Integrative Biology

Dynamic Article Links

Cite this: DOI: 10.1039/c2ib20095c

www.rsc.org/ibiology

CRITICAL REVIEW

# Microbial evolution *in vivo* and *in silico:* methods and applications

**Vadim Mozhayskiy and Ilias Tagkopoulos***

***Received 25th April 2012, Accepted 14th September 2012***
**DOI: 10.1039/c2ib20095c**

Downloaded by RSC Internal on 11 December 2012
Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C

Microbial evolution has been extensively studied in the past fifty years, which has lead to seminal discoveries that have shaped our understanding of evolutionary forces and dynamics. It is only recently however, that transformative technologies and computational advances have enabled a larger in-scale and in-depth investigation of the genetic basis and mechanistic underpinnings of evolutionary adaptation. In this review we focus on the strengths and limitations of *in vivo* and *in silico* techniques for studying microbial evolution in the laboratory, and we discuss how these complementary approaches can be integrated in a unifying framework for elucidating microbial evolution.

## Introduction

All life forms, from microbes to higher vertebrates, are constantly under the influence of evolutionary processes that lead to adaptation and phenotypic variation. Whether these forces result in new and rapidly evolving species, as in the case of adaptive radiation, or are responsible for phenotypic divergence within a species, the underlying mechanism by which complex behaviour arises remains the same: gradual accumulation of selected genetic mutations and epigenetic changes gives rise to a myriad of anatomical, physiological and behavioural expressions. Although the notion that evolution, niche adaptation, and phenotypic variation leads to "endless forms most beautiful" can be traced back to Darwin,[1] it was only in the last decades that with the advent of high-throughput sequencing and profiling techniques, we were able to gain a better understanding of the mechanisms by which mutations give rise to novel traits. Remarkably, it has been shown that even single mutations, such as nucleotide polymorphisms, can yield phenotypes that are significantly dissimilar.[2] The same holds for the rewiring of the gene regulatory and biochemical networks, as they were found to exhibit a high degree of evolvability,[3,4] yet preserve phenotypic robustness when under stabilizing selection and in the presence of disrupting mutations.[5,6]

Laboratory evolution of microbial strains has been used with great success to elucidate biological and evolutionary phenomena such as the dynamics of evolutionary processes,[7,8] the emergence of complex traits,[9,10] the role of epistatic interactions during evolution,[11] the reproducibility and the order of mutation fixation,[12] and the quantification of the fundamental evolutionary parameters such as mutation rates and distribution of fitness effects.[13–16] Microbes are the choice of preference for these studies, since they are easy to cultivate, they have short generation times, large population sizes, and small genome sizes. Despite its many advantages, laboratory evolution has many limitations that stem from the nature and the timescale of evolution: in all cases it is time-consuming and laborious, with experiments lasting from months to decades, only to provide data collection for a few thousands of generations. Changes are difficult to observe, record and analyze. Mutations and population structure changes cannot be readily predicted, and results cannot be easily generalized since evolutionary trajectories can differ considerably in replicate experiments.

*Department of Computer Science, UC Davis Genome Center, University of California Davis, Davis, California, 95616, USA. E-mail: itagkopoulos@ucdavis.edu; Fax: +1 (530) 752-4767; Tel: +1 (530) 752-7707*

### Insight, innovation, integration

Microbial evolution has been studied extensively over the past decades due to the importance of microbes to the environment, human health, industrial applications and agriculture. However, it was only in the last decade that advances in high-throughput technologies and computational infrastructure allowed us to investigate the evolutionary processes and mechanistic underpinnings of microbial adaptation in complex environments. This review gives an overview the temporal and spatial scales involved when studying microbial evolution, it summarizes the techniques used for both *in vivo* and *in silico* studies of microbial evolution, it presents the most important findings so far, and finally it discusses the challenges and possible future directions in the field.

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

For these reasons, a number of *in silico* models have been developed that simulate the evolution of "artificial" organisms that live, mutate and compete in structured environments. While these simulations are still simplified and abstract, they have been successfully applied to address questions regarding the interplay between robustness and evolvability,[6,17] the evolution of complexity,[18] the taxis response pathways,[19] the evolution of evolvability,[4,20] and the effect of stochasticity on microbial adaptation.[21]

In this review, we discuss the scales of space and time in which microbial evolution takes place and what their implications are to both computational and experimental methods. Then we present a comprehensive summary of the *in vivo* and *in silico* methods and we review microbial evolution studies that utilize these methods. Finally, we conclude with suggestions on how the two complementary directions can be integrated to provide a unified view of microbial evolution.

## The scale of microbial evolution

Microbial populations enjoy exponential growth, large population numbers and fast division rates. These characteristics, together with a high mutation rate due to the lack of the error correcting mechanisms that is present in eukaryotic cells, result in high evolution rates and a remarkable ability to cope with environmental variation. The continuous evolution of antibiotic resistance in pathogenic strains is a notorious example of the capacity of microbes to adapt in stressful environments. Fig. 1 summarizes the temporal and spatial scales involved in microbial evolution, and the *in vivo* and *in silico* methods discussed in this review.

### Mutation rates in an organism

Mutations can be a substitution of a single nucleotide, also known as a single nucleotide polymorphism (SNP), or insertions and deletions of DNA fragments. The latter are caused by many factors that include the movement of mobile elements within the genome (*e.g.* transposons), horizontal gene transfer, recombination and gene amplifications events. SNPs are the most frequent types of mutations in the microbial world and account for about 70% of mutations observed in the adaptive laboratory evolution (ALE) experiments.[8] In asexual organisms, mutations can be introduced during the DNA replication stage of cell division or in the presence of certain environmental conditions (*e.g.* radiation or mutagenic chemicals).

The mutation rate in microbes is about $10^{-8}$ to $10^{-10}$ mutations per nucleotide per DNA replication (generation)[22,23] and it is generally within two orders of magnitude for most species.[24] In *Escherichia coli*, which has a genome size around $5 \times 10^6$ nucleotides, this translates to a mutation rate of approximately $10^{-2}$ mutations per replication per organism. In other words, on average we expect one mutation in an *E. coli* cell in every hundred generations. This mutation rate is consistent with recent experimental work where it was shown that bacteria accumulate approximately 8–10 mutations for every 1000 generations of adaptation under various selection pressures.[8] Mutation rates can vary significantly around this average, depending on various biotic and abiotic factors in their environment.[25,26] In addition, environmental parameters affect the distribution of fitness effects for the various mutations, and therefore govern which mutations are selected and fixed.

Given the relatively high rates of microbial adaptation, laboratory experiments have a duration that ranges from a few hundred generations[9,12,27,28] to several thousand generations.[10,29–34] A notable exception is the 24-year long evolution experiment which already surpassed 50 000 generations for *E. coli*.[7] Through the work of various groups, we have been able to observe cases of evolutionary rewiring to cope with nutrient limitations,[7,29,30,35] temperature shifts,[31,32,36,37] pH variations,[38] high ethanol concentrations[10,27] and antibiotics.[12] We are still missing clear cases of a *de novo* emergence of molecular pathways or speciation, mainly because of the great number of genome alterations that are needed for such phenomena to be observed.

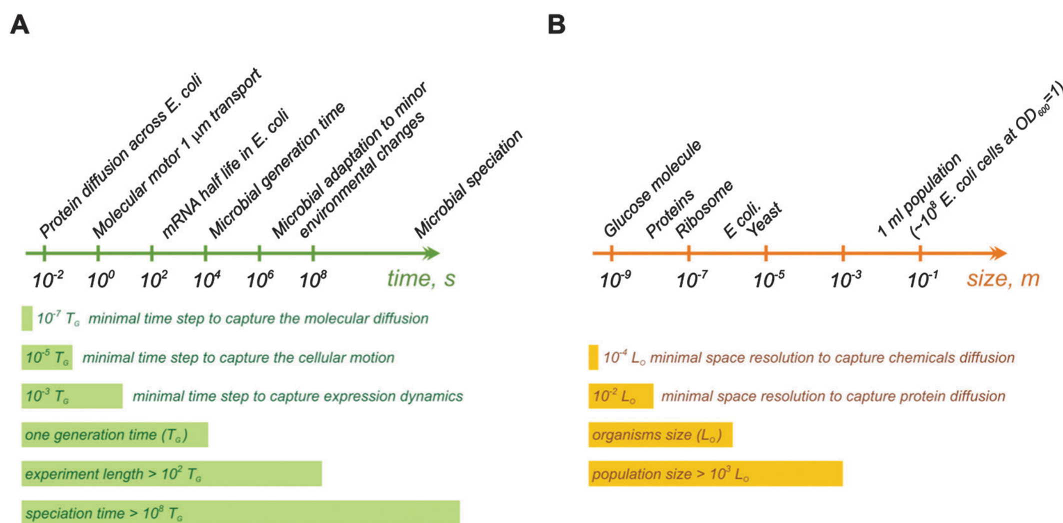


**Fig. 1** Temporal and spatial scales of microbial evolution. A model that tries to capture cellular events and evolutionary forces has to be able to effectively operate in a timescale that varies from milliseconds to years. More information about the relative time and spatial scales in biology can be found in the BioNumbers database.[39,136]

 This journal is © The Royal Society of Chemistry 2012

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

**Mutation rates in a population**

Population size has a significant effect on the rate of adaptation. While each individual organism carries a small number of mutations in its genome, the capacity of microbes to adapt fast in environmental fluctuations stems from having a large population size. A 10 ml flask of *E. coli* culture at $OD_{600} = 1$ contains $10^9$ to $10^{10}$ cells.[39] While the mutation rate is only about $10^{-2}$ mutations per organism per division, the entire population accumulates $\sim 10^7$ mutations at every division (taking into account only SNP mutations). Note that this number of mutations is more than the length of the *E. coli* genome. Therefore, if we assume that random mutations are uniformly distributed in the absence of selection, virtually all possible SNP mutations are always present in that population. Deleterious mutations are quickly removed from the population, but repeatedly reappear at later divisions. Discarded mutations can potentially become beneficial and can be fixed in a population, if the environment fluctuates considerably. Despite the large availability of SNPs that may exist in a bacterial population, this number is still insignificant if we consider the combinatorial explosion of even a handful of SNPs (*e.g.* there are $10^{30}$ possible combinations of any 5 SNP combinations in *E. coli*'s genome).

For small population sizes, the effect of the genetic drift, *i.e.* the change in the frequency of a gene variant in the population due to random sampling and not selection, plays an important role in the rate of mutation fixation. Random drift becomes less important for larger populations, as the probability of mutation fixation by the genetic drift is inversely proportional to the effective population size.[24] Another phenomenon that is important during microbial evolution is clonal interference: in large population sizes, beneficial mutations arise independently and in the absence of genetic recombination they compete with each other for fixation in the population. This results in an increase of the rate of evolution with the square root of the population size, instead of its linear dependence to it for smaller population sizes. Clonal interference has been observed in laboratory settings for both viruses and bacteria,[40–42] its dynamics have been investigated in computational models of evolution,[43–45] and its effect has been recently measured experimentally in eukaryotic yeast cells.[46]

## *In vivo* laboratory evolution

The combinatorial explosion of mutation sets and the clonal interference limit on the effective population size are the main limitations in ALE experiments, as it is infeasible to search the genotypic space in a systematic way. This is also the reason that replicate experiments give very different results when it comes to the specific mutations that are present in the adapted strains, as the appearance of neutral or near-neutral mutations largely resembles a random walk. For example, in massively parallel ALE studies[9,32,34] when several hundred biological replicates are evolved simultaneously, the observed mutations across the different cell lines have a low overlap. This overlap can increase when the selection pressure is strong and there are limited ways that the cell can cope with environmental variation, as in the recent example of 115 biological replicates of *E. coli* evolved over 2000 generations under heat stress.[32] In this experiment, although individual fixed mutations were mostly unique over a hundred of biological replicates (only 3% of mutations are shared between any two evolved strains), the target gene and protein functional groups had a high overlap (25% of affected operons shared between any two strains). In addition, in certain environments, the order by which mutation appear is also conserved over multiple biological replicates.[12]

**Model organisms**

Virtually any cultivatable microorganism could be used for *in vivo* evolutionary experiments. However, some model organisms are more preferred due to the existence of reference genome annotation, known gene functions, partial gene regulatory network and technical expertise for genetic manipulation. *E. coli*, the first gram-negative bacterium to be sequenced and with the most extensive database of gene function and regulation,[47,48] is the workhorse for laboratory microbial evolution.[7,12,26,27,29–32,38,45,49–56] Other organisms that used for ALE experiments include yeasts,[9,10,34,57,58] which are of a particular interest as they can reproduce both asexually and sexually, *L. lactis* a bacterium with high importance due to its present in skimmed milk,[33] *P. aeruginosa*, which is often responsible for many bacterial infections, especially in lungs,[59] cyanobacteria and other more exotic organisms (*e.g. B. boroniphilus* used to engineer improved boron resistance[28]). Often studies of microbial evolution are accompanied or complemented by studies in viruses, which exploit their high adaptive potential and plasticity to understand the fitness landscape and distribution of fitness effects during evolution.[13–16]

**Experimental methods for laboratory evolution**

Two common techniques that are applied in laboratory evolution are serial passages and chemostat cultivations of tightly controlled environments[8] (see Fig. 2). As mutations occur mostly during the DNA replication, the overall number of accumulated mutations, and therefore the speed of adaptation, is a function of the total number of microbial divisions. In order to maximize the evolution rate over a given time, the population growth can be maintained close to the exponential limit: the culture can be keep growing in a flask and then be diluted with fresh media once the nutrients are exhausted (serial passages with dilution ratio from 1 : 50 to 1 : 500), or alternatively a constant growth in the exponential phase is maintained through a chemostat cultivation.[9,10,60]

The use of chemostats is attractive, as they are semi-automated and allow for a much more precise control of the environmental parameters (nutrient abundance, pH, temperature, *etc.*). These parameters can be varied over time with a desired periodicity for evolution studies in variable environments. Recently a chemostat-like setting that was called a "morbidostat", the stressor concentration (antibiotics in this case) was automatically increased proportionally to the adaptation level of the population, so that a constant population size and division rate is achieved.[12]

The drawback for use of semi-automated chemostats is the high cost of the experimental setup, the high maintenance and

View Article Online

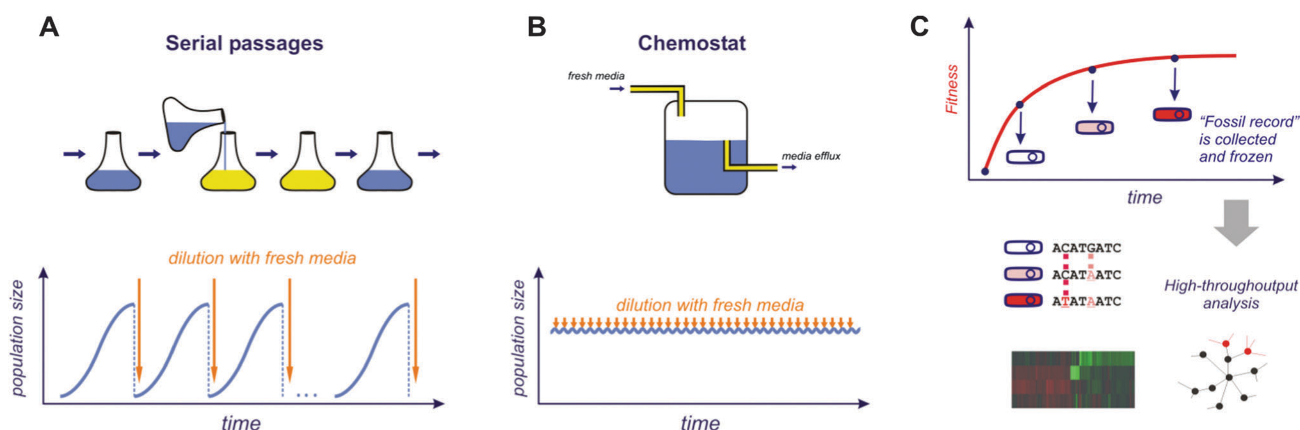


**Fig. 2** Methods used for studying microbial evolution *in vivo*. The most common techniques used to maintain control over environmental conditions and microbial growth rates for prolonged times are: (A) serial passages and (B) chemostat cultivations. In a chemostat it is possible to continuously monitor and control the growth conditions at a desired level; however this approach requires a complex experimental setup. When a large number of strains need to be evolved in parallel, serial passages are frequently used to reduce the cost of the experiment. In the latter case, microbial cultures are periodically diluted with a fresh media (usually daily) to limit the microbial concentrations and supply the growing populations with new nutrients. Inevitably, some of the environmental parameters fluctuate between the passages (*e.g.* the nutrient concentration and pH), but other parameters like temperature and oxygen concentration can be kept constant. In a proper experimental setting the effect of the environmental fluctuations on the population growth is kept at a minimal level. (C) Intermediate phenotypes that emerge during evolution can be sampled and stored for later analysis. Although we are still limited in our ability to study population structure in large scale, keeping a ''fossil'' record of the evolutionary trajectory is paramount for re-tracing evolutionary steps, ''replaying'' evolution, and for further analysis in the future.

Downloaded by RSC Internal on 11 December 2012
Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C

operation costs, and the need of multiple vessels to run parallel experiments and controls. For this reasons, serial passages in shake flasks or multi-well plates are a simpler and more popular alternative for performing laboratory evolution. In these settings the environmental conditions are more loosely controlled, with some parameters being easily controllable (*e.g.* temperature and oxygen concentration), while others varying considerably within every serial passage cycle (nutrient concentration, pH, cell density, *etc.*). Another difference between serial passages and the chemostats is that homogeneous populations are more frequently observed in serial passages.[8]

To create a selective environment during the laboratory evolution, various levels and combinations of stressors are usually applied. Some common stressful environments include temperature shifts, pH fluctuations to acidic or alkaline environments, oxygen deprivation, antibiotics, carbon source and amino acid limitation, and various types of disrupting chemical stressors.[10,12,27,31,32,36–38] Both mild and strong levels of these stressors are applied, with the latter leading to mutations that are associated with larger fitness effects and therefore faster fixation of the beneficial mutations.[7] To shorten adaptation time, a library of mutants is sometimes used as a starting point for the various cultures.[9,10,28] More advanced approaches to ALE may include sequential adaptation to multiple[9] and/or time-dependent[52] stresses to study the cross-protection and the interaction between different stress response pathways or non-constant environments. Table 1 summarizes the various approaches so far.

## Systems biology analysis

The high-throughput techniques that have revolutionized genome sequencing and transcriptional profiling over the past decade provide researchers with a powerful arsenal for mapping genetic and certain epigenetic changes that appear during evolution. In addition, the annotated genomes of all model organisms are readily available, and in many cases functional gene networks help to infer the functional categories of many genes. However, even in the case of the most studied model organism, the bacterium *Escherichia coli*, only about half of the genes are included in the known functional gene network,[47,48] while more than one fourth of the genes are currently not associated with any sigma factor. Even for the genes that are included in the network, many links are currently missing as new condition-specific experiments constantly reveal novel associations between genes that previously were thought to be uncorrelated. The data is much more sparse for other model organisms and while networks of protein–DNA interactions are available, they usually do not include the sign of regulation (inhibition or activation) and never include the weight (the strength of regulation). Clearly, we have still a lot to learn about bacterial physiology and the mechanistic underpinnings behind complex microbial behaviors.

The lack of the complete information in the ''*omics*'' databases is compensated with the development and the wider availability of the high-throughput techniques at a decreasing cost. The explosive growth of the technology advanced the field of ALE over the last decade and recently published experiments usually include high-throughput analysis at some level.[8] Whole genome resequencing makes it possible to obtain a complete list of mutations of a microbe at any point of the evolutionary trajectory relative to the reference.[8] Using RNA-Seq technology[61] one can analyze the expression patterns of the entire organism in different environmental conditions and with the help of ChIP-Seq[62] it is possible to infer new gene regulatory pathways in model organisms.[63,64] The integration of advanced high-throughoutput tools with the gradual accumulation of the knowledge in the ''*omics*'' databases helps to study


This journal is © The Royal Society of Chemistry 2012

**Table 1** An overview of *in vivo* recent microbial evolution studies

| *In vivo* laboratory evolution method | (Publication year) Objective of the studies | Results |
|---|---|---|
| *Species*: *E. coli* REL606<br>*Time*: $> 50\,000$ generations<br>*Environment*: minimal media + limiting glucose, 37 °C<br>*Populations*: 10 ml, $5 \times 10^8$ cells before each dilution; 12 parallel populations<br>*Technique*: serial passages; daily 1 : 100 dilutions | (2009) *E. coli* long term evolution experiment.[7] | Large number of mutations were accumulated over first 2 K generations ($\sim$25 mutations); between 2 K and 20 K the mutation accumulation rate was almost constant; after 20 K generations 45 mutations were accumulated (SNPs account for 64% of all mutations). Four populations evolved a $\sim$100-fold increase in mutation rates. All 12 populations evolved large increase in average cell size. |
| *Species*: *E. coli* MG1655<br>*Time*: $\approx$1100 generations<br>*Environment*: M9 minimal media + lactate, 30 °C<br>*Population*: 250 ml, $5 \times 10^5$–$5 \times 10^8$ cells transferred at each dilution; 11 parallel populations<br>*Technique*: serial passages; dilutions to maintain the exponential growth | (2009) Genetic basis of *E. coli* adaptation to the minimal lactate media.[29] | 2–8 mutations per strain (4.8 on average) were accumulated; 90% increase in the growth rate observed after 1100 generations; significant variance in mutation types between replicates; majority of mutations relate to metabolism, regulation and cell envelope functions. |
| *Species*: *E. coli* (MG1655, transposon and over-expression libraries)<br>*Time*: short time selection (5–10 generations) of mutants; 80 generations for the wild type evolution<br>*Environment*: LB or M9 + glucose media; supplemented with ethanol (4–7% v/v)<br>*Technique*: fitness profiling of the transposon and over expression libraries; serial passages for the short time evolution | (2010) Regulatory and metabolic rewiring during evolution of ethanol tolerance in *E. coli*.[27] | Fitness profiling was used to estimate the effect of single locus perturbations (transposon insertion and over-expression libraries) on ethanol tolerance; fitness effects of transposon insertions were more pronounced; naturally accessible ethanol resistance pathways were analyzed based on the short time evolution experiment; the correlation between fitness profiling and evolution results was reported. |
| *Species*: *E. coli* MG1655<br>*Time*: $\approx$660 generations<br>*Environment*: M9 minimal media + 0.2% glycerol, 30 °C<br>*Population*: 250 ml; $5 \times 10^5$–$5 \times 10^8$ cells transferred at each dilution; 5 parallel populations<br>*Technique*: serial passages; dilutions to maintain the exponential growth | (2006) Genetic basis of *E. coli* adaptation to the glycerol minimal media.[30] | 13 mutations fixed in 5 sequenced stains; single, double and triple site-directed mutants were created to confirm fitness effect of the observed mutations. |
| *Species*: *E. coli* DH1ΔleuB<br>*Time*: up to 7560 generations<br>*Environment*: M63 medium + leucine, 36.9–44.8 °C<br>*Population*: 5 ml, two parallel populations evolved at constant (37 °C) and increasing (37 °C → 45 °C) temperatures<br>*Technique*: serial passages; daily dilutions to maintain the exponential growth | (2010) Positive to neutral transition in mutation fixation in thermal adaptation.[31] | After every temperature increase the fitness of the population dropped, but each time was almost completely recovered to the initial level after $\approx$50 generations. Two phase fitness dynamics was observed: positive was followed by the neutral; however constant increase in fitness was observed in both phases; whole genome sequencing confirmed transition from one phase to another. |
| *Species*: *E. coli* B REL1206 pre-adapted for 2000 generations at 42.2 °C<br>*Time*: 2000 generations<br>*Environment*: Davis minimal medium + glucose, 42.2 °C<br>*Population*: 10 ml, 115 parallel populations<br>*Technique*: serial passages; daily 1 : 100 dilutions | (2012) Massively parallel adaptation of *E. coli* to the high temperature.[32] | After 2000 generations 115 parallel evolved strains acquired on average 11 mutations (6.9 point mutations, 2.3 short insertions or deletions, 1.0 long deletion, 0.6 insertions, 0.2 large duplications). Most point mutations were unique, but 69% of large deletions were identical. Once genes with most mutations are divided into 10 functional groups (containing 34% of all mutations) any two lines shared on average 32% of the functional units. |
| *Species*: *E. coli* MG1655<br>*Time*: 20 days<br>*Environment*: M9 minimal media + glycose + antibiotics, 30 °C<br>*Population*: 12 ml, morbidostat<br>*Technique*: morbidostat | (2012) Evolution of antibiotic resistance under dynamically sustained selection.[12] | Sequential accumulation of mutations was observed under increasing drug selection (to maintain a constant growth rate). |
| *Species*: *L. Lactis* KF147<br>*Time*: 1000 generations<br>*Environment*: 10 ml skimmed milk<br>*Population*: $10^{10}$ cells before each dilution; 3 parallel populations<br>*Technique*: serial passages, 1 : 10 000 dilutions | (2012) Evolution of lactic acid bacterium *L. lactis*.[33] | 6 to 28 mutations were accumulated after 1000 generations; Reproduction of the transition from plant to dairy niche in the laboratory evolution reveals several signatures that resemble those seen in the strains isolated from wither niche. |

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

View Article Online

Downloaded by RSC Internal on 11 December 2012
Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C

Table 1 (*continued*)

| *In vivo* laboratory evolution method | (Publication year) Objective of the studies | Results |
|---|---|---|
| *Species*: *P. Aeruginosa*<br>*Time*: up 39 000 generations *in vivo*<br>*Environment*: 3 individuals with chronic cystic fibrosis<br>*Population*: *in vivo* airway infection<br>*Technique*: *in vivo* sampling | (2010) Parallel evolution of *P. aeruginosa in vivo* during chronic lung infection.[59] | Transcriptome profiles were obtained *over 8 years from 3 individuals with chronic airway infection*; genetic data was not collected. Genes with the common expression changes were identified in three populations, and authors hypothesized that some of these genes encode adaptive traits. However the number of traits changed in parallel is smaller than the number of unique changes (in one of the populations only 6% of changes belong to parallel evolved set). |
| *Species*: *B. boroniphilus* DSM 17376<br>*Time*: 50 days<br>*Environment*: TSB medium + boron ($H_3B0_3$), 30 °C<br>*Population*: starting population: EMS mutagenized<br>*Technique*: batch selection with daily dilutions | (2011) Application of the *in vivo* evolutionary engineering to improve boron-resistance.[28] | Highly boron-resistant mutants were obtained in the study. *Cross-resistance* of evolved strains to iron, copper, and salt (NaCl) was identified suggesting common stress-resistant pathways. |
| *Species*: *T. fusca*<br>*Time*: 40 days ($\approx$220–284 generations)<br>*Environment*: continuous exposure to cellobiose, or daily alternating exposure to cellobiose and glucose<br>*Population*: two parallel populations in different environments<br>*Technique*: serial passages | (2011) Laboratory evolution of cellulolytic actinobacterium *T. fusca*.[137] | Strain evolved in *alternating environments* shows increased cell yield for growth on glucose and shows more generalist phenotype. |
| *Species*: yeast strain DBY15084<br>*Time*: 1000 generations (10 generations per day)<br>*Environment*: 96-well plates; no shaking; 30 °C<br>*Population*: $10^5$–$10^6$ cells; 592 parallel populations<br>*Technique*: Serial passages in a laboratory automation workstation; 1 : 32 dilution every 12 hours or 1 : 1024 dilutions every 24 hours | (2011) Massively parallel study of the fate of the beneficial mutations.[34] | Flow cytometry was used to detect evolved sterility (fitness advantage). Found that fitness advantage by individual mutations plays surprisingly small role; underlying background genetic variation is quickly generated in the initially clonal populations and is crucial for determining the final fate of beneficial mutations. *Different scenarios where observed: sweep, clonal interference, multiple reappearance, frequency-dependent selection*. |
| *Species*: *S. cerevisiae* CEN.PK 113.7D<br>*Time*: $\approx$10 days under each batch selection/stress<br>*Environment*: Oxidative ($H_2O_2$), heat, ethanol, and freezing-thawing stresses<br>*Population*: 96-well plates; several parallel populations for each stress<br>*Technique*: batch selection of chemically mutagenized cells in a laboratory automation workstation; also a short time (68 generations) chemostat selection | (2005) Evolutionary engineering of multiple-stress resistant *S. cerevisiae*.[9] | Among all chemostat and batch selection strategies, batch selection under freezing-thawing stress showed the most highly improved *multiple-stress-resistance*. |
| *Species*: *S. cerevisiae* W303-1A<br>*Time*: 200 days<br>*Environment*: ethanol stress, batch growth<br>*Population*: mutagenised and non-mutagenised starting populations<br>*Technique*: chemostat | (2010) Generation of stable ethanol-tolerant mutants of *S. cerevisiae*.[10] | Both pre-mutagenised and non-pre-mutagenised starting populations show remarkable similarity across a range of ethanol stress conditions. The mutants produced significantly more glycerol than the parent; it was suggested that it is the mechanism for the observed increased ethanol tolerance. |
| *Species*: φX174 Microviridae<br>*Environment*: mostly *E. coli* hosts, but also *Shigella* and *Salmonella* was used; 32–42 °C<br>*Population*: population sizes smaller than the host population sizes to limit co-infection and recombination<br>*Technique*: serial passages every 30 min for 40–90 hours; total, or chemostat cultures for 10–180 days | (2010) Review of Microviridae experimental evolution.[138] | φX174 is a common model system for experimental evolutionary studies; review summarizes 10 experiments with 58 lineages in total; strong *parallel evolution* is observed (50% of substitutions occurs at sites with three or more events). |

the genotypes and phenotypes of the evolved microbes at the level of details that it was not possible to achieve before.

### Challenges for the *in vivo* evolution

While recent advances have resulted in enabling high-throughput technologies (Fig. 3), there is little that we can do to address the gap between the temporal scales under which evolution operates, and what we can observe in the laboratory. Inevitably, ALE studies tend to be labour-intensive and/or costly. The experiments have to last for months or years, and it is still just a blink at evolution timescales. In addition, regardless on the experimental setting, constant human involvement and control is required. A significant number of biological

 This journal is © The Royal Society of Chemistry 2012

View Article Online

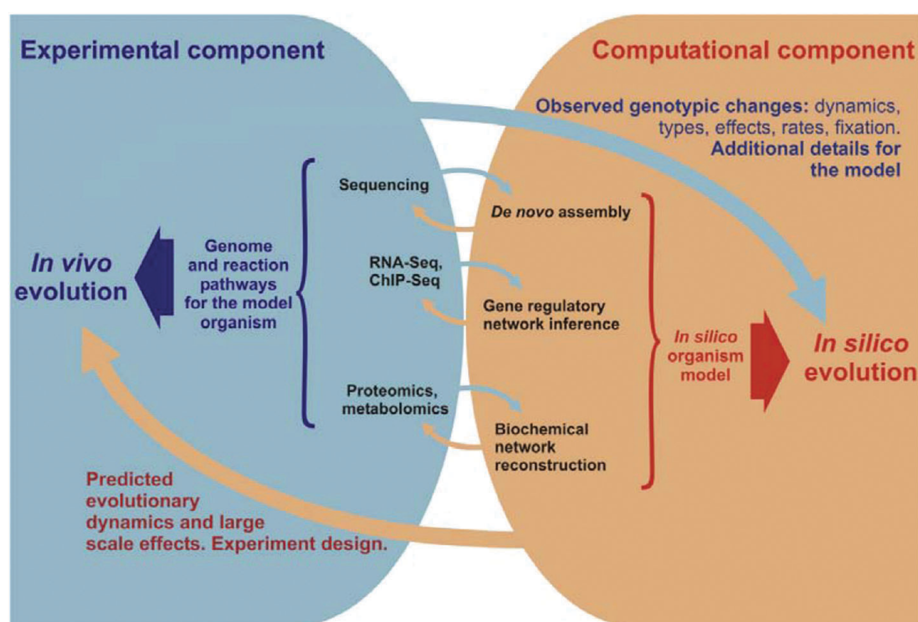


**Fig. 3** Interactions between theory and experimentation. Laboratory evolution of microbial strains results in a large number of biological datasets that include phenotypic information, mutation maps, transcriptional, proteomic and metabolomic profiling of the evolved strains. These large datasets are further processed by a number of computational techniques that include the *de novo* or referenced assembly of the sequencing reads, network inference and computational modelling of the various biological pathways. The latter are further used to gain a deeper insight on the mechanistic basis and interaction between their components. These models are further utilized to generate biological hypotheses that in turn provide the basis of further experimentation.

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

replicates are necessary to ensure that the evolutionary trajectories in a multi-dimensional genotypic space are not a random coincidence of events. If additional verifications or replicates are desired after the first round of adaptation is completed are analysed, significant time (equal to the time of the original experiment) is required for these additional experiments.

In addition, even with the existence of high-throughput technologies we are unable to have a high resolution map of the various changes that appear in an evolving population across various "*omics*" scales. While the cost of the high-throughput output methods constantly decreases, the overall cost of the analysis of the ALEs adds up quickly due to the combinatorial explosion that takes place when we consider number of necessary replicates, environmental conditions, and temporal samples that have to included for a comprehensive analysis of microbial evolution.

## *In silico* microbial evolution models

The many challenges that are associated with laboratory microbial evolution have motivated the development of variety of computational tools to simulate evolution *in silico* (Table 2). Computational evolution models attempt to closely replicate biological and evolutionary processes both in single cell and population level. In order to maintain the proper balance between biological realism and computational feasibility, the various models include multiple levels of abstraction and crude parameterization. The resolution of each model depends both on the questions that it aims to address and the type of parameters that are available through experimental measurements. In this section we describe the parameters, components, and models of various *in silico* evolution models and discuss their role in the realistic description of the *in vivo* processes.

The wide spectrum of approaches implemented in *in silico* evolutionary models fall in one of three main categories, depending on how the cellular organization is represented in the model (Fig. 4): methods that use a pseudo-alphabet, methods that include a representation of the gene regulatory and biochemical network, and methods that use other means to represent the evolution of ordered computation (as in the structured sequence of instructions present in the "digital organisms" of the Avida simulator). The bottleneck that all *in silico* models share is how to realistically map the organisms' genotype to the observed phenotype/fitness in any given environment.

One of the abstraction extremes is the population models that include only a simplistic organism model, often consisting of just the organism fitness and allelic frequencies.[65–68] These models are best suited to simulate population proprieties at short timescales, but without the inner structure of the organisms their ability to realistically associate mutation events to phenotypic differences is limited.

To realistically capture evolutionary dynamics and their effect in cellular organization, the ideal model should capture multiple scales of biological organization. First, rudimentary models of basic biological processes such as transcription, translation, protein modification and gene regulation, should be incorporated in the model. These processes can act as functions within the model of a cell. Each cell should also be described by a number of parameters (size, energy levels, doubling time, *etc.*) that are relevant to the biological questions that the simulation aims to address. In addition to the cellular functions, another set of actions

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

**Table 2** An overview of *in silico* microbial evolution models and applications

| *In silico* models | (Publication year) Objective of the studies | Results |
|---|---|---|
| *Gene network model*<br>*Organism*: genes regulatory network of ~4 abstract genes either in 'active' or 'inactive' state depending on the incoming regulation<br>*Population*: fixed size $N = 10^3..10^4$; sexual or asexual; random initialization<br>*Fitness* depends on the mismatch of the gene states (phenotype) with the optimum<br>*Environment*: fluctuating or stable<br>*Mutations*: network connectivity or weights<br>*Selection*: reproduction with probability ~fitness | (2009) Hypothesis: natural selection favors high evolvability, especially in sexual populations.[87] | Fluctuating natural selection increases the capacity to adjust to new environments. Shifts in the directional selection cause evolvability to increase. |
| *Evolution of signaling pathways*[18,88]<br>*Organisms*: signal transduction, biological pathways; network of protein interactions; up to 2..18 proteins in one of two states<br>*Population*: 1000 randomly initialized pathways<br>*Fitness*: almost the same for all pathways, decreased for larger pathways | A. (2008) Evolution of taxis responses.[19] | A. Concludes that the evolved taxis pathways can contain features of both adaptive and non-adaptive dynamics, depending on the stimuli conditions during evolution. |
| *Environment*: signals are the ligand concentrations<br>*Mutations*: during replication of a randomly selected organism<br>*Selection*: varies; *e.g.* the magnitude of the pathway's transient response | B. (2006) Evolution of complexity in signaling pathways.[18] | B. Found that stringent selection pressure result in a smaller, less complex pathways; in contrast to more complex pathways evolved under simple response requirements. |
| *EVE* (*Evolution in Variable Environments*)[94,115,139]<br>*Organism*: gene regulatory and biochemical network of 10..100 'triplets' (mRNA-protein-modified protein nodes)<br>*Population*: constant size, 100..1000+ organisms, random initialization | A. (2008) Predictive Behavior Within Microbial Genetic Networks.[94] | A. Both *in silico* models and *E. coli* transcriptional responses show predictive behavior and reflect an associative learning paradigm. |
| *Fitness*: correlation between the metabolic pathway expression levels and nutrients availability<br>*Environment*: input signals varying over time; nutrients availability is a delayed function of the signals | B. (2011) The effect of horizontal gene transfer (HGT) on the evolution of gene regulatory networks.[93] | B. Distribution of Fitness Effect (DFE) for HGT events is analyzed. The effect of HGT on organism regulatory organization is studied. |
| *Mutations*: modifications of the network topology and weights<br>*Selection*: quality of the nutrient availability predictions leads to faster replication | C. (2012) A rate of evolution of a population exposed to a sequence of environments depends on a relative similarity and complexity of those environments.[92] | C. Evolution can be accelerated by evolving cell populations in a sequence of environments that are increasingly more complex; weak environmental correlations decelerate evolution. |
| *Avida*[74,75]<br>*Organism*: short codes with 100..300 instructions;<br>*Population*: limited by the size of a 2D grid (~60 × 60), offsprings are replicated in to the neighboring cells replacing the existing organisms; | A. (2001) High mutation rates promote the survival of the 'flattest'.[140] | A. At high mutation rates organism with a high fitness can be outcompeted by those with a lower fitness that is more robust with respect to mutations. |
| *Fitness*: increases exponentially with the number of evolved binary functions as the organism approaches a preset response function<br>*Environment*: input of random binary strings | B. (2008) Effect of genetic robustness on evolvability.[17] | B. Likely conclusion: robustness relaxes the selection; in the short the evolvability is hampered, in the long term genetic diversity is accumulated and the evolvability increases. |
| *Mutations*: lines of the code are randomly modified<br>*Selection*: organisms are limited in CPU time, more fitted organisms are rewarded with bonus CPU time | C. (2011) The effect of low impact mutations on evolution of digital organisms.[141] | C. Digital organisms do not evolve if the predefined fitness benefit is below a certain threshold. Therefore the selection breaks down for fitness effect below the selection threshold. |
| RAevol (Regulatory Aevol)[72,73]<br>*Organism*: Gene regulatory networks: DNA, RNA, Proteins<br>*Population*: fixed size, ~1000 gene regulatory networks; random initialization<br>*Fitness*: "metabolic error", the difference between organism's and target phenotypes | (2010) Robustness and variability in static and variable environments[72,142] | Amount of accumulated non-coding sequences is inversely proportional to the mutation rate in log–log space; the total number of nodes and edges also scales down with the increase of the mutation rate;[142] applied to produce the artificial datasets, knock out maps, *etc.*[72] |

 This journal is © The Royal Society of Chemistry 2012

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

**Table 2** ***(continued)***

| *In silico* models | (Publication year) Objective of the studies | Results |
|---|---|---|
| *Phenotype*: a function of metabolic proteins; mapped from a genotype *via* artificial chemistry<br>*Environment*: a target phenotype; time dependent<br>*Mutations*: alternations to the genome sequence; introduced at every generation<br>*Selection*: number of offsprings is based on the fitness | | |
| *Evolution of gene regulatory networks*[107]<br>*Organism*: gene regulatory network (DNA, protein), *cis*-sites (100 cites per gene, unique mapping between *cis*-sites and proteins); two phenotypic genes; ∼10 genes<br>*Population*: ∼10 000 organisms; random initialization<br>*Fitness*: a proximity of organism's phenotype to the target phenotype;<br>*Phenotype*: expression level of two phenotypic genes;<br>*Environment*: random or cyclic fluctuating target phenotype;<br>*Mutations*: alternations to network parameters and *cis*-sites; duplication, deletion<br>*Selection*: division probability proportional to fitness | (2010) Evolution of gene regulatory networks in fluctuating environments.[107] | Studied the effect of various model parameters on the GRN evolution in fluctuating environments: mutations rates of different type, gene expression costs, network node degree distributions, *etc*. |
| *AG model* (*Artificial genome*)[86,89]<br>*Organisms*: A string in a 4-letter alphabet; 4-character promoters, ∼30 of 6-character genes; expressed genes bind according to their sequence and regulate first downstream gene; binary gene expression; organism can be represented as a gene regulatory network<br>*Population*: ∼50 random genomes<br>*Fitness*: assigned score based on a set of organism's proprieties (varies: network size, node degrees, *etc*.)<br>*Phenotype*: binary expression table of all genes at a given time point<br>*Environment*: defined by the set of proprieties used in fitness evaluation<br>*Mutations*: Base in AG can be miss-copied at the reproduction step<br>*Selection*: sexual reproduction of high scoring genomes | (2006) Study the model parameters and proprieties of the evolved organisms.[86] | Internal constraints and biases of the model are reported. |
| *Evolution of gene regulatory networks*[21,90]<br>*Organisms*: A set of genes with *cis*-binding sites for regulators; one-to-one map between genes and binding sites; Boolean network dynamics[85]<br>*Population*: fixed size; 1000 organisms; pre-evolved under 'neutral evolution' (no selection) to produce realistic networks<br>*Fitness*: based on the production of the 'biomass' by 'biomass pathway' genes<br>*Environment*: input signals are food processed by special proteins<br>*Mutations*: changes in genes and binding site types, amplification, deletion, HGT<br>*Selection*: organisms with higher fitness are more probably to replicate | (2010) Stochasticity *vs*. determinism.[21] | Genomes evolved under the stochastic gene expression model are smaller than ones evolved with deterministic model. Stochastic model is more capable of evolving solutions. |
| *Evolution of gene regulatory networks*[91,95,96]<br>*Organisms*: a network of 11 species (DNA, input proteins, transcription factors, mRNA, proteins, RNAP, *etc*.). Allowed *reactions* are described with ODE or a stochastic algorithm<br>*Population*: 40 individuals, random sexual reproduction | (2008) Robustness as emergent propriety of evolution.[95] | Effect of starting populations (with various degree of the transcription factor specificity), input types, and objective function were studied. It is observed that robustness is evolves along with the network evolution even without a direct selection pressure towards more robust organisms. |

View Article Online

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

**Table 2** (*continued*)

| *In silico* models | (Publication year) Objective of the studies | Results |
|---|---|---|
| *Fitness*: based on the difference in the expression levels of a special proteins (biomass production) and the rest of the proteins; as an option can be also inversely proportional to the total number of network links<br>*Phenotype*: protein expression values<br>*Environment*: inputs connected to the sensor proteins and the type of fitness function itself<br>*Mutations*: changes to network; artificially high mutation rates are used to speed up the convergence<br>*Selection*: $\frac{3}{4}$ of organisms are replaced with $\frac{1}{4}$ of fittest organisms (to protect fit organisms from high mutation rates) | | |
| *Artificial chemistry and evolution of metabolic networks*[71]<br>*Organisms*: genome (cyclic RNA 5000 bases long, genes are 100 bases long) + metabolic system; "sequence → structure → function" model; sequence is transformed into a 3D structure; function is assigned using a heuristic algorithm. Metabolic reactions are described with artificial chemistry model<br>*Population*: ~100 organisms<br>*Fitness*: based on flux balance analysis, a maximum of the yield function<br>*Environment*: chemical composition of the food source; time variable<br>*Mutations*: applied to the genome sequence directly<br>*Selection*: direct competition for the resources | (2011) Evolution of early metabolism.[71] | Model attempt to provide "genome → phenotype → fitness" map for a given genomic sequence using artificial chemistry model. Sowed that shapes the metabolic networks at early stages. Later stages keep the core set of metabolic pathways unchanged and alter the enzyme recruitment patterns. |

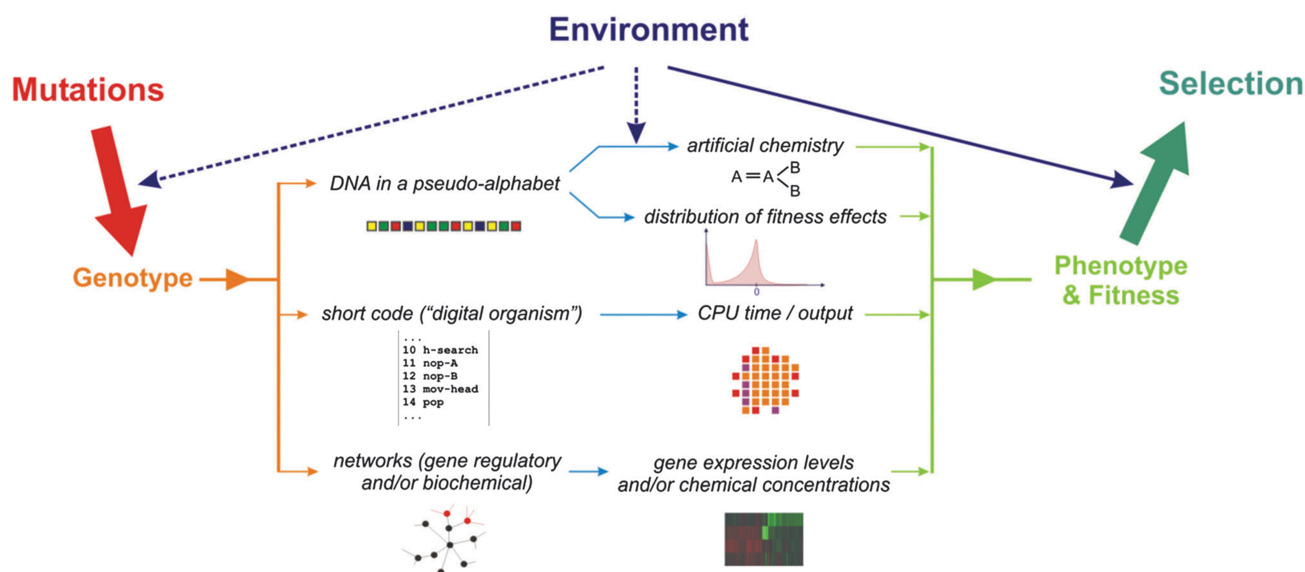


**Fig. 4** Mapping genotype-to-phenotype for microbial evolution *in silico*. The most common approaches to map the genotypic and phenotypic relationships are (a) using a pseudo-alphabet together with a fitness effect distribution or artificial chemistry models, (b) the representation of an organism's genotype as a list of programming commands that are further evaluated in a specific task or objective function, (c) the inclusion of a gene regulatory, biochemical and metabolic network with a distinct temporal profile that is related to phenotypic fitness. In all models, mutations affect the genotypic space, while selection acts on the various phenotypes, with the environment affecting both these processes.

that can modify cellular organization and parameters is needed to model the evolutionary processes that the model tries to capture. These include random mutation that can be of different kind (*e.g.* hits in the promoter region *vs.* the protein domains) and magnitude (*e.g.* fluctuation around the existing parameter value *vs.* parameter randomization), recombination, insertion

 This journal is © The Royal Society of Chemistry 2012

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

of genetic fragments through lateral gene transfer, *etc*. Each cell can act as an independent agent that receives environmental cues, processes their information through its internal pathways and reacts accordingly. Another important aspect of any multi-scale evolutionary model is to incorporate a representation of the environment that the simulated organisms occupy. The environmental model should include the various dynamic signals present to the environment and environmental parameters that affect the organism-specific parameters in the model.

It is evident that such level of multi-scale modeling faces significant hurdles in many dimensions. First it will have to incorporate processes that have fundamentally different time-scales that range from milliseconds to decades (Fig. 1A). The same is true but at a lesser extend for the spatial component of the model, which may have to range over six orders of magnitude. In addition, since most evolutionary simulations are performed in a discrete time, the choice of the time step is not trivial. A rule of thumb is that the time step should be smaller than the frequency of mutations and divisions (the lifetime of a single generation) and that the overall length of a simulation should be sufficient to describe mutation fixation and natural selection. If the model has to include gene expression and molecular interactions then the time step must be further reduced at least $10^{-3}$ relative to the generation time.

### Modeling of genetic information

Models which include the description of an organism at the genome level, have to either assign the fitness effect to each random mutation randomly by drawing from a predefined distribution of fitness effects[69,70] or to design a chemical representation which map the genotype to fitness.[71–73] These models are almost always restricted to abstract, pseudo-genomic alphabets to reduce the genomic space. However, the advantage of having a DNA level in the model, even though described through an abstract alphabet, is its flexibility to incorporate more realistic representations of different types of mutations and maintain experimentally-derived relative probabilities for these events.

As it is difficult to calculate the fitness effect of random mutations if only the organisms' genome is known, many models include hard-coded, genotype-fitness or network-fitness associations. A common model is based on gene regulatory networks, which are discussed in the next section. Alternative representations include associating each organism with a short code (a sequence of instructions) as it is implemented in the Avida digital organism simulator.[74,75] In that case, mutations modify and replace the instructions while the fitness is based on the output of the evolving program.

### Gene regulatory and biochemical network

Biological functions can be described with a high degree of realism using a systems biology approach. The interaction between the molecules inside the cell can be described as a graph[76] (usually directed and weighted). The nodes may represent molecules (*e.g.* DNA, mRNA, proteins, small metabolites). Links between the nodes correspond to regulatory associations. The direction of the link may represent the source–target relationship of the interaction pair while the edge weight captures the strength of the regulation. The values assigned to the nodes of the graph may correspond to the number of molecules of each type or respective concentrations. With a realistic expression model the dynamics of the gene regulatory and biochemical network can closely resemble the dynamics of chemical reactions in a living cell. Mutations are introduced as the alternations of the network parameters and topology. For example, the change in the weight of a protein-to-DNA interaction may correspond to a mutation which alters the binding affinity of a transcription factor to its binding site and a replication of a network fragment is equivalent to the gene duplication. This approach helps to assign fitness to an organism using direct evaluation of its network dynamics, as compared to randomly sampling a fitness distribution.

One of the challenges associated with these methods is to provide realistic parameter ranges for the various parameters in the network, such as the reaction rates and binding affinities. Another challenge is to maintain a realistic ratio between different mutation types (various alternations to the organisms' network), as it directly affects the effective distribution of fitness effects during a simulation run. Hence, network mutation models should be adjusted to produce distribution of fitness effects similar to ones observed in nature.

The choice of the initial state of the evolving organisms is a challenge in itself. Hopefully, in the nearest future it would be possible to construct an initial network based on the propriety of a particular model organism. However, we still do not have enough information to represent the complete expression dynamics even in the case of well-studied model organisms. The obvious alternative is to start with a random network, which resembles in its proprieties the gene regulatory and biochemical network of the organisms that we would like to study.[77–79]

### Expression models

For small regulatory networks and pathways, reaction constants can be well studied and the dynamics of the expression can be described efficiently by systems of differential equations or a variety of stochastic methods. An overview of the mathematical models for dynamics of the biological pathways and networks can be found in a number of reviews.[80–84] However, due to a the large scale of simulations with respect to population sizes and number of generations, *in silico* evolution simulators are generally limited to the use of simplified gene expression and regulatory models. Boolean network[85] of genes in "on" or "off" state is one of the most frequent choices.[18,21,86–90] Alternatively continuous concentrations of cellular components can be simulated deterministically and/or stochastically[91–96] as a function of the total incoming regulation to each node.

When it comes to modeling the dynamics of biological networks several options exist. In deterministic models of biological systems,[75,97,98] a certain input always produces the same output, and these often utilize ordinary differential equations (ODE). In contrast, stochastic differential equations (SDE) are used when system behaviour may be sensitive to noise.[21,99–102] Additionally, various biological systems such as transcriptional networks include reactions that are slow and with a low number of interacting molecules, which make the use of differential equation methods unsuitable. In such cases, stochastic simulations of discrete intervals have been found to be

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

more accurate, thus matching better the biological dynamics and providing valuable insight to experimentalists on the dynamic system behaviour, usually at the cost of computationally more intensive simulations.[103] Invariably, stochastic models use a probabilistic framework to capture the stochastic nature of the respective biological process, although the underlying algorithm may vary depending on the accuracy and performance that we want to achieve. Popular simulation algorithms include dynamic Monte Carlo approaches, Gillespie variants,[69,100] and discrete event runs.[104] Some models integrate stochastic and deterministic models: for example TinkerCell[105] is a modular CAD tool for systems and synthetic biology, where the user can create designs and simulate their behaviour by utilizing deterministic (ODEs), stochastic and flux-balance models. Genotype–phenotype relationships are reconstructed for metabolic networks using a variety of constraint-based reconstruction and analysis methods.[106] Hopefully, with further development of highly scalable *in silico* evolution codes and continuous accumulation of the experimental parameters, microbial evolution models will be capable to include expression models at a level of precision which is already achieved for smaller, non-evolving pathways.

### Fitness

Generally, fitness characterizes the suitability of an organism to grow in a specific environment. Experimentally, measures of an organism's fitness include the growth rate (or doubling time), the survival rate after environmental perturbation, and the ability to outcompete other organisms under a controlled environment.

Various *in silico* models propose a number of approaches to evaluate and use fitness and/or energy. Fitness is rarely a pure metric or an output parameter of the model.[71,94] More frequently it is used to adjust the population dynamics and serves as both an output and an input parameter of the model: the replication probability[86,87,107] or the number of offsprings[21,72,73,90] can be set proportional to the calculated fitness, alternatively the computational time can be rewarded to an organism based on its fitness.[74,75] There is an important difference in these approaches as they postulate one set of parameters and evaluate the other.

Since fitness is always a function of the phenotypic profile under a given environmental context, *in silico* models need to encode the structure of the environment within their models. This is usually achieved by representing the environment as a collection of input signals, which are connected to and processed by each individual organism[18,74,75,88,92–94] or by adjusting the target phenotype directly[72,73,107] and therefore the genotype-to-fitness mapping function is modified to represent the environmental changes. If an artificial chemistry is part of the model, an environment may include the nutrients of various chemical composition metabolized by each evolving organism.[71] Any model can be easily extended by varying the signals over time to represent non-constant environments.

The overall fitness function of the genotypic and environmental parameters can be represented as a multi-dimensional fitness landscape.[108] Evolution therefore can be visualized simply a search for the maximum fitness on this multi-dimensional surface. However, even if an analytical shape of the fitness landscape would be given, its large dimensionality and the complex shape with multiple local optima would make the exhaustive search for the global maximum impossible. Moreover, the fitness landscape surface can be continuous only with respect to either the genotype or the fitness, but not both. Most heuristic algorithms employ the Monte Carlo approach in attempt to sample the parameter space and the sampling size is rarely sufficient to cover the solution space.

### Distribution of fitness effects and mutation rates

The distribution of fitness effects (DFE) is particularly useful on discriminating between lethal, deleterious, neutral, and advantageous mutations. The shape of DFE is an important characteristic of evolution and cannot be ignored if the natural selection is being modeled. DFE, together with the known mutation rates and population sizes can provide a solid basis for a model of mutation fixation dynamics. It is hard to measure DFE experimentally,[13–16,49,109] as one is usually able to observe only fixed mutations (usually with a strong positive effect on the organism's fitness) after the natural selection 'discarded' most negative and all the lethal mutations. However, DFE was measured for random single point mutations in viruses[13–16] and bacteria.[49,109] DFE follows the general trend which is in agreement with nearly-neutral evolutionary model:[110] most mutations are neutral, with the distribution slightly skewed towards the deleterious mutations, significant number of mutations is lethal, and few have a strong positive effect.[111] The distribution of beneficial mutations is believed to be exponential[112] (*i.e.* with a longer 'tail' then a Gaussian distribution would have), therefore few mutations with very strong positive effect on fitness could be observed in any environment. *In silico* microbial evolution models, which cannot directly map the genotype to phenotype, randomly draw fitness effects of new mutations from a predefined DFE curve.

Although both distributions of different mutation types and mutation rates are reasonably well studied *in vivo*,[22,23] adaptation of these parameters by the *in silico* models is complicated by the abstract nature of the models. In some cases, such as the Avida simulator,[74,75] mutation rates cannot be directly mapped as the underlying substrate is not the same (*e.g.* programming code *vs.* DNA changes). Trying to quantify mutations in the gene regulatory network is not a much easier task: the topology and weights attempt to represent the molecular dynamics within the organism, but not the underlying genotype. Models with an implicit DNA level have an advantage of a direct control on the mutation types and rates, but suffer from the undefined fitness effects for these mutations as it was described above.

### Population-level effects

A detailed cellular model which describes its short term dynamics is usually too complex for evolutionary simulations in a population level. However the mutation fixation dynamics strongly depend on the population size: in small populations the random drift plays a significant role,[43] and even deleterious mutations can be fixed by a random chance; large microbial populations are subject to the effect of clonal interference.[34,113] The latter effect results in a higher relative heterogeneity of large bacterial populations and in the parallel occurrence of equally viable but

 This journal is © The Royal Society of Chemistry 2012

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

genetically different organisms in one population. While the mutation rate (per organism, per generation) is usually constant in *in silico* experiments, the rate of mutation fixation and the type of fixed mutations strongly depends on the selection pressure. The presence of stress induces a higher effective fixation rates.[114] Selection pressure *in silico* is controlled by the cost/benefit balance between the complexity of the environment and the reward level for the partial adaptation.

Another common problem for all computational models is the difference between "biological" and "computational" time. Evolving organisms can shrink or expand in size during evolution which directly affects the time that it is needed to simulate their behavior. To cope with this discrepancy, *in silico* implementations either sequentially apply mutations and selection to the entire population or include synchronization points among parallel processes that calculate the values of every cell in the population.[115] Alternatively, the individual selection pressure is adjusted according to the computational time used by an organism.[75]

### Parallel simulations of microbial evolution

The level of details that could be included in any *in silico* microbial simulation, population sizes and the evolution time is always limited by the available computational power. As computational clusters and cloud computing resources become widely available, the natural direction to expand the capabilities of the *in silico* simulators is to utilize parallelization. While this is yet virtually unexplored possibility, simulations of the microbial evolution can be parallelized quite easily. Most simulators describe evolution of single organisms in a population independently, with sequential application of the selection pressure to favour most viable organisms. Population can be split into groups of organisms each of those can be evolved in parallel[115] with frequent updates regarding the state of the environment, nutrients availability, population state, *etc.* in order to keep the time synchronized between the population sub-groups. This approach makes the parallel implementation almost trivial. However such a naive approach faces significant problems with the scalability of the code. As organisms evolve, variation in computation times between computational groups becomes significant, and the need for synchronisation increases the idle time as all computational process wait for the slowest one to complete the evolutional task. Once the number of parallel processes reaches thousands or tens of thousands, the idle time starts to dominate over the actual computing time.

### Challenges for microbial evolution models

As discussed above, one of the main challenges in the *in silico* modelling of microbial evolution is to achieve the right balance between biological realism and computational feasibility. This becomes a particularly formidable problem when we consider that the various processes that should be incorporated in the model have a wide range of temporal and spatial scales. It is ultimately left to the discretion of the researcher to decide the level of detail and biological organization that should be included in the model, which in turn it will determine the questions that it will be able to address.

Another challenge that is common for all modelling efforts in general, and for microbial evolution models specifically, is the availability of environmental and cellular parameters (Fig. 3). Any multi-scale effort to model microbial evolution is bound by our knowledge for cellular organization and processes, as well as the value of the various evolutionary parameters. As microbial populations and environmental context may vary considerably, an accurate estimation of the model parameters is both necessary for biological realism and difficult to achieve.

Ultimately, the predictive ability of a model is evaluated through comparison to experimental observations. In the case of microbial evolution, there are some examples where theory, computational simulation, and biological observation complement nicely each other and present a compelling case on the emergence of complex microbial behaviours.[94] In other cases however, it is notoriously hard to validate computational predictions, since the experimental validation necessary to achieve this is prohibitingly time-consuming.

## Conclusion remarks and future directions

In the past decades, there has been a paradigm shift on how we perceive microbial life. Initially regarded as passive life forms with minimal interaction with their environment, microbes have evolved intricate genetic, proteomic and metabolomic programs that enable them to exhibit sophisticated behavioural repertoires. Extensive work has been done at characterizing these behaviours that are both individual (stochastic switching,[116–118] persistence,[119] bet-hedging,[120] anticipation[94,121]) and social (symbiosis and cooperation,[122–124] defection,[125] biofilm formation[126,127] and parasitism,[128] among others[129]).

Despite these efforts, our understanding of the dynamics and emergence of these behaviors during evolution is very limited. Gathering hints on why organisms evolved certain processes is challenging, since we usually are oblivious of most environmental parameters in their habitats. Laboratory evolution of microorganisms provides a more controllable way to observe *in vivo* evolution and it has considerably increased our understanding of microbial physiology and evolutionary processes. However, ALE experiments are still bound by the slow timescales in which evolution operates, and our limited ability to record, process, and store biological information through measurements and experimentation.[7,130–133] Mathematical multi-scale modeling and supercomputer simulation can overcome these shortcomings, and given sufficient biological realism and computational resources, we can build a framework that can serve as a test bench for generating and testing evolutionary hypotheses.

For this to happen, there are several conditions that need to be met. First, we have to develop integrative and multi-scale models that incorporate tightly coupled processes from a wide range of sub-cellular, cellular and evolutionary functions. This goal calls for models that have balanced biological resolution and simplifying assumptions. This can be achieved by (a) avoiding the inclusion in the model of processes and parameters that do not play a significant role for the phenomena we are interested, (b) providing realistic parameter ranges to the model that stem from experimental measurements, (c) refraining from

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

oversimplifying the model since this would introduce bias to the simulation and may conceal system-level phenomena, (d) identifying the optimal level of abstraction for the temporal and spatial timescales that are involved in the phenomena that we study. From a technical standpoint, we have to develop scalable algorithms and flexible data structures capable to capture disparate objects (molecular species, biological networks, organisms, populations, environments). The code should be freely available and the models should adhere to modeling standards (*e.g.* SBML-compatible[134,135]). A simulation framework that is modular, architecture-independent, has an intuitive visualization interface, and can work both in standard computer systems as well as in parallel supercomputer systems for large scale runs would be useful both to the research and teaching community. Recently, there has been work performed towards the direction to integrate high-performance computing, multi-scale modeling and experimental validation to study complex phenomena and evolutionary dynamics in microbial evolution.[92–94,115] Attempts such as the above have the potential to elucidate biological dynamics through an iterative procedure of modeling-validation-refinement with far-fetching applications in the various biotechnological and medical fields.

## Abbreviations

| | |
|---|---|
| ALE | adaptive laboratory evolution |
| DFE | distribution of fitness effects |
| HGT | horizontal gene transfer |
| SNP | single nucleotide polymorphism |

## Acknowledgements

This work was supported by the UC Davis Opportunity Fund.

## Notes and references

1 C. Darwin, *On the Origin of Species*, John Murray, London, 1859.
2 S. H. Lee, J. H. J. van der Werf, B. J. Hayes, M. E. Goddard and P. M. Visscher, *PLoS Genet.*, 2008, **4**, 11.
3 A. Gardner and W. Zuidema, *Evolution*, 2003, **57**, 1448–1450.
4 J. Draghi and G. R. Wagner, *Evolution*, 2008, **62**, 301–315.
5 S. Ciliberti, O. C. Martin and A. Wagner, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 13591–13596.
6 A. Wagner, *Proc. R. Soc. B*, 2008, **275**, 91–100.
7 J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski and J. F. Kim, *Nature*, 2009, **461**, 1243–1247.
8 T. M. Conrad, N. E. Lewis and B. O. Palsson, *Mol. Syst. Biol.*, 2011, **7**, 509.
9 Z. P. Cakar, U. O. S. Seker, C. Tamerler, M. Sonderegger and U. Sauer, *FEMS Yeast Res.*, 2005, **5**, 569–578.
10 D. Stanley, S. Fraser, P. J. Chambers, P. Rogers and G. A. Stanley, *J. Ind. Microbiol. Biotechnol.*, 2010, **37**, 139–149.
11 T. F. Cooper, S. K. Remold, R. E. Lenski and D. Schneider, *PLoS Genet.*, 2008, **4**, e35.
12 E. Toprak, A. Veres, J. B. Michel, R. Chait, D. L. Hartl and R. Kishony, *Nat. Genet.*, 2012, **44**, 101–105.
13 R. Sanjuan, *Philos. Trans. R. Soc., B*, 2010, **365**, 1975–1982.
14 J. B. Peris, P. Davis, J. M. Cuevas, M. R. Nebot and R. Sanjuan, *Genetics*, 2010, **185**, 603–609.
15 R. Sanjuan, A. Moya and S. F. Elena, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 8396–8401.
16 P. Domingo-Calap, J. M. Cuevas and R. Sanjuan, *PLoS Genet.*, 2009, **5**, 7.
17 S. F. Elena and R. Sanjuan, *BMC Evol. Biol.*, 2008, **8**, 9.
18 O. S. Soyer and S. Bonhoeffer, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 16337–16342.
19 R. A. Goldstein and O. S. Soyer, *PLoS Comput. Biol.*, 2008, **4**, 7.
20 A. Crombach and P. Hogeweg, *PLoS Comput. Biol.*, 2008, **4**, 13.
21 D. J. Jenkins and D. J. Stekel, *J. Mol. Evol.*, 2010, **70**, 215–231.
22 J. W. Drake, B. Charlesworth, D. Charlesworth and J. F. Crow, *Genetics*, 1998, **148**, 1667–1686.
23 H. Ochman, S. Elwyn and N. A. Moran, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 12638–12643.
24 M. Lynch, *Trends Genet.*, 2010, **26**, 345–352.
25 B. A. Bridges, *Philos. Trans. R. Soc., B*, 2001, **356**, 29–39.
26 P. D. Sniegowski, P. J. Gerrish and R. E. Lenski, *Nature*, 1997, **387**, 703–705.
27 H. Goodarzi, B. D. Bennett, S. Amini, M. L. Reaves, A. K. Hottes, J. D. Rabinowitz and S. Tavazoie, *Mol. Syst. Biol.*, 2010, **6**, 12.
28 M. Sen, U. Yilmaz, A. Baysal, S. Akman and Z. P. Cakar, *Antonie Van Leeuwenhoek*, 2011, **99**, 825–835.
29 T. M. Conrad, A. R. Joyce, M. K. Applebee, C. L. Barrett, B. Xie, Y. Gao and B. O. Palsson, *Genome Biol.*, 2009, **10**, R118.
30 C. Herring, A. Raghunathan, C. Honisch, T. Patel, K. Applebee, A. Joyce, T. Albert, F. Blattner, D. van den Boom, C. Cantor and B. Palsson, *Nat. Genet.*, 2006, **38**, 1406–1412.
31 T. Kishimoto, L. Iijima, M. Tatsumi, N. Ono, A. Oyake, T. Hashimoto, M. Matsuo, M. Okubo, S. Suzuki, K. Mori, A. Kashiwagi, C. Furusawa, B.-W. Ying and T. Yomo, *PLoS Genet.*, 2010, **6**, e1001164.
32 O. Tenaillon, A. Rodriguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long and B. S. Gaut, *Science*, 2012, **335**, 457–461.
33 H. Bachmann, M. J. C. Starrenburg, D. Molenaar, M. Kleerebezem and J. Vlieg, *Genome Res.*, 2012, **22**, 115–124.
34 G. I. Lang, D. Botstein and M. M. Desai, *Genetics*, 2011, **188**, 647–661.
35 S. E. Finkel and R. Kolter, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 4023–4027.
36 A. M. Leroi, A. F. Bennett and R. E. Lenski, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 1917–1921.
37 M. M. Riehle, A. F. Bennett and A. D. Long, *Physiol. Biochem. Zool.*, 2005, **78**, 299–315.
38 B. S. Hughes, A. J. Cullum and A. F. Bennett, *Evolution*, 2007, **61**, 1725–1734.
39 R. Milo, P. Jorgensen, U. Moran, G. Weber and M. Springer, *Nucleic Acids Res.*, 2010, **38**, D750–D753.
40 R. Miralles, P. J. Gerrish, A. Moya and S. F. Elena, *Science*, 1999, **285**, 1745–1747.
41 J. De Visser and D. E. Rozen, *J. Evol. Biol.*, 2005, **18**, 779–788.
42 J. de Visser, C. W. Zeyl, P. J. Gerrish, J. L. Blanchard and R. E. Lenski, *Science*, 1999, **283**, 404–406.
43 J. H. Gillespie, *Evolution*, 2001, **55**, 2161–2169.
44 Y. Kim and H. A. Orr, *Genetics*, 2005, **171**, 1377–1386.
45 P. J. Gerrish and R. E. Lenski, *Genetica*, 1998, **102–103**, 127–144.
46 K. C. Kao and G. Sherlock, *Nat. Genet.*, 2008, **40**, 1499–1504.
47 I. M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muniz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus and P. D. Karp, *Nucleic Acids Res.*, 2011, **39**, D583–D590.
48 S. Gama-Castro, H. Salgado, M. Peralta-Gil, A. Santos-Zavaleta, L. Muniz-Rascado, H. Solano-Lira, V. Jimenez-Jacinto, V. Weiss, J. S. Garcia-Sotelo, A. Lopez-Fuentes, L. Porron-Sotelo, S. Alquicira-Hernandez, A. Medina-Rivera, I. Martinez-Flores, K. Alquicira-Hernandez, R. Martinez-Adame, C. Bonavides-Martinez, J. Miranda-Rios, A. M. Huerta, A. Mendoza-Vargas, L. Collado-Torres, B. Taboada, L. Vega-Alvarado, M. Olvera, L. Olvera, R. Grande, E. Morett and J. Collado-Vides, *Nucleic Acids Res.*, 2011, **39**, D98–D105.
49 S. F. Elena, L. Ekunwe, N. Hajela, S. A. Oden and R. E. Lenski, *Genetica*, 1998, **102–103**, 349–358.
50 R. Woods, D. Schneider, C. L. Winkworth, M. A. Riley and R. E. Lenski, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 9107–9112.
51 M. M. Riehle, A. F. Bennett, R. E. Lenski and A. D. Long, *Physiol. Genomics*, 2003, **14**, 47–58.
52 S. C. Sleight, C. Orlic, D. Schneider and R. E. Lenski, *Genetics*, 2008, **180**, 431–443.
53 D. Alcantara-Diaz, M. Brena-Valle and J. Serment-Guerrero, *Mutagenesis*, 2004, **19**, 349–354.

 This journal is © The Royal Society of Chemistry 2012

Published on 18 September 2012 on http://pubs.rsc.org | doi:10.1039/C2IB20095C
Downloaded by RSC Internal on 11 December 2012

54 S. Atsumi, T.-Y. Wu, I. M. P. Machado, W.-C. Huang, P.-Y. Chen, M. Pellegrini and J. C. Liao, *Mol. Syst. Biol.*, 2010, **6**, 449.
55 P. Charusanti, T. M. Conrad, E. M. Knight, K. Venkataraman, N. L. Fong, B. Xie, Y. Gao and B. O. Palsson, *PLoS Genet.*, 2010, **6**, e1001186.
56 D.-H. Lee and B. O. Palsson, *Appl. Environ. Microbiol.*, 2010, **76**, 6327–6327.
57 D. B. Berry, Q. Guan, J. Hose, S. Haroon, M. Gebbia, L. E. Heisler, C. Nislow, G. Giaever and A. P. Gasch, *PLoS Genet.*, 2011, **7**, e1002353.
58 Z. P. Cakar, C. Alkim, B. Turanli, N. Tokman, S. Akman, M. Sarikaya, C. Tamerler, L. Benbadis and J. M. Francois, *J. Biotechnol.*, 2009, **143**, 130–138.
59 H. K. Huse, T. Kwon, J. E. A. Zlosnik, D. P. Speert, E. M. Marcotte and M. Whiteley, *mBio*, 2010, **1**, 8.
60 T. Ferenci, *Adv. Microb. Physiol.*, 2008, **53**, 169–229.
61 Z. Wang, M. Gerstein and M. Snyder, *Nat. Rev. Genet.*, 2009, **10**, 57–63.
62 P. J. Park, *Nat. Rev. Genet.*, 2009, **10**, 669–680.
63 R. De Smet and K. Marchal, *Nat. Rev. Microbiol.*, 2010, **8**, 717–729.
64 M. Hecker, S. Lambeck, S. Toepfer, E. van Someren and R. Guthke, *Biosystems*, 2009, **96**, 86–103.
65 R. A. Fisher, *The genetical theory of natural selection*, 1930.
66 G. R. Price, *Nature*, 1970, **227**, 520–521.
67 S. Leibler and E. Kussell, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 13183–13188.
68 J. A. Draghi, T. L. Parsons, G. P. Wagner and J. B. Plotkin, *Nature*, 2010, **463**, 353–355.
69 J. H. Gillespie, *Theor. Popul. Biol.*, 1983, **23**, 202–215.
70 J. H. Gillespie, *Evolution*, 1984, **38**, 1116–1129.
71 A. Ullrich, M. Rohrschneider, G. Scheuermann, P. F. Stadler and C. Flamm, *Artificial Life*, 2011, **17**, 87–108.
72 G. Beslon, D. P. Parsons, J. M. Pena, C. Rigotti and Y. Sanchez-Dehesa, *Intell. Data Anal.*, 2010, **14**, 173–191.
73 Y. Sanchez-Dehesa, D. Parsons, J. M. Pena and G. Beslon, *Mathematical Modelling of Natural Phenomena*, 2008, **3**, 27–66.
74 C. Adami, *Nat. Rev. Genet.*, 2006, **7**, 109–118.
75 C. Ofria and C. O. Wilke, *Artificial Life*, 2004, **10**, 191–229.
76 G. Karlebach and R. Shamir, *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 770–780.
77 H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A. L. Barabasi, *Nature*, 2000, **407**, 651–654.
78 A. Wagner and D. A. Fell, *Proc. R. Soc. B*, 2001, **268**, 1803–1810.
79 A. L. Barabasi and Z. N. Oltvai, *Nat. Rev. Genet.*, 2004, **5**, 101–113.
80 H. De Jong, *J. Comput. Biol.*, 2002, **9**, 67–103.
81 S. Rosenfeld, *Prog. Biophys. Mol. Biol.*, 2011, **106**, 400–409.
82 N. Tenazinha and S. Vinga, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2011, **8**, 943–958.
83 H. Bolouri and E. H. Davidson, *Bioessays*, 2002, **24**, 1118–1129.
84 P. Smolen, D. A. Baxter and J. H. Byrne, *Neuron*, 2000, **26**, 567–580.
85 S. A. Kauffman, *J. Theor. Biol.*, 1969, **22**, 437–467.
86 A. P. Quayle and S. Bullock, *J. Theor. Biol.*, 2006, **238**, 737–753.
87 J. Draghi and G. P. Wagner, *J. Evol. Biol.*, 2009, **22**, 599–611.
88 O. S. Soyer, T. Pfeiffer and S. Bonhoeffer, *J. Theor. Biol.*, 2006, **241**, 223–232.
89 T. Reil, *Advances in Artificial Life, Proceedings*, 1999, **1674**, 457–466.
90 D. J. Jenkins and D. J. Stekel, *Artificial Life*, 2009, **15**, 259–291.
91 F. Hayot and C. Jayaprakash, *J. Theor. Biol.*, 2005, **234**, 133–143.
92 V. Mozhayskiy and I. Tagkopoulos, *BMC Bioinf.*, 2012, **13**, S10.
93 V. Mozhayskiy and I. Tagkopoulos, *BMC Bioinf.*, 2012, **13**, S13.
94 I. Tagkopoulos, Y. C. Liu and S. Tavazoie, *Science*, 2008, **320**, 1313–1317.
95 A. Krishnan, M. Tomitaa and A. Giuliani, *Phys. A*, 2008, **387**, 2170–2186.
96 P. J. Ingram, M. P. H. Stumpf and J. Stark, *BMC Genomics*, 2006, **7**, 12.
97 N. Kashtan, E. Noor and U. Alon, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 13711–13716.
98 P. Francois and V. Hakim, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 580–585.
99 P. S. Swain, M. B. Elowitz and E. D. Siggia, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12795–12800.
100 D. T. Gillespie, *J. Comput. Phys.*, 1976, **22**, 403–434.
101 M. B. Elowitz, A. J. Levine, E. D. Siggia and P. S. Swain, *Science*, 2002, **297**, 1183–1186.
102 T. B. Kepler and T. C. Elston, *Biophys. J.*, 2001, **81**, 3116–3136.
103 D. J. Wilkinson, *Nat. Rev. Genet.*, 2009, **10**, 122–133.
104 S. Ghosh, P. Ghosh, K. Basu, S. K. Das and S. Daefler, in Proceedings. 39th Annual Simulation Symposium, IEEE Computer Society, Editon edn., 2006, p. 8.
105 D. Chandran, F. T. Bergmann and H. M. Sauro, *J. Biol. Eng.*, 2009, **3**, 19.
106 N. E. Lewis, H. Nagarajan and B. O. Palsson, *Nat. Rev. Microbiol.*, 2012, **10**, 291–305.
107 M. E. Tsuda and M. Kawata, *PLoS Comput. Biol.*, 2010, **6**, 22.
108 S. Wright, *Proc Sixth Internat Congr Genetics Ithaca New York*, 1932, **1**, 356–366.
109 T. T. Kibota and M. Lynch, *Nature*, 1996, **381**, 694–696.
110 M. Kimura, *Nature*, 1968, **217**, 624–626.
111 A. Eyre-Walker and P. D. Keightley, *Nat. Rev. Genet.*, 2007, **8**, 610–618.
112 H. A. Orr, *Genetics*, 2003, **163**, 1519–1526.
113 M. Imhof and C. Schlotterer, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 1113–1117.
114 J. R. Roth, *Curr. Biol.*, 2010, **20**, R15–R17.
115 V. Mozhayskiy, R. Miller, K.-L. Ma and I. Tagkopoulos, *TeraGrid'11*, Salt Lake City, UT, edn, 2011.
116 M. Acar, J. T. Mettetal and A. van Oudenaarden, *Nat. Genet.*, 2008, **40**, 471–475.
117 M. Kaern, T. C. Elston, W. J. Blake and J. J. Collins, *Nat. Rev. Genet.*, 2005, **6**, 451–464.
118 R. Losick and C. Desplan, *Science*, 2008, **320**, 65–68.
119 N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik and S. Leibler, *Science*, 2004, **305**, 1622–1625.
120 J.-W. Veening, E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers and L. W. Hamoen, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 4393–4398.
121 A. Mitchell, G. H. Romano, B. Groisman, A. Yona, E. Dekel, M. Kupiec, O. Dahan and Y. Pilpel, *Nature*, 2009, **460**, 220–224.
122 A. S. Griffin, S. A. West and A. Buckling, *Nature*, 2004, **430**, 1024–1027.
123 S. P. Diggle, A. S. Griffin, G. S. Campbell and S. A. West, *Nature*, 2007, **450**, U411–U417.
124 L. V. Hooper, *Nat. Rev. Microbiol.*, 2009, **7**, 367–374.
125 J. Gore, H. Youk and A. van Oudenaarden, *Nature*, 2009, **459**, 253–256.
126 J. S. Webb, M. Givskov and S. Kjelleberg, *Curr. Opin. Microbiol.*, 2003, **6**, 578–585.
127 L. Hall-Stoodley and P. Stoodley, *Cell. Microbiol.*, 2009, **11**, 1034–1043.
128 D. Blumenthal, C. E. Mitchell, P. Pysek and V. Jarosik, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 7899–7904.
129 S. A. West, A. S. Griffin, A. Gardner and S. P. Diggle, *Nat. Rev. Microbiol.*, 2006, **4**, 597–607.
130 T. F. Cooper and R. E. Lenski, *BMC Evol. Biol.*, 2010, 10.
131 M. T. Stanek, T. F. Cooper and R. E. Lenski, *BMC Evol. Biol.*, 2009, **9**, 302.
132 Z. D. Blount, C. Z. Borland and R. E. Lenski, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 7899–7906.
133 A. F. Bennett and R. E. Lenski, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 8649–8654.
134 L. P. Smith, F. T. Bergmann, D. Chandran and H. M. Sauro, *Bioinformatics*, 2009, **25**, 2452–2454.
135 Y. Cai, B. Hartnett, C. Gustafsson and J. Peccoud, *Bioinformatics*, 2007, **23**, 2760–2767.
136 R. Phillips and R. Milo, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 21465–21471.
137 Y. Deng and S. S. Fong, *J. Biol. Chem.*, 2011, **286**, 39958–39966.
138 H. A. Wichman and C. J. Brown, *Philos. Trans. R. Soc., B*, 2010, **365**, 2495–2501.
139 I. Tagkopoulos, *Emergence of predictive capacity within microbial genetic networks*, PhD thesis, Princeton University, 2008, ISBN: 978-0-549-74801-4.
140 C. O. Wilke, J. L. Wang, C. Ofria, R. E. Lenski and C. Adami, *Nature*, 2001, **412**, 331–333.
141 C. W. Nelson and J. C. Sanford, *Theor. Biol. Med. Modell.*, 2011, **8**, 17.
142 G. Beslon, D. P. Parsons, Y. Sanchez-Dehesa, J. M. Pena and C. Knibbe, *Biosystems*, 2010, **102**, 32–40.