Endless evolutionary paths to Virtual Microbes

Thomas Cuypers[1] and Paulien Hogeweg[1]

[1]Theoretical Biology group, Utrecht University, padualaan 8, 3584CH Utrecht, the Netherlands
t.d.cuypers@uu.nl

Abstract

How long term evolution shapes genome function and size is unknown. Factors that will likely play a role are population size and the environment that organism live in. In addition, the emergence of ecological interactions in spatially structured environments is expected to feed back on organism fitness, changing the functional demands on the genome. Evolution experiments in laboratory settings provide insights into the remarkable abilities of micro-organisms to adapt, showing high evolvability through genomic and physiological changes as well as evolving novel ecological interactions. Unravelling the underlying mechanisms of this rapid adaptation remains challenging.

Here, we develop a computational model of evolving microbes to gain insight in long term patterns of genome evolution and the evolution of evolvability. The virtual microbes take up nutrients from their environment and convert them to useful building blocks. Populations discover novel metabolic and regulatory functions through rare horizontal gene transfer events. Regulation of protein expression can serve to tune metabolic fluxes, increasing their production rate, and avoid costly protein expression as well as toxic build up of internal metabolites. Modelling evolution in silico allows storing a perfect 'fossil record' of mutations in the line of descent.

Preliminary experiments show continued incorporation of novel metabolic and transport genes into the metabolic networks on evolutionary timescales. Even after reaching the maximum limit of resources consumed in the environment, parts of the metabolic and transport network remained in flux, showing that no single end-point of evolution is expected.

Despite being energetically costly to express, in some populations TFs evolve strongly connected regulatory networks, suggesting a role in optimizing production and avoiding toxicity. Interestingly, when experiments were run in larger populations, the fraction of TFs seemed to be lower. This result mimics a theoretical prediction that reduced efficacy of selection in small populations allows accumulation of (slightly) costly genomic features.

Introduction

One of the big questions in evolutionary biology is much of the genome of organisms has been shaped by adaptation and how much by neutral events. A related question that could be crucial in finding an answer to the former is to which extent genomes reflect the need for organisms to generate variation and allow evolution of novel functions. In other words, how does the ability to evolve feed back on evolution and vice versa.

When studying the evolution of evolvability it may be especially important to consider a complex evolutionary dynamics arising from an interplay of evolution of metabolic as well as regulatory capacity in populations of interacting individuals. Interactions among microbes e.g. in microbial mats, are rampant. Often interactions cannot be avoided. Bringing together a large enough set of species with varying metabolic capabilities is almost guaranteed to result in cross-feeding interactions. It is an open question whether microbes might want to actively avoid or even encourage such commensalism on evolutionary time scales.

Specific questions that could be addressed with such a model are: How are genome and regulatory network organization shaped under various degrees of species interdependence in terms of cross-feeding interactions, fluctuating environmental conditions or addition of novel metabolic compounds. What are the roles of various mutational operators as well as HGT during adaptation towards novel environmental conditions? What is the role of selfish elements such as transposons or various forms of viral particles in enabling long term evolvability within lineages.

Virtual Microbes Model

The Virtual Microbes model is an individual based evolutionary model that incorporates a structured genome encoding genes involved in metabolism, transportation and transcription regulation. It extends the Virtual Cell model that has been used to study dynamics of genome evolution, by adding ecological interactions and a richer metabolismNeyfakh et al. (2006); Cuypers and Hogeweg (2012). Microbes interact with each other by competing for metabolites and an opportunity to reproduce in a spatial environment. Microbes shape their local environment by enzymatically converting metabolites through internal metabolism and active transport as well as passive diffusion of metabolites over the cell membrane. Changes in local, external metabolite con-

centrations due to cellular metabolism propagate in the environment by diffusion. In addition, influx rates of the various metabolites change stochastically over time. The spatial structure and environmental fluctuations lead to strong temporal and spatial variation in environmental conditions, creating variable selection pressures, often demanding fast adaptive responses. Such demand for rapid change may leave evolutionary signatures in the genome, which could be selected for increased evolvabilityCrombach and Hogeweg (2007, 2008). Alternatively, evolution of population diversity could lead to the emergence of a structured ecosystem-Boer and Hogeweg (2007); de Boer and Hogeweg (2010); Le Gac et al. (2012), potentially foregoing the need for rapid adaptation in single lineages.

The reproductive success of microbes depends on their ability to obtain essential building blocks, either directly from the environment or by producing them from precursor metabolites through metabolism. For active transport of metabolites microbes require genes that encode specialized transporter enzymes. Transport is costly, requiring the consumption of special energy carrier molecules. Therefore, successful microbes should typically partition metabolic fluxYou et al. (2013) both to the production of building block as well as energy molecules. Active transport can, however, increase the risk of accumulating toxic levels of internal resources. Toxicity, measured as the time integral of metabolite concentration above their safe levels, increases the death rate of microbes.

Microbes can employ several potential strategies to reduce the risk of toxic metabolite levels. One of these possibilities is to incorporate exporter enzymes that reduce metabolite concentrations that are at toxic levelsRuiz and Levy (2014). Secondly, connecting internal metabolites to each other via metabolic pathways can ensure that any excess concentrations can be dissipated to downstream parts in the pathways. Thirdly, the expression rates of enzymes and transporters can be regulated by transcription factors (TFs), binding to the genesProulx and Smiley (2010). Transcription factors encode a binding motif that can recognize sequences in the operators regions of any gene. This binding is fuzzyPayne and Wagner (2014), depending on a partial bit string match, with the accuracy of the match modifying the strength of the binding interaction. TFs can interact with specific ligand molecules that modulate the regulatory effect that they have on the transcription rate of downstream genes. This constitutes a sensing mechanism for internal concentrations of the ligand molecules and responding to changes in concentration by adjusting the gene expression state of genes regulated by ligand binding TFs. During the evolution of microbes the different strategies can be expected to exist concurrently and interact.

Mutations happen at different genomic levels. The genome can consist of multiple chromosomes, that can independently undergo duplication, deletion, fission or fusion to another chromosome. At the subchromosome level, stretches of genes can be duplicated in tandem, deleted, translocated and become inverted. Mutations at the gene level affect enzymatic parameters, binding motifs and operator sequences. Importantly, transporters can undergo a mutation that changes the direction of transport, allowing enzymes that import a specific metabolite to become exporters of the metabolite and vice versa. Mutations can, however, not alter the basic reaction type of enzymes or transporters nor the type of ligand molecules recognized by TFs. Instead, novel enzymes can be 'discovered' at a low rate through horizontal gene transfer (HGT) from an external sourceKoonin and Wolf (2012); Polz et al. (2013). Microbes can also acquire genomic material from close by neighbours at a rate that is much faster than that of external HGT. The explicit genome structure combined with mutational operators working at different genomic scales provide relatively many degrees of freedom to the evolutionary process. Earlier work on the Virtual Cell model has shown how successfully adapting populations typically experience a surge in genomic material early in evolution, through a burst of large scale duplication eventsCuypers and Hogeweg (2012). The largely neutral, early genome inflation crucially increases the degrees of freedom for adaptive evolution by providing the raw genomic material for innovation.

To what extent genomes increase in size depends on explicit and implicit costs of genome size. An implicit cost of maintaining a large genome is the mutational load it imposes. Because the rates of most types of mutations is proportional to the number of genes, the load in terms of the per generation amount of deleterious mutations increases with genome size. In addition, protein expression bears an explicit energetic cost for organisms. Here we model it as a small rate of energy molecule consumption to express genes. The two costs are fundamentally different in the sense that the costs of expression can be mediated by transcription regulation, whereas mutational load is intrinsic to having more genomic material. With respect to these small costs of genome size, population genetics theory predicts that selection to reduce these costs is more effective in larger populations. Under this prediction, small populations would be more likely evolve larger genomesLynch and Conery (2003).

scaling of competition

In order to maintain a functional selection coefficient as the population evolves higher levels of production we scale the chance of reproduction during competition in the local neighbourhood. The scaling factor depends on the historical maximum of the population median production value. As a consequence, populations that at any point in time reach a new production maximum are subsequently subjected to harsher conditions when trying to reproduce locally. On the one hand this ensures continued selection for increasing the efficiency of building block production. On the other hand,

however, population numbers may start to fluctuate more, increasing the risk of extinction. Moreover, stronger bottlenecking can have consequences for the evolutionary dynamics at the phenotypic and genotypic level. For example, it could negatively affect metabolic diversity in the population, but also lead to increased drift, allowing genome size to expand.

evolutionary simulations

In the current preliminary research we ran simulations with different maximum population sizes and studied the effects of continuing evolved populations under various mutational regimes. For initial simulations we used maximum population sizes of 1024, 2500 and 4900 individuals. These were evolved for 100000 time steps or approximately 5000 generations. Subsequently simulations of the evolved 1024 and 2500 populations were continued under regimes with, respectively, no mutations, only stretch deletions or no external HGT for another 100000 time steps. In addition, we continued evolution unchanged until extinction.

The metabolic universe was initialized with 16 metabolites, two of which are essential building blocks and one designated as an energy carrier. During all simulations all metabolites except building blocks and the energy carrier fluxed into the environment at a a stochastically changing rate, taking values that differ up to two orders of magnitude. The degradation rate of external metabolites was chosen in the order of the average influx rate. Toxic levels for internal molecules were chosen to be within one order of magnitude upwards of ambient concentrations. The levels were chosen to lead to near zero toxicity when metabolites only pass into cells by passive diffusion. However, during evolution, a varying fraction of the population suffers from toxicity, as selection to increase production can lead internal concentrations of certain metabolites to (transiently) become higher through the addition or increased expression rates of metabolic enzymes and transporters.

To create starting populations a single individual was initialized and then cloned to fill up the population. The individual has a genome capable of synthesizing the two essential building blocks and the energy carrier and a starting set of TFs that can sense a subset of the metabolites produced and consumed in the metabolism. The wiring pattern of TFs to operators is randomly initialized.

open-endedness

The Virtual Microbes model couples environmental variation and spatial structure with an explicit genome organisation. The model explores the impact of continuous environmental variability on the evolutionary dynamics of metabolic and gene regulatory networks (GRN). An important source of innovation is a low rate of external HGT. In early evolution, this influx of externally provided novel genes leads to an expansion of the metabolic network. In all but the smallest populations, adsorption of novel metabolic genes continues until the full set of environmentally available metabolites can be converted by enzymes and channelled into the metabolism. Because all metabolites have a small passive rate of diffusion over the membrane, it appears beneficial to incorporate metabolic steps to channel as many metabolites as possible into the existing metabolic pathways that provide growth and energy.

The expansion of metabolic capacities is less pronounced for transportation. This is likely due to the energetic cost of transport reactions. Producing energy requires diverting metabolic flux away from producing building blocks. In addition, increasing the influx of some metabolites through active transport can lead to the build-up of internal metabolite concentrations to toxic levels. Interestingly, import capacity typically keeps increasing during long term evolution, after the metabolic capabilities have saturated.

The Virtual Microbes model allows for evolution of metabolic diversity and crossfeeding at the population level. We sought to measure diversity of metabolism by looking at the differences in metabolic capacities between subpopulations. First, we defined an individuals metabolic type by composing the unique set of metabolites consumed, produced, imported and exported. Next, we computed, for the five most abundant metabolic types in the population, the fractions of overlap between the respective capabilities. Differences are expected to increase when novel metabolic genes become established in a subset of the population. When new enzymes sweep to complete fixation, however, the difference measure should decreases again. Moreover, as more enzyme types become fixated in the whole population, subsequent novel genes establishing in the population yield proportionally smaller differences. As an overall trend population diversity in consumed and produced metabolites tends to very low levels, suggesting that the evolved metabolic network converges on a local optimum with selection acting strongly against deviation from the optimum. In contrast, diversity in actively transported metabolites remains at much higher levels. It is most pronounced in the set of exported metabolites. This may in fact reflect a neutral process in which exporters continuously evolve from more conserved importers by direction changing mutations. Alternatively, selection could be acting transiently to evolve exporters of a metabolites that are temporarily abundant and reach toxic internal concentrations. Currently, we have not yet been able to distinguish between both scenarios. However, the experiment where all mutations are stopped provided an indication that the neutral explanation may not fully account for the presence of exporter variation, which will be discussed in more detail below.

Crossfeeding has been observed to evolve rapidly and repeatedly in evolutionary experiments initiated with isogenic populationsGac et al. (2012). Theoretical approaches also suggest that, rather than being a delicate balance between

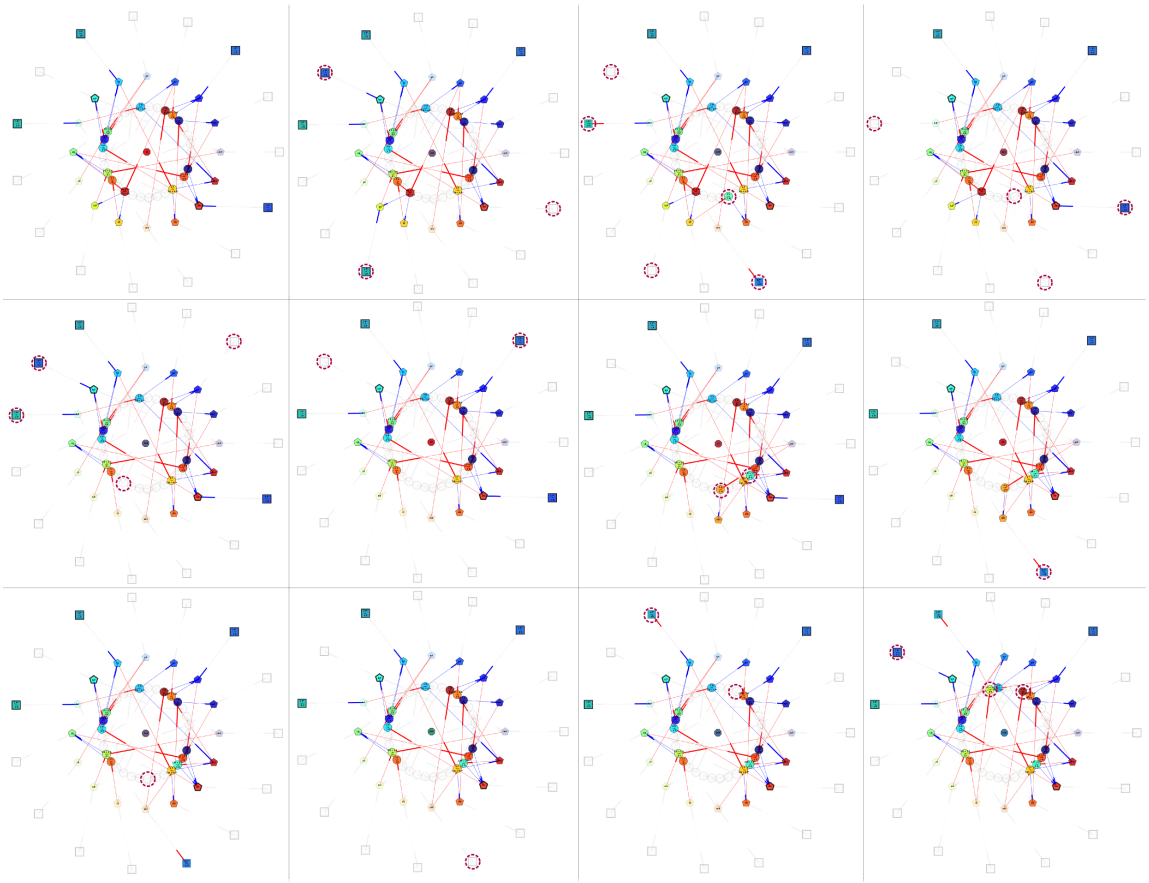

Figure 1: Continuous changes in metabolism over time. Metabolomes of best individuals in the population at 10000 timestep intervals, starting at t=100000. Outer circle: transporters, middle circle: metabolites, inner circle: enzymes. Red arrows indicate consumption and export of metabolites, blue arrows indicate production and import. Additions and deletions in the metabolic network are marked by dashed circles.

mutualistic partners, it is often an expected generic property of microbes sharing a common environment and having different capabilities for using metabolites and producing them as by productsKlitgord and Segre (2010). To probe for potential crossfeeding in our simulations we compare the metabolic capacities of microbes that are neighbours on the grid. The first measure detects the simultaneous presence in local neighbourhoods of microbes that can produce and export a metabolite with those that import and consume that same metabolite. Large variation exists between simulations in the extent to which local metabolisms evolve this type of crossfeeding, ranging from an average of 0 to more than 3 metabolites potentially exchanged in local neighbourhoods. It must be noted, however, that crossfeeding could be detected between closely related individuals that have identical capacities for both producing, exporting, importing and consuming a metabolite. When a stricter measure is used, requiring that an individual does not itself produce an imported and consumed metabolite, no significant evolution of crossfeeding is observed. From our preliminary research, so far we were not able to conclude that crossfeeding is a selected ecosystem property.

Different views may exist on what constitutes open-endedness in the context of evolutionary computer simulations or evolution in general. Simply put, biological evolution is a prototype of open-endedness. The history of life on earth is testimony to the unlimited capacity of evolution to keep adapting to novel conditions, generate new forms, and generating new levels of selection, as evidenced by, but not limited to the major transitions in evolutionSmith and Szathmary (1997); Koonin (2007). Computational models of evolution will be hard pressed to show the same 'unexpected' escape from 'seemingly' fixed reaction structures, if only because it is intrinsically hard to anticipate which lower level properties of evolutionary systems would enable the emergence of higher levels entities for selection to act upon. Nevertheless, many examples exist, especially in the case of spatially extended models, of the impact that struc-

turing into higher level patterns has on evolutionary dynamicsde Boer and Hogeweg (2010); Takeuchi and Hogeweg (2012); Colizzi and Hogeweg (2014). In our model, the coexistence of lineages with diverse metabolic capacities and potential crossfeeding interactions can in fact be the result of higher level selection acting on the ecosystem. Taking another perspective, open-endedness may be attributed to systems that keep evolving novel adaptive structures over prolonged evolutionary timescales, for example, the evolution novel metabolic reactions and pathways. Although our model explicitly fixes the metabolic and reaction universe, this does not preclude continued evolution of metabolic pathways. Because microbes alter their local environment by metabolising, the fitness landscape keeps changing. This can result in continuous change of metabolic pathways even though no change in productivity is observed in the population (see fig.1).

population size effect

Population size is thought to play a key role in the evolution of genome size and organismal complexity. In populations with high 'effective population sizes' slightly deleterious mutations are removed more effectively. Because protein coding genes are associated with a (small) energetic cost for protein expression as well as imposing a mutational load, duplications and HGT that do not yield immediate adaptive benefit are more likely to be negatively selected in large populations. Conversely, in small populations these slightly deleterious mutations are more likely to become fixed through neutral drift, making their genomes more prone to increase in size. In a predecessor of the current model we showed that neutral increases in genome size early in evolution enhance the long term adaptive success of populationsCuypers and Hogeweg (2012). Instrumental for this result is the enhanced evolvability in genomes with large numbers of near neutral TFs that eventually evolve adaptive regulatory interactions through rewiring. Here, we wanted to test whether small populations could eventually be more successful due to neutrally drifting genome sizes that favour the evolution of adaptive gene regulation. Simulations with maximum population sizes of 1024, 2500 and 4900 were initialized by creating isogenic clone populations. We used two different genome seeds and evolved them with two different environmental seeds each. The same seed combinations were used for all different population sizes, bringing the total to 12 initial simulations. The number of simulations being too small to expect any statistically significant differences we here report our qualitative observations.

For the first aspect, genome size, we found that two of the smallest populations ranked first and second of all simulations. When the simulations with population sizes of 1024 and 2500 were continued until twice the evolutionary time, the simulations that ranked highest initially had declined in genome size and lost their first and second rank. This pattern is compatible with the idea that small populations can have faster increases in genome size. However, some of the larger populations continued to increase in genome size in the extended simulations. This could be a sign that at least part of the genome size increases are due to direct adaptive effects, for example an adaptive dosage increase, both in small and in large populations.

Metabolism in organisms is often tightly regulated, to avoid toxic buildup of intermediates or wasteful expression of unneeded enzymes or prevent branching into suboptimal metabolic pathways. Instrumental for this regulation are TFs that sense metabolite concentrations. Because regulatory interactions are randomly initialized in our simulations, TFs are unlikely to provide immediate benefits. Therefore, their retention during the initial phase of evolution depends on the balance between the speed at which beneficial mutations arrive in the GRN and the time needed for negative selection to purge the small costs imposed by their retention. As mentioned before, such negative selection is likely to be less efficient in small populations. Comparing the average node connectivityBeineke et al. (2002) of networks G using:

$$\overline{\kappa}(G) = \frac{\sum_{u,v} \kappa_G(u,v)}{\binom{n}{2}} \tag{1}$$

with $\kappa_G(u,v)$ the local node connectivity defined as the minimum number of nodes in the graph that have to be removed to disconnect nodes u and v, we found a striking difference between populations of maximally 1024 individuals and those with 2500 and 4900. The smaller populations had an average evolved connectivity of 0.29 compared to 0.036 and 0.042 in the respective larger populations. Moreover, two of the small populations with the largest genomes, had reached connectivities $> 1.$ at an intermediate evolutionary time point, whereas none of the larger populations ever had connectivity > 0.35. Interestingly, the highly connected networks for which these maxima were recorded appeared to be a transient state during evolution. When we continued the simulations beyond the initial experiment, connectivities declined for three out of four simulations. The exception was a simulation that initially had an intermediate connectivity, a small genome and low fitness within the small population set. It subsequently increased genome size and connectivity and had a jump to the maximum fitness in the set. Following the overall trend for small populations but with a delay, the connectivity, after having peaked at a value $> 1.$, decreased towards the end.

We found the highest connectivities in the populations with the largest genomes. This can partly be explained by considering that the average degree of a TF is expected to increase as the genome grows and more binding sequences become available as potential (independent) targets. Therefore, if the proportion of TFs remains constant over genome size, the average degree of all genes increases and, assuming randomly dispersed binding sequences, subsequently the

connectivity in the GRN. The small population simulations showed an increased tendency for fast genome size increases and high network connectivity. A measure of productivity showed that some large gains in this score were made after the network connectivities had peaked and that the population with the lowest historic productivity was also the least connected. However, we can not conclude that high connectivity is necessary to increase productivity, since the average productivity level was higher in the larger populations while connectivity was lower. Rather, the search for adaptive mutations is likely enhanced both by an increased sampling rate in large populations and higher evolvability of large, connected GRNs. The latter could arise more readily in smaller populations whose genomes are more prone to neutral size increases.

stopping mutations

To gain insight in the importance of mutations for remaining adapted we ran continuation experiments where mutations were (partly) stopped. One hypothesis was that of various evolved genome structures, those with weaker regulatory features would be less likely to survive when mutations stopped as less regulated genomes may depend on continuous mutations to keep readapting to the environmental changes. We did not, however, see any beneficial effect of regulation on population survival. In fact, the only population that survived to nearly the end of the simulation had a significantly lower capacity to metabolise and transport. Its production capacity was the second lowest of all eight simulations (4 at 1024 and 4 at 2500 maximum population size) considered. Due to its low import capacity it most likely avoided most of the risk of accumulating toxic levels of resources. Moreover, the low production value reached in this population means that selection pressure from competition scaling is less severe, likely avoiding large population fluctuations (see). Concluding, long term survival crucially depends on the ability to avoid toxic levels. As competition increases due to adaptive increases of metabolic capacities, being able to generate mutants that escape toxicity likely becomes increasingly important. Our expectation that a highly regulating network structure would increase population survival in the absence of adaptation by mutations could not be confirmed.

deletions only

In a variant of the protocol without mutations, we continued the same initial simulations with only deletions of gene stretches. Interestingly, in this case the survival of populations increases dramatically to 5 out of 8 simulations, compared to none when no mutations happen. All surviving populations undergo extensive genome reduction and lose part of their metabolic capacity, both in converted and transported metabolites. Strikingly, populations lose almost all regulatory interactions. Extinction in the remaining runs happens quickly after changing the mutation protocol, when genomes have not yet lost many genes and still contain about four times as many genes as the evolved surviving populations. As can be expected, none of the surviving populations has significant gains in production. An increase of the time to coalescence indicates a reduced rate of adaptive sweeps. Concluding, we find that evolving with just deletions enables significantly better survival compared to no mutations, by enabling removal of metabolic and transport enzymes that cause (temporary) buildup of toxic metabolites. Avoiding the risk of toxicity by deleting metabolic genes may subsequently reduce the need for gene regulation, leading to the loss of TFs.

no external HGT

Innovations in metabolism are possible when microbes incorporate new genes by HGT. Consequently, when we stop HGT from an external source the population can not gain novel gene types, although individual lineages may still take up new genes by HGT from their neighbours on the grid. Even though the external HGT rate was moderate at 0.002 per generation, the effect of stopping it was noticeable with respect metabolic capacity in the population. In particular, the amount of different transporters in the population declines, although much more capacity was retained compared to the deletion only mutation protocol. Curiously, two of four small populations still had significant productivity gains after stopping external HGT, although their final productivity remained below those of the other two populations that had no more gains. In contrast to the two populations with the highest productivity that retained significantly connected GRNs, those that gained productivity without becoming the best producers eventually lost all connectivity. The small population with the highest fitness retained a highly connected GRN (0.7). However, in the parallel run with continuous external HGT this population had a significant drop in connectivity while not showing any productivity increase. In sharp contrast, none of the 2500 populations retained significantly connected GRNs when external HGT was stopped, despite their much higher productivity scores. Again, we see a complex interplay between population size, genome size and connectivity and the level at which populations can exploit their environment while avoiding effects of toxicity. External HGT keeps the metabolism in a state of flux. The sampling of novel metabolic genes allows populations to increase production. The flux in metabolic pathways may however render previously evolved regulation non-functional, which may explain why the same evolved population lost regulation when evolving in unchanged conditions, while strongly consolidating its regulation when external HGT was stopped. Thus, evolution of regulation may be a two step process in which initially weak negative selection can allow connectivity to increase mostly neutrally and consolidation of the structure only happens when it remains

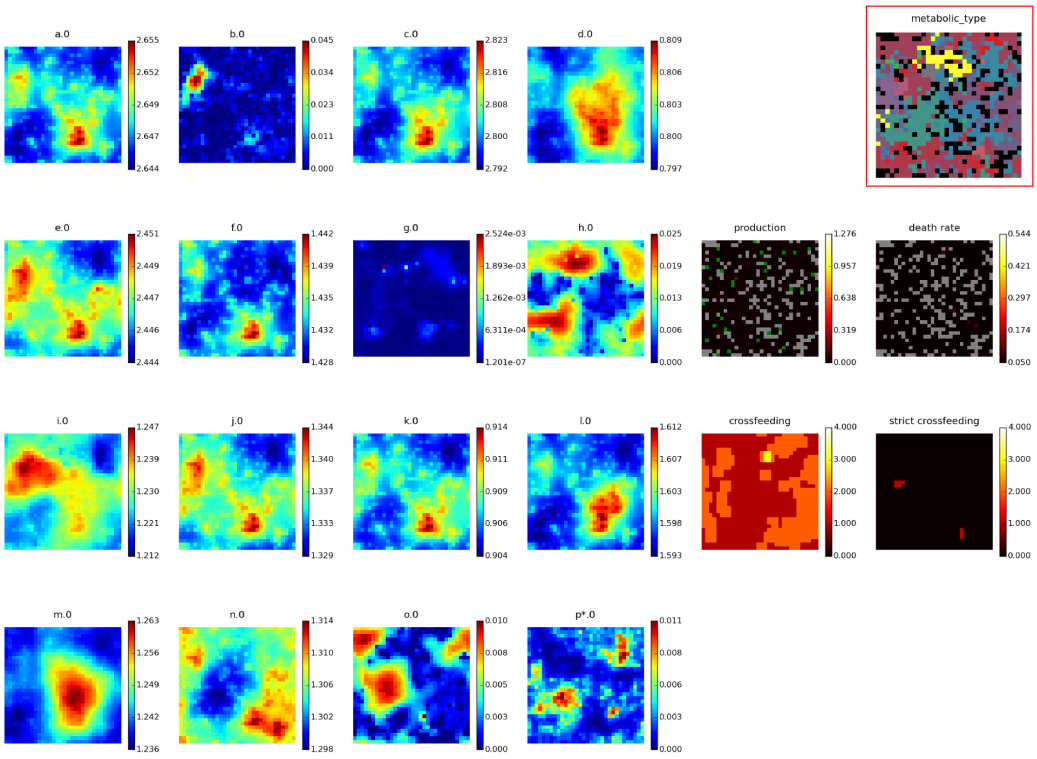

Figure 2: Snapshot of ecology. At t=150000 metabolite concentrations on the grid are plotted together with the metabolic type of individuals on the grid (red outline). An arbitrary colouring is used to indicate individuals with identical metabolic capabilities. These are defined as unique combinations of sets of metabolites consumed, produced, imported and exported, respectively, by the individual. Note that it is possible to identify patterns in metabolite concentrations that correspond to particular metabolic types.

adaptive for a prolonged evolutionary time frame.

fitness and robustness

The fitness of real organisms is a complex function of their ability, amongst others, to exploit resources faster and more efficiently than their neighbours, avoid dying from toxic effects and adapting to changes in the environment by regulation or mutation. Therefore, when our model mimics these complex demands that the environmental imposes on virtual organisms it becomes much harder to quantitatively assess their fitness compared to using a straightforward optimization criterion. Ultimately, the fittest individuals are those found in retrospect to lie on the line of descent to the future population. Individuals are subject to different selection pressures due to environmental variability in space and time. For example, increased import capacity may be adaptive when influx is low or when neighbours are also actively importing a resource but can be dangerous when environmental concentrations become higher due to increased influx or if neighbours stop uptake. In this case it can be beneficial to evolve a low affinity exporter for the same metabolite, despite its associated energetic cost, to counter the toxic buildup. We indeed see the presence of exporters in genomes on the line of descent. However, their presence is mostly transient, which could indicate that mutational changes to other parts of metabolism make their presence redundant. Evolving regulatory control of metabolism could be one other way to mediate toxic effects. But tight regulation may not even be strictly necessary, for example if reactions downstream in the metabolism become more efficient so that toxic buildup of intermediates is prevented. Indeed, we have seen that populations evolving high productivity do not necessarily maintain highly connected GRNs.

The single most important mechanism for the evolved populations to be robust was by generating adaptive mutants. This was clearly demonstrated by the 100 % extinction rate when all mutations were stopped. Even when allowing just deletion mutants, the survival rate could be dramatically increased, in striking parallel to recent experimental findingsHottes et al. (2013). In addition, populations with less metabolic and in particular transport capability were less prone to go extinct. This is in part due to competition scaling becoming steeper as populations reach higher levels of production. Steeper scaling can cause more severe population

bottlenecks and ultimately extinction when populations cannot reach the high production demand during unfavourable environmental conditions.

variability

As previously mentioned, metabolic networks showed considerable variability over evolutionary time (fig. 1). We found that this variability also extends to the population level. Distinct subpopulations with different metabolic capabilities coexist in space where they shape their local environments in different ways 2]. This can also be shown in phylogenetic trees that show distinct metabolic types persisting in evolutionary time, punctuated by adaptive sweeps and bottlenecks that reduce temporarily reduce diversity.

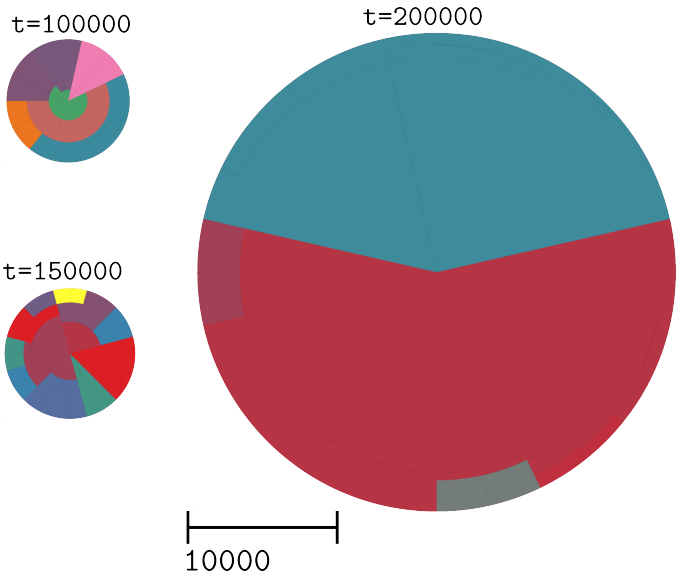

Figure 3: Phylogenetic tree representations of metabolic type evolution. At three evolutionary time points the metabolic types of ancestors are imposed on the phylogenetic tree. Note that the tree at t=150000 corresponds with the snapshot of ecology in fig. 2. At t=100000 and 150000 the tree is much shallower than for the tree at t=200000. This last time points illustrates the long term coexistence of distinct metabolic types, while the earlier time points show that high diversity can be generated at relatively short timescales.

Conclusions

Summarizing, we have seen that populations continuously adapt in response to environmental change, even when the full metabolic potential is reached. Without the ability to evolve populations rapidly go extinct. However, allowing deletions rescues most populations. Despite strong selection, variability in the population in space and correspondingly in the environment remains large over the entire evolutionary time. Population size strongly impacts the evolutionary dynamics. Small populations are are more likely to evolve large genomes with high connectivity, but do not reach the highest levels of productivity evolved in some larger populations. The most counterintuitive result was the high overall fitness reached by non regulating genomes, that must continuously adapt through duplication, deletion and HGT.

a tool to guide evolutionary thinking

The Virtual Microbes model has a complicated structure and high level of detail when compared to classical population genetics models of evolution. As a consequence, results are much harder to interpret and may sometimes look counterintuitive. Explaining the model results can even become as complicated as interpreting measurements from wet lab experimental evolution or phylogenetic studies on sequence data. An important benefit, however, of developing and using models with this level of detail is that we can develop a better intuition about how genetopic, phenotypic and ecosystem processes may interact on evolutionary timescales. In fact, gaining awareness of these mechanism during exploratory research could be at least as important as testing specific hypotheses in a small parameter space. For example, the finding of both regulated and unregulated metabolic solutions as alternative successful evolutionary solutions was an unexpected result warranting further investigation. Having a detailed record of intermediate evolutionary states will enable us to analyse the importance of the ecosystem in either solution and perform competition experiments between different solutions.

Our model provides a framework in which to test the impact of various aspects of microbial physiology and ecology on long term evolution of genome architecture. Because no single solution for evolving a virtual microbe exists, we may get evolution of evolvability almost for free.

References

Beineke, L. W., Oellermann, O. R., and Pippert, R. E. (2002). The average connectivity of a graph. *Discrete Mathematics*, 252(13):31–45.

Boer, F. K. d. and Hogeweg, P. (2007). The role of speciation in spatial coevolutionary function approximation. In *Proceedings of the 2007 GECCO conference companion on Genetic and evolutionary computation*, pages 2437–2441, London, United Kingdom. ACM.

Colizzi, E. S. and Hogeweg, P. (2014). Evolution of Functional Diversification within Quasispecies. *Genome Biol Evol*, 6(8):1990–2007.

Crombach, A. and Hogeweg, P. (2007). Chromosome Rearrangements and the Evolution of Genome Structuring and Adaptability. *Mol Biol Evol*, 24(5):1130–1139.

Crombach, A. and Hogeweg, P. (2008). Evolution of Evolvability in Gene Regulatory Networks. *PLoS Comput Biol*, 4(7).

Cuypers, T. D. and Hogeweg, P. (2012). Virtual Genomes in Flux: An Interplay of Neutrality and Adaptability Explains Genome Expansion and Streamlining. *Genome Biol Evol*, 4(3):212–229.

de Boer, F. K. and Hogeweg, P. (2010). Eco-evolutionary dynamics, coding structure and the information threshold. *BMC Evol. Biol.*, 10:361.

Gac, M. L., Plucain, J., Hindré, T., Lenski, R. E., and Schneider, D. (2012). Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with Escherichia coli. *PNAS*, 109(24):9487–9492.

Hottes, A. K., Freddolino, P. L., Khare, A., Donnell, Z. N., Liu, J. C., and Tavazoie, S. (2013). Bacterial Adaptation through Loss of Function. *PLoS Genet*, 9(7).

Klitgord, N. and Segre, D. (2010). Environments that Induce Synthetic Microbial Ecosystems. *PLoS Computational Biology*, 6(11). 00053 PMID: 21124952.

Koonin, E. V. (2007). The Biological Big Bang model for the major transitions in evolution. *Biol. Direct*, 2:21.

Koonin, E. V. and Wolf, Y. I. (2012). Evolution of microbes and viruses: a paradigm shift in evolutionary biology? *Front Cell Infect Microbiol*, 2.

Le Gac, M., Plucain, J., Hindre, T., Lenski, R. E., and Schneider, D. (2012). Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with Escherichia coli. *Proc Natl Acad Sci U S A*, 109(24):9487–9492.

Lynch, M. and Conery, J. S. (2003). The origins of genome complexity. *Science*, 302:1401–1404.

Neyfakh, A. A., Baranova, N. N., and Mizrokhi, L. J. (2006). A system for studying evolution of life-like virtual organisms. *Biol Direct*, 1:23.

Payne, J. L. and Wagner, A. (2014). The Robustness and Evolvability of Transcription Factor Binding Sites. *Science*, 343(6173):875–877.

Polz, M. F., Alm, E. J., and Hanage, W. P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics*, 29(3):170–175.

Proulx, S. R. and Smiley, M. W. (2010). The evolutionary origins of gene regulation. *J. Exp. Zool.*, 314B(4):327–340.

Ruiz, C. and Levy, S. B. (2014). Regulation of acrAB expression by cellular metabolites in Escherichia coli. *J Antimicrob Chemother*, 69(2):390–399.

Smith, J. M. and Szathmary, E. (1997). *The major transitions in evolution*. Oxford University Press, USA.

Takeuchi, N. and Hogeweg, P. (2012). Evolutionary Dynamics of RNA-like Replicator Systems: A Bioinformatic Approach to the Origin of Life. *Phys Life Rev*, 9(3):219–263.

You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C. W., Wang, Y.-P., Lenz, P., Yan, D., and Hwa, T. (2013). Coordination of bacterial proteome with metabolism by cAMP signalling. *Nature*, 500(7462):301–306.