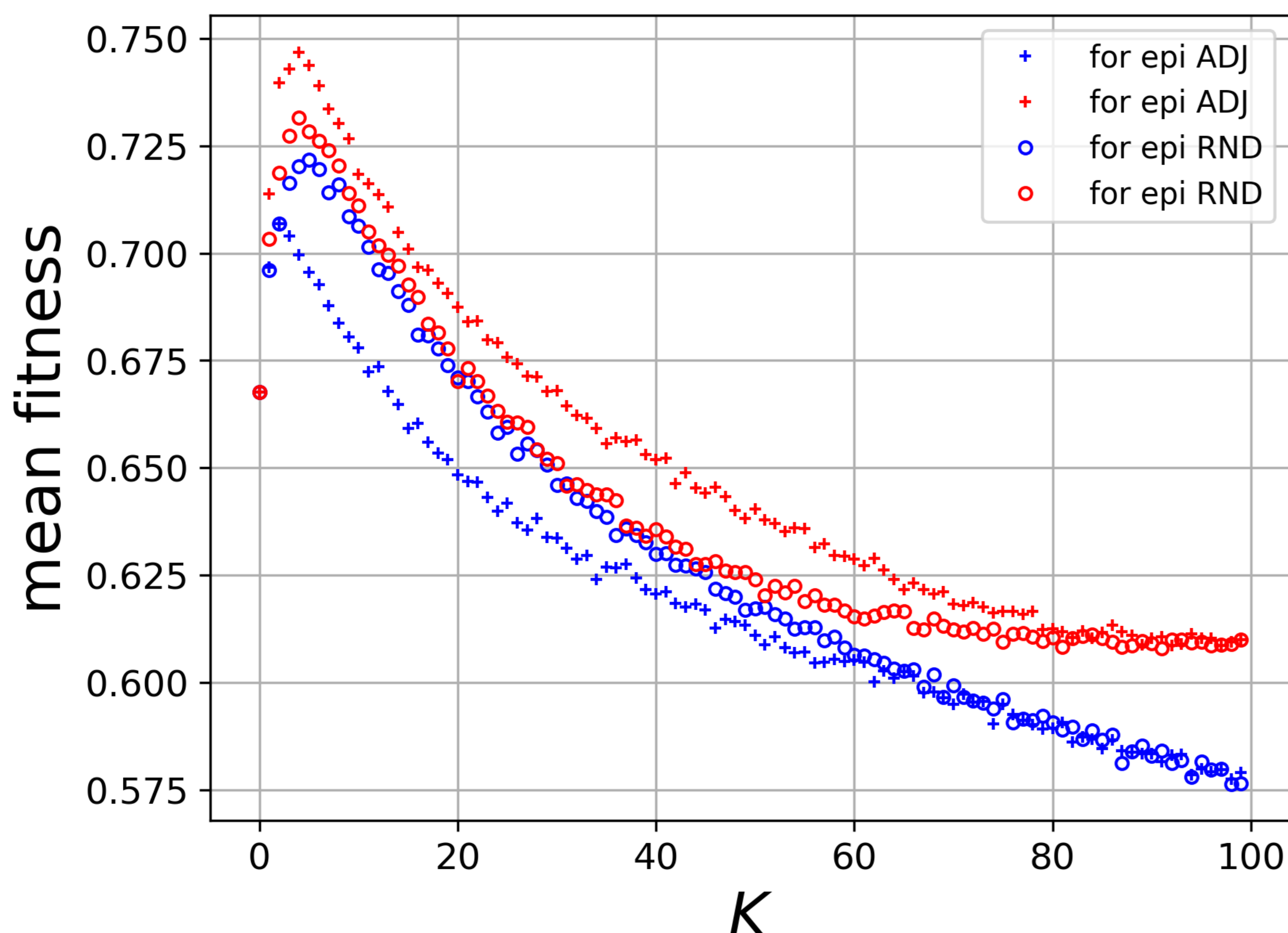
NK fitness networks

Representatives instances of the NK model for $N = 4$ and their fitness networks in layered representation. Node colours correspond to fitness values, increasing from left to right. The shortest paths from lowest to highest fitness genotypes are highlighted in red. The top row shows the cases for point mutations, while the bottom row corresponds to inversion mutations. The landscapes ruggedness go from single peaks $K = 0$, intermediate ruggedness $K = 1, K = 2$, and the full rugged case $K = 3$.



Fitness and epistatic interaction

Behaviour of the mean fitness of genotypes with size $N=100$ for different rugged landscapes, tuned through $K=0, \dots, 99$. For adjacent (+) and random (o) neighbourhoods interactions. Red color corresponds to inversion mutations and blue to point mutations.



Summary

This work provided computational evidence that once an adaptive walk reaches a local fitness peak, then inversion mutations can redirect the evolutionary process towards genotypes with higher fitness values. The way our model was conceived, also proves that escape from a local peak of fitness can occur in constant environments without contingencies. Our model for inversion mutations not only elucidated an escape mechanism, but have also made it possible to uncover interesting aspects about the combinatorics of inversion mutations. We have been able to contrast the isomorphism of mutational graphs for point mutations and their isomorphism with the genotype space. Finally, it was possible to verify that for a very simple model of population genetics in the strong weak mutation selection limit –with rugged landscapes simulated with the well know NK model– our adaptive walks simulations showed that when the evolutionary dynamics is driven by inversions, higher fitness values can be reached.

Bibliography

- Wright S. *The roles of mutation, inbreeding, crossbreeding, and selection in evolution*. In: Proc. Sixth International Congress on Genetics. vol. 1, na; 1932. p. 356.
- Cervera H, Lalic J, Elena SF. *Efficient escape from local optima in a highly rugged fitness landscape by evolving RNA virus populations*. Proc. Royal Society B: 2016;283(1836):20160984.
- Weinreich DM, Chao L. *Rapid evolutionary escape by large populations from local fitness peaks is likely in nature*. Evolution. 2005;59(6):1175.
- Iwasa Y, Michor F, Nowak MA. *Stochastic tunnels in evolutionary dynamics*. Genetics. 2004;166(3):1571.
- Guo Y, Vucelja M, Amir A. *Stochastic tunneling across fitness valleys can give rise to a logarithmic long-term fitness trajectory*. Science Advances. 2019;5(7):3842.
- Gillespie JH. *Molecular evolution over the mutational landscape*. Evolution. 1984;38(5):1116.
- Kauffman SA. *The origins of order: Self-organization and selection in evolution*. Oxford University Press, USA; 1993.
- Trujillo L, Banse P, Beslon G. *Simulating short- and long-term evolutionary dynamics on rugged landscapes*. The ALIFE conference 2021 (arXiv:2105.05520).

