In Silico Evolution of Multi-scale Microbial Systems in the Presence of Mobile Genetic Elements and Horizontal Gene Transfer

Vadim Mozhayskiy and Ilias Tagkopoulos[*]

Department of Computer Science and Genome Center
University of California Davis
One Shields Avenue, Davis, CA 95616, USA
`itagkopoulos@ucdavis.edu`

Abstract. Recent phylogenetic studies reveal that Horizontal Gene Transfer (HGT) events are likely ubiquitous in the Tree of Life. However, our knowledge of HGT's role in evolution and biological organization is very limited, mainly due to the difficulty tracing HGT events experimentally, and lack of computational models that can capture its dynamics. Here, we present a novel, multi-scale model of microbial populations with the capacity to study the effect of HGT on complex traits and regulatory network evolution. We describe a parallel load-balancing framework, which was developed to overcome the innate challenges of simulating evolving populations of such magnitude and complexity. Supercomputer simulations of *in silico* cells that mutate, compete, and evolve, show that HGT can significantly accelerate, but also disrupt, the emergence of advantageous traits in microbial populations. We show that HGT leaves a lasting imprint to gene regulatory networks when it comes to their size and sparsity. In any given experiment, we observed phenotypic variability that can be explained by individual gain and loss of function during evolution. Analysis of the fossil mutational and HGT event record, both for evolved and non-evolved populations, reveals that the distribution of fitness effect for HGT has different characteristics in terms of symmetry, shape and bias from its mutational counterpart. Interestingly, we observed that evolution can be accelerated when populations are exposed in correlated environments of increased complexity, especially in the presence of HGT.

Keywords: Horizontal Gene Transfer, Microbial Evolution, Biological Networks, Simulation, Multi-scale Modeling, High Performance Computing.

1 Introduction

Horizontal Gene Transfer (HGT) is the process of horizontal transfer of genetic material within and across species. It is a mechanism of genetic exchange complementary to vertical transfer, which occurs through cell division and results in the transfer of genetic information from an ancestor to its offspring cells. Although largely ignored in the past, recent phylogenetic evidence suggests that its impact on bacterial evolution is significant and should be investigated more thoroughly [1, 2].
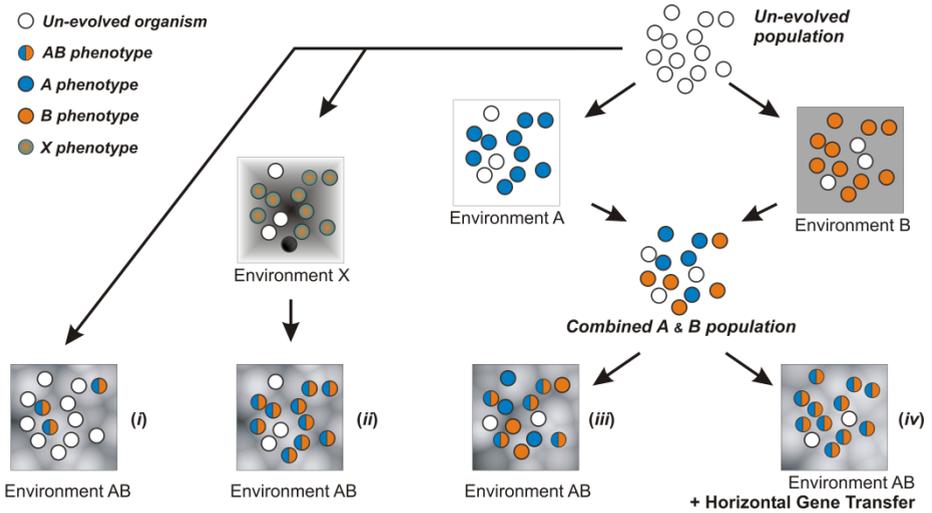
---

[*] Corresponding author.

Fig. 1. General overview of the simulated ecological setting: Microbial populations evolve under environment AB either directly (*i*), or indirectly through an intermediate environment X of lower complexity (*ii*). In addition, we test whether pre-exposing to environments A and B and then merging the respective populations without (*iii*) or with (*iv*) HGT changes the rate and characteristics of evolution.

For instance, it has been estimated that up to a 32% of the bacterial genome is acquired by HGT [3]. However, even this number is a lower bound of the HGT events that take place through bacterial evolution, since only a small fraction of transferred material is positively selected, fixed and, consequently, observable through phylogenetic analysis [4].

Due to our limited ability to observe HGT dynamics in an experimental setting, theoretical models have been traditionally employed to elucidate the impact of HGT on evolution. Continuous kinetic [5, 6] and stochastic models [7-9] were developed to fit experimentally observed short-term dynamics of HGT in twenty-four-hour experiments [9] and to analyze the interplay between rates of HGT, mutations, and selection pressure parameters. It was shown that transferred genes can be successfully fixed in a population if the HGT rate is comparable to the mutational inactivation rate [7] and that high rates of HGT may affect evolution rate in a simple population model [8]. Previous models, although insightful, have a limited scope as they lack any notion of gene regulation, cellular networks and processes, multi-scale structures, and temporal expression dynamics. To address these issues, we extended our previous work [10] to develop a multi-scale simulation framework, capable of simulating the evolution of unicellular organisms in the presence of HGT, which is described in Sec. 2 and 3.

An overview of the general simulation setting discussed in this paper is illustrated in Fig. 1. We start with a random initial population of cells and three dynamic environments, namely A, B and AB, where the latter is the combination of the first two, and hence of higher complexity. The un-evolved initial population is exposed in one of following three settings: (a) it is directly placed into environment AB, (b) it is first placed in environment X, which is of lower complexity, (c) it is initially evolved in

environments A and B, which leads to two distinct populations, that are subsequently randomly sampled (keeping the same effective size) to form a final population that is then placed in the AB environment with and without HGT. This setting allows us to investigate questions related to HGT and evolution in environments that are both correlated and increasingly more complex.

2   Biological Model

In our model, each *in silico* organism encompass functions and parameters that model basic biological phenomena, while its core consists of a gene regulatory and biochemical network with abstract molecular representations. The model has been extended to incorporate Horizontal Gene Transfer in addition to the other cellular (transcription, translation, modification, growth, death, etc.) and evolutionary (mutation and natural selection) processes. In a simulation, a fixed-size population of cells mutates, competes and evolves in well-defined, temporal, multivariate environments. Each cell comprises three types of nodes: Gene/mRNA, Protein, and Modified Protein (Fig. 2a). The Promoter/Gene/RNA node captures gene regulation and transcription, while the Protein and Modified Protein nodes capture translation and post-translational modification (acetylation, phosphorylation, etc.), respectively.
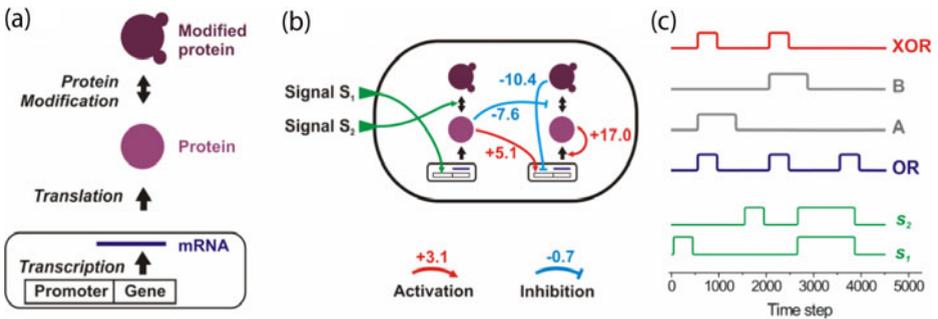

Fig. 2. (a) Basic cellular modeling in our simulation framework; a "triplet" captures processes of transcription, translation, and post-translational modification. (b) Example of a gene regulatory and biochemical network in an organism where environmental signals (e.g. oxygen, temperature, etc.) regulate the expression of certain genes/proteins. (c) Environmental signals (s_1 and s_2) and nutrient abundance for four environments (bottom to top: OR, A, B, XOR) shown as a function of time steps within one epoch. Nutrient presence is a delayed function of the two signals.

A "triplet" consists of a specific gene node and its products, i.e. the corresponding protein and modified protein node, and generally captures the "central dogma" of molecular biology (Fig. 2a). Each organism has its own distinct gene regulatory and biochemical network (i.e. a collection of various triplets and weighted regulatory edges) that can be depicted as a directed weighted graph (see Fig. 2b). There exists a set of "special triplets", which are common in all cells, and encode physiological responses. It is important to note that we do not impose any objective function or arbitrary

selection. Instead, we model the environment in which synthetic organisms live and evolve, which consists of signals, nutrients and other chemicals (e.g. toxic compounds) with concentrations that can fluctuate over time. In this work, every environment has only one nutrient type and each organism possess one special triplet, whose expression allows the organism to metabolize the nutrients that are present. Since nutrients are present for a short duration, organisms that evolve the capacity to infer their presence and be prepared (e.g. express the metabolic triplet) have a selective advantage, in analogy to real microbial systems. We utilize this framework to address questions regarding the impact of HGT on trait evolution, and gene regulatory network organization.

In the setting discussed here, two signals s_1 and s_2 carry information regarding the presence of nutrients in the environment (Fig. 2c). For example, in a XOR environment, the "Nutrients Presence [XOR]" = Delayed (s_1 XOR s_2). Similarly, the correlation-structure of environments A and B is "Nutrients Presence [A]" = Delayed (s_1 AND NOT(s_2)) and "Nutrients Presence [B]" = Delayed (NOT(s_1) AND s_2) respectively. Despite the fact that the combined AB environment (delayed XOR) is a simple combination of the A and B environments, its complexity is significantly higher when compared to the other two (A and B) as it is not linearly separable [11]; in contrast to both A and B environments that can be separated linearly. To assess the fitness level of each organism, we report the Pearson correlation between nutrient abundance and response protein expression level over a predefined interval of time, which we call an "epoch" (4,500 time units in our simulations). We stress that this similarity measure is used for visualization purposes as a proxy to each organism's fitness, and at no point participates or interferes with the selection or evolutionary trajectory of cells during the simulation.

The probability of molecule creation at each node and at each time step is a function of the regulatory effect of other nodes on that specific node, and the availability of substrate molecules. We model the molecule production probability as a two-level sigmoid function that captures saturation effects for any given regulator and for the expression of any given node. As such, the molecule production probability of node i is given by:

$$G_i = basal_i + (1 - basal_i) \cdot tanh\left(\frac{\sum_{j=1}^{n}\left(w_{ij} \cdot f_{ij}\left(v_j, \tilde{m}_{ij}, \tilde{s}_{ij}\right)\right) - m_i}{s_i}\right), \qquad (1)$$

where the sigmoid function f_{ij} describes the regulatory effect of node j on node i:

$$f_{ij}\left(v_j, \tilde{m}_{ij}, \tilde{s}_{ij}\right) = \frac{1}{2} \cdot \left[1 + tanh\left(\frac{v_j - \tilde{m}_{ij}}{\tilde{s}_{ij}}\right)\right], \qquad (2)$$

where w_{ij} is the regulatory matrix element (i.e. the strength and direction that exerts node j to node i), v_j is the value of node j, m_i and s_i the midpoint and slope of the target-specific sigmoid function, $\tilde{m}_{ij}$ and $\tilde{s}_{ij}$ the midpoint and slope of the regulator specific sigmoid function, n is number of regulating nodes, $basal_i$ is the basal expression parameter.

3   Parallel Simulation Framework

The simulations described here are of unprecedented scale and scope, with integrated models of the environment, population, organism, biological network and molecular species. This level of detail is necessary in order to model phenomena that transcend multiple scales, as in the case of Horizontal Gene Transfer. We had to develop efficient algorithms for HPC communication, balancing and process migration, as cell death and division creates unforeseen loads to the various computational cores. In addition, as organisms adapt and evolve, the complexity of their internal networks constantly increases, and with that the need for computational power. Cells with larger networks can be more efficient in nutrients metabolism and therefore grow and divide faster in real time. On the contrary, the computational time for cells with extended genomes is always larger, and scales with $O(N^2)$, where N is the number of nodes within the cellular network. This calls for a synchronization point at each time-point during our simulations, which may lead to poor scalability due to load imbalance (Fig. 3).

   Initially, cells were distributed to MPI processes with one cell per process per computational core; MPI processes were synchronized at the end of each time step. However, in this initial implementation the imbalance was a problem even for a small number of cells, and the code did not scale beyond 64 cores. The model was improved when a group of cells were assigned to each MPI process, because of averaging effects (i.e. the average computational load was similar among processes). Strong scaling results (Fig. 4a) showed that for our problem size, a load of 8 cells per core was ideal as the imbalance between processes was minimal.
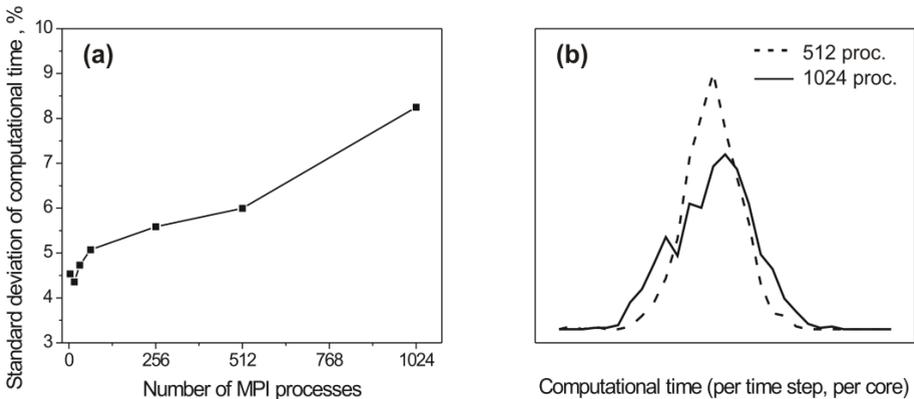

Fig. 3. Variation of the computational time per MPI process increases with the number processes. (a) Standard deviation of the computational time (per time step, per core) five epochs after cell distribution was balanced, at a constant load of 8 cells/process. (b) Distributions of computational time across MPI processes for jobs with 512 and 1024 processes (4096 and 8192 cells, respectively). In larger populations, the higher variance between cell size results in unbalanced computational loads and increased idling time at synchronization points.
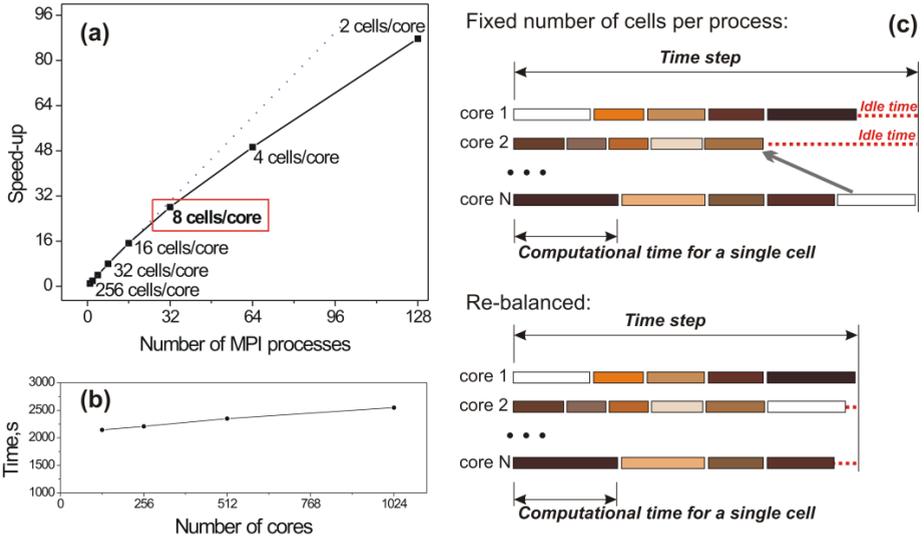
Fig. 4. (a) Strong scaling for a population 256 cells. Code scales well for loads of 8 cells/core or more. (b) Weak scaling up to 8192 cells for the hybrid MPI/OpenMP model. (c) Dynamic MPI load balancing. Top: in an evolving population, computational time for cells varies with the cell network size. This results in idle cores (dashed lines) with smaller (i.e. fast-to-compute) cells, as they synchronize at each time-point. Each bar segment depicts the computational time of a single cell, and multiple cells, of various complexities, are assigned to a single core. The maximum of these loads (here, the load of core N) defines the speed of the simulation. Bottom: with the addition of a dynamic load balancer, cells are redistributed to minimize idling time.

Next, we further extended our model by implementing a hybrid MPI/OpenMP solution: each MPI process is executed on a multi-core computational node; cells assigned to each MPI process are stored in the node's shared memory; computational cycles for each cell update in an MPI process are dynamically distributed between available cores. Dynamic cell distribution is carried out by creating a pool of cycles and cores, and aims to eliminate the idling time during communication. Weak and strong scaling shows scalability up to 8192 cells with near-linear speedup (Fig. 4b).

Finally, by adding dynamic MPI load balancing (Fig. 4c), computational time is monitored for each cell and is used to redistribute cells between MPI processes in order to have a more balanced load between cores. The cell growth rate is used to predict cell division/death events. Although the current load-balancing implementation is a distributed process, a scalable hierarchical implementation can further increase the performance of the simulator.

4   Application: Horizontal Gene Transfer

In our model, genes and its products are represented by triplets, and therefore HGT can be treated as inter-cellular transfer of one or more triplets (mobile genetic elements). For every HGT event a random fragment (i.e. subset of triplets) is copied from a donor cell

and inserted into the regulatory network of a recipient cell. Fragment size is chosen using a probability density function (normalized sigmoid function):

$$P(n) = \frac{1-tanh\left(\frac{n-m}{s}\right)}{m\cdot\left(2+ln\left(e^{-\frac{2m}{s}}+1\right)\right)},$$ (3)

where n is the fragment size m and s are the middle point and slope of the probability density function, respectively. In most cases $s=m$ was used, and therefore 67% of all transferred fragments were not larger than m triplets (Fig. 5a). The original regulation of the response protein by the transferred sub-network is preserved (Fig. 5a, insert). Experimentally observed HGT rates between bacteria in natural environments vary between 10^{-7} and 10^{-11} per generation per cell [12-14], while in some cases the rate climbs between 10^{-3} to 10^{-1} [14, 15]. Default HGT rate used throughout the paper is $5\cdot10^{-6}$ per cell per time step.
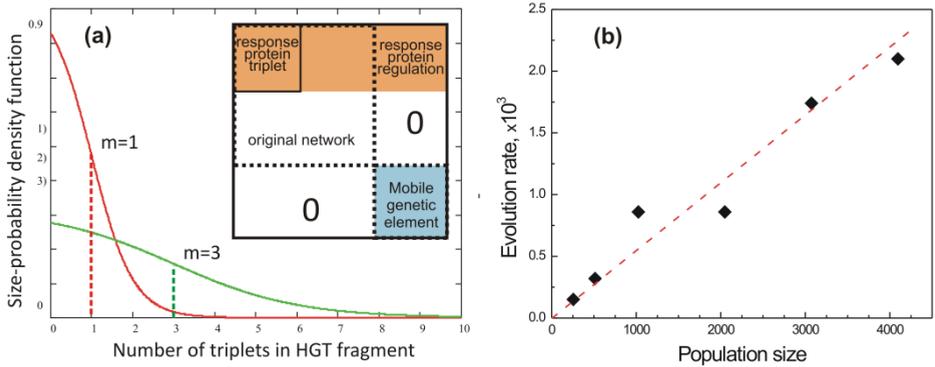

Fig. 5. (a) HGT transfer probability and genome integration: Probability density function profile used to select fragment sizes; (insert) incorporation of the transferred fragment into the regulatory matrix, where only the response pathway regulation is conserved. (b) Evolution rate is a linear function of the population size. Rate is calculated as an average slope of the maximum fitness increase for population sizes up to 4096 cells (16 replicates). Initial random populations evolved in the XOR environment until the maximum fitness is stabilized.

First, we looked at how the rate of evolution scales as a function of the population size. It is believed that evolution speed increases linearly with population size N for small populations, and with ln(N) for intermediate population sizes [16], while it approaches a saturation limit for large populations (>10^9) [17] . We observed a linear dependence of evolution rate to population size (Fig. 5b) in agreement with theoretical predictions. Although it is possible that HGT can eventually be beneficial at high population sizes, where simultaneous emergence of competing beneficial mutations may decrease the rate of evolution, we did not observe this effect in our simulations due to the smaller population size.
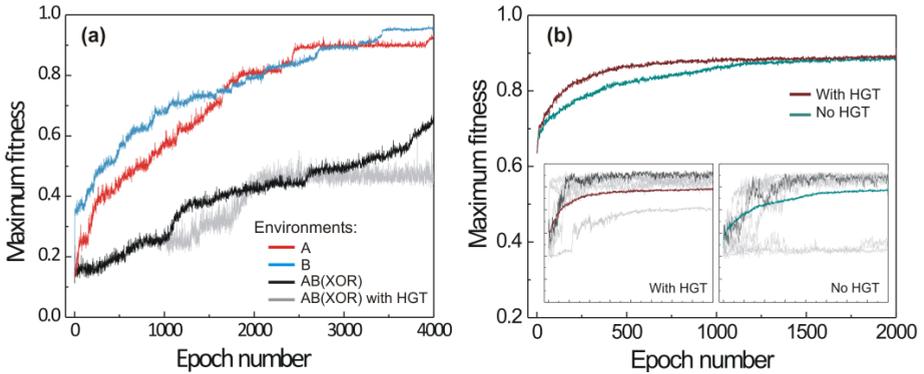
Fig. 6. (Color) Multi-step acceleration and HGT effect in "dual-step" evolution: (a) Evolution of random population of cells in *A*, *B* and *XOR* environments shown in red, blue, and gray respectively. Maximum fitness averaged over 64 simulations, (b) evolutionary trajectory under "dual-step" evolution, where population of evolved cells in A and B environments show remarkably fast adaptation to environment AB (64 simulations). HGT confers an additional acceleration of adaptation to new settings. Inset: Maximum fitness curves for 8 out of 64 individual simulations with (left inset) and without (right inset) HGT are shown in gray. One curve is highlighted with dark gray for clarity.

4.1   Evolution in Coupled Environments of Increasing Complexity

Recent theoretical predictions [18, 19] suggest that evolution generalizes to new environments through facilitated variation, a process in which genetic changes are channeled in useful phenotypic directions. Here we hypothesize that evolution can be accelerated by exposing evolving populations in similar, correlated environments of increasing complexity, and we assess whether HGT further accelerates evolution in such settings. When random populations are exposed directly to environment AB, more than 4,000 epochs are needed to evolve the delayed *XOR* function (Fig. 6, gray curve). In contrast, populations evolve faster in environments of lower complexity, such as the environments A and B (Fig. 6, red/blue lines). Remarkably, if we sample equal amounts of cells from A and B and expose the new population in the complex environment AB with all other parameters being equal (size of population, average nutrient concentration, etc.), XOR phenotypes of high fitness appear surprisingly fast (Fig. 6b). This effect is even more pronounced in the presence of HGT, where the fittest phenotype arises twice as fast as those without HGT present (Fig. 6b, insert). Analysis of individual simulation runs results in similar observations, with all experiments leading to phenotypes of increased fitness in the presence of HGT.

   Detailed statistics of the evolution probability and speed is shown in Table 1. In "single-step" evolution (un-evolved → XOR) only 18 of 32 (56%) experiments were successful and terminated with an evolved *XOR* population (after 4,000 epochs). Success probability of the "dual-step" adaptation process was estimated as a product of "single-step" probabilities and equals 91% and 82% percent with and without HGT, respectively. HGT accelerates emergence of the combined phenotype in {*A*, *B*}

Table 1. Rate of adaptation a complex *XOR* environment in different experimental scenarios. The probability and the speed of phenotype emergence are shown for two fitness thresholds 0.75 (evolved organism) and 0.90 (refined evolved organism). Average speed is the average epoch number at which maximum fitness surpasses the threshold.

| | Emergence of the organism with fitness *w* | | | |
| | *w*>0.75 | | *w*>0.90 | |
	Success Rate	Average speed, *epochs*	Success Rate	Average speed, *epochs*
Un-evolved → *XOR*	18/32	2485	15/32	2489
Un-evolved → *OR*	29/32	1179	13/32	>4,000
OR → *XOR*	30/32	210	5/32	2093
Acceleration by stepwise adaptation		1.8		–
Un-evolved → *A*	30/32	1043	29/32	1067
Un-evolved → *B*	31/32	1217	31/32	1319
{*A & B*} → *XOR*	58/64	234	47/64	448
Acceleration by stepwise adaptation		1.7		1.4
{*A & B*} → *XOR* + **HGT**	64/64	138	48/64	406
Acceleration by HGT		1.7		1.1

mixed populations by a factor of 1.7. However the probability and the speed of the phenotypic refinement for fitness levels above 0.9 is less affected by HGT relative to the initial emergence of the phenotype above the 0.75 threshold (note that any phenotype with 0.75 Pearson correlation between metabolic pathway expression and nutrients exhibits the XOR I/O characteristic). This is to be expected, since subsequent fine-tuning is due to mutations, and not insertion of new functional fragments from other organisms. Evolution through a single environment of intermediate complexity (un-evolved → OR → XOR) accelerates the evolution of a XOR phenotype by a factor 1.8, but with a lower probability of highly fit cells to appear in the final population (only in 5 out of 32 experiments, cells with fitness higher than *w*>0.90 emerged).

4.2   Effect of Horizontal Gene Transfer on the Network Organization

The full gene regulatory and biochemical networks of evolved cells are usually too complex to analyze since many of the connections are not relevant to the observed phenotype. To address this, we employed a reduction algorithm described elsewhere [10] to extract the "minimal" network that encompasses only essential connections. As shown in Table 2, average fitness of reduced minimal networks is at least 95% of the full network's fitness, however the average number of regulatory edges is significantly reduced: from 338 to 14.1 and from 335 to 10.6 with and without HGT respectively. Presence of HGT events results in larger networks that are considerably more sparse (0.39 *vs.* 0.22), but with the same average sparsity and reduced network size difference when it comes to their minimal counterparts.

Table 2. Complete and minimal network statistics for populations evolved in a XOR phenotype with and without HGT

	Full network				Minimal network			
	no HGT		HGT		no HGT		HGT	
Fitness *(St. Dev.)*	0.81	*(0.052)*	0.79	*(0.044)*	0.78	*(0.006)*	0.75	*(0.006)*
Triplets	8.8		13.8		5.5		6.7	
Links *(St. Dev.)*	335	*(157)*	338	*(136)*	10.6	*(0.03)*	14.1	*(0.03)*
Sparsity	0.39		0.22		0.11		0.10	
Modularity	3.8		10.1		3.3		3.1	

4.3 Distribution of Fitness Effect of Mutational and HGT Events

Mutations and HGT events differ in magnitude and direction when it comes to their fitness effect. Traditionally, models rely on theoretical or experimentally constructed distributions of fitness effect (DFE) when introducing mutations in a population. For mutations, these distributions have been measured experimentally for viruses and bacteria (e.g. [20-22]) and have also been obtained theoretically (e.g. [23] and references therein). In general, it is assumed that most mutations have a neutral or nearly neutral effect and the vast majority of mutations have a negative fitness effect [23]. In bacteriophage F1, 20% of single point mutations were found to be lethal, while the mean fitness decrease was around 11% [21]. In *E. coli*, the average effect of spontaneous deleterious mutations and random insertions is less than 1% and 3%, respectively [20, 22].

Table 3. Proprieties of Distribution of Fitness Effect for mutational and HGT events in evolved and un-evolved populations

	Mean fitness change, %	Fitness variance	Skewness	Kurtosis	Percent of lethal events
Mutations:					
Un-evolved populations	-6.6%	0.109	-0.014	3.10	33.0 %
Evolved populations	-4.5%	0.032	-1.028	6.39	3.2 %
Horizontal Gene Transfer:					
Un-evolved populations	-5.1%	0.063	-0.129	4.49	14.4 %
Evolved populations	-6.6%	0.048	-0.064	4.70	4.6 %

Here, we use our *in silico* simulation framework to investigate the shape and changes in the DFE for both mutations and HGT. Since each organism has its own regulatory network that results to a distinct phenotypic behavior, we are able to calculate fitness before and immediately after any HGT event by looking at the expression levels of the response pathway. This allows us to profile the shape of DFE along the evolutionary trajectory and to account for genetic drift, which can be a significant force in small populations. In both mutational and HGT DFEs, there is a profound decrease in the number of lethal events (i.e. fitness effect equal to -1) in evolved populations versus the non-evolved populations (Table 3), a clear indication of mutational robustness. Furthermore, we observe a decrease in variance and increase in kurtosis (sharpness) of the DFE as populations evolve, both for HGT and mutations, although the effect is more profound

in the latter case. As population evolves, mutational DFE becomes more skewed towards negative fitness effects, which is to be expected as most mutations in an evolved organism result in a decreased fitness. Interestingly, the DFE of HGT events becomes more symmetric in evolved populations, as the probability for HGT to transfer a beneficial or disrupting fragment increases (the first because of the availability of beneficial sub-networks, the second because of the high ratio of already fine-tuned cells in the population which can be disrupted by a HGT event).

5   Discussion

To elucidate the effect of Horizontal Gene Transfer in bacterial evolution, we used *in silico* microbial organisms that compete and evolve under dynamic environments in the presence of HGT. The simulation framework presented here is the first that incorporates models of cellular and evolutionary processes, together with representations of the environment, population, organism, biological network and molecular species. This allows us to address questions that transcend many levels of biological organization and investigate the impact of phenomena, such as HGT and environmental perturbations in an unprecedented scale, albeit at the price of increased computational complexity.

Our results show that multi-step evolution accelerates the emergence of complex traits, especially in the presence of HGT, and illustrate its effect on adaptation and network organization. We showed that the distribution of fitness effects for HGT presents some notable differences from its mutational counterpart. There are many future directions to explore in order to increase the scope and biological realism of our simulator. The current framework will benefit from the addition of a spatial component, since in the current setting we assumed a well-mixed, homogeneous environment which clearly limits us on the number of hypotheses we can test. This will allow us to investigate individually the various HGT mechanisms, whose effect vary greatly with the spatial landscape of the environment and population structure. Furthermore, the biological realism of the underlying network can be improved by refining the models that capture cellular processes, as well as adding a metabolic layer and its corresponding models. Despite its limitations, this work is ground-breaking by creating an overarching model of biological phenomena, a synthetic environment, where hypotheses can be tested or automatically generated. This, in conjunction with advanced HPC techniques can prove to be transformative in predicting evolution and microbial behavior in general.

References

1. Ragan, M.A., Beiko, R.G.: Lateral genetic transfer: open issues. Philosophical Transactions of the Royal Society B-Biological Sciences 364, 2241–2251 (2009)
2. Boto, L.: Horizontal gene transfer in evolution: facts and challenges. Proc. Biol. Sci. 277, 819–827 (2010)

3. Koonin, E.V., Makarova, K.S., Aravind, L.: Horizontal gene transfer in prokaryotes: Quantification and classification. Annual Review of Microbiology 55, 709–742 (2001)
4. Gogarten, J.P., Doolittle, W.F., Lawrence, J.G.: Prokaryotic evolution in light of gene transfer. Molecular Biology and Evolution 19, 2226–2238 (2002)
5. Koslowski, T., Zehender, F.: Towards a quantitative understanding of horizontal gene transfer: A kinetic model. Journal of Theoretical Biology 237, 23–29 (2005)
6. Nielsen, K.M., Townsend, J.P.: Monitoring and modeling horizontal gene transfer. Nature Biotechnology 22, 1110–1114 (2004)
7. Novozhilov, A.S., Karev, G.P., Koonin, E.V.: Mathematical modeling of evolution of horizontally transferred genes. Molecular Biology and Evolution 22, 1721–1732 (2005)
8. Levin, B.R., Cornejo, O.E.: The Population and Evolutionary Dynamics of Homologous Gene Recombination in Bacteria. PLoS Genetics 5, Article No.: e1000601 (2009)
9. Philipsen, K.R., Christiansen, L.E., Hasman, H., Madsen, H.: Modelling conjugation with stochastic differential equations. Journal of Theoretical Biology 263, 134–142 (2010)
10. Tagkopoulos, I., Liu, Y.C., Tavazoie, S.: Predictive behavior within microbial genetic networks. Science 320, 1313–1317 (2008)
11. Duda, R.O., Hart, P.E.: Pattern classification and scene analysis. Wiley-Interscience, Hoboken (1973)
12. Ando, T., Itakura, S., Uchii, K., Sobue, R., Maeda, S.: Horizontal transfer of non-conjugative plasmid in colony biofilm of Escherichia coli on food-based media. World Journal of Microbiology & Biotechnology 25, 1865–1869 (2009)
13. Baur, B., Hanselmann, K., Schlimme, W., Jenni, B.: Genetic transformation in freshwater: Escherichia coli is able to develop natural competence. Appl. Environ. Microbiol. 62, 3673–3678 (1996)
14. Jiang, S.C., Paul, J.H.: Gene transfer by transduction in the marine environment. Applied and Environmental Microbiology 64, 2780–2787 (1998)
15. McDaniel, L., Young, E., Delaney, J., Ruhnau, F., Ritchie, K., Paul, J.: High Frequency of Horizontal Gene Transfer in the Oceans. Nature 330, 1 (2010)
16. Park, S.C., Simon, D., Krug, J.: The Speed of Evolution in Large Asexual Populations. Journal of Statistical Physics 138, 381–410 (2010)
17. Gerrish, P.J., Lenski, R.E.: The fate of competing beneficial mutations in an asexual population. Genetica 102-103, 127–144 (1998)
18. Kashtan, N., Noor, E., Alon, U.: Varying environments can speed up evolution. Proceedings of the National Academy of Sciences of the United States of America 104, 13711–13716 (2007)
19. Parter, M., Kashtan, N., Alon, U.: Facilitated Variation: How Evolution Learns from Past Environments To Generalize to New Environments. Plos Computational Biology 4 (2008)
20. Elena, S.F., Ekunwe, L., Hajela, N., Oden, S.A., Lenski, R.E.: Distribution of fitness effects caused by random insertion mutations in Escherichia coli. Genetica 102-103, 349–358 (1998)
21. Peris, J.B., Davis, P., Cuevas, J.M., Nebot, M.R., Sanjuan, R.: Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in Bacteriophage f1. Genetics 185, U308–U603 (2010)
22. Kibota, T.T., Lynch, M.: Estimate of the genomic mutation rate deleterious to overall fitness in E-coli. Nature 381, 694–696 (1996)
23. Eyre-Walker, A., Keightley, P.D.: The distribution of fitness effects of new mutations. Nature Reviews Genetics 8, 610–618 (2007)