Mutation Rate as a Selection Mechanism in Continuous Evolution of Bacteriophage

Peter Reintjes

Museum of Life and Science
Innatrix, Inc.

Abstract

Engineering novel proteins via continuous evolution of bacteriophage currently requires the host bacteria to be transformed with two additional functions: A mutagenesis vector to provide an elevated rate of viral mutation, and a selection mechanism to increase the number of infectious progeny for the individual encoding the improved product [**?**]. I propose using a mutagenesis suppressor as the selection mechanism whereby the desired activity results in a higher number of clones, rewarding the genotype by increasing the number of identical progeny.

Introduction

Continuous evolution of bacteriophage has the potential to become a potent protein engineering tool [**?**] [**?**]. Continuous evolution rapidly produces a viral genome containing a gene which has undergone many generations of mutation and selection for a particular property. The generality of this approach to protein engineering is limited by our ability to insert an expressable initial gene into the phage and create a selection mechanism for the desired activity.

Specifically, the Phage Assisted Continuous Evolution (PACE) [**?**] system requires:

- A modified viral genome replacing a crucial phage gene with the gene to be evolved.
- A transformed host with inducible mutagenesis.
- A host plasmid containing a selection mechanism to provide the crucial phage gene in proportion to the desired activity of the evolving gene.

1 M-Selection

Extremely high, yet controllable, in vivo mutation rates are now possible [**?**]. This broad-spectrum mutagenesis with a (claimed) 320,000:1 dynamic range is generated and can be controlled within an individual host cell. I suggest that if this mutagenesis is initiated by infection, say with the phage shock promotor but then reduced by the desired activity, the two principal components of evolution will be satisfied: The desired activity will be selected for as the number of phage progeny of the individual exhibiting that activity will be amplified by more faithful reproduction. I propose calling this mechanism Mutagenesis-selection or M-selection. So far, all PACE-derived procedures

induce a uniform mutagenesis in the lagoon: M-selection doesn't even require the small-molecule (Ara) induction mechanism.

We require 2 clones per generation to infect hosts and produce phage to prevent washout of a genotype. Assuming phage production of 100/hour and a lagoon transit time of one hour, we calculate calculate a mutation rate of about 4 mutations per phage genome. The percentage of phage progeny which are exact copies of the parent is given by the Poisson distribution where $\mu = 0$ is the expected number of mutations and λ is the mutation rate per virion (mutation rate/base * 6kbp/ genome). A mutation rate of $\lambda = 4$ gives a 1.8% probability of zero mutations.

$$\mathrm{P}\left(\mu, \lambda\right) = \frac{e^{-\lambda}\lambda^{\mu}}{\mu!} \rightarrow P(0,4) = \frac{e^{-4}}{1} = .018 \tag{1}$$

So the maximum per-base mutation rate that a genome could tolerate and still stay in the lagoon is $6400/4 = 6.25X10^{-}4$. Liu et al. claim 322,000 times basal mutation rate. Basal rate is possibly ambiguous here if they are referring to the basal rate of E. coli producing E. coli $(4x10 - 10)$ vs. E coli producing M13 phage $(7.2x10 - 7)$, but in either case their claimed dynamic range covers the required phage mutation crossover rate required to avoid washout.

PACE is designed to keep the transit time of host cells through the lagoon on the order of one generation. However, at very high levels of mutation it seems unlikely that succeeding generations of E. coli mutants or phage will remain viable. Furthermore, the continuous flow environment ensures that there can be no other selection mechanism beyond the number of faithful copies of a particular genome. The obvious exception to this would be a colony or biofilm which remains somehow attached in the lagoon, but the population size for such entities would be limited by the surface areas within the lagoon, and not the volume.

Taking a phage production average of 100/hour for normal M13 infection requires us to have a cloning rate of at least 2%: two faithful copies per infection in order to avoid washout. The previous consideration of high mutation rates would seem to allow slower flow rates, which will lower this minimum fraction.

An important difference with PACE is that mutagenesis is not induced externally and therefore in uninfected cells, so there may be a minimum number of faithful copies produced before the mutagenesis products are expressed by the infected host. It may prove useful to retain the arabinose-induced mutagenesis as a way of achieving a baseline mutation rate in the lagoon, or an ante-chamber for the lagoon where induction can begin before exposure to phage.

As the evolving protein approaches a high level of desired activity, the proportionally lowered mutation rate will ensure that more viral progeny contain exact copies of the parent genotype, while a baseline mutation rate ensures that our mechanism to lower the mutation rate will not result in stagnation.

A possible drawback to this technique is that the mutation space is less agressively searched as the desired activity improves. But there is no particular reason that the modulation of mutation rate necessary to produce the population shift will be a particularly low (e.g. natural) mutation rate.

2 Possible Benefits of M-Selection

1. The virus is not crippled, that is, no critical genes are removed from the virion which should improve our ability to propagate phage prior to the evolution experiment. Care must be taken to ensure that the evolving gene is not discarded,

but M13 plus one additional gene or domain is exactly what has been perfected in phage display.

2. No phage genes are required in a host plasmid. This avoids the problem associated with the presence of gIII in the host.

3. High rates of mutation will be associated with individuals exhibiting little or none of the desired activity, and these lines of descent (both E. coli and phage) will be subject to catastrophic levels of mutation and low viability. This may reduce the importance of a carefully controlled flow rate.

4. Clonal Interference [**?**], the inabiltiy of a single genotype to gain sufficient advantage to become fixed in a population due to weak selection pressure is viewed as a problem in experimental evolution (one mutant has an important beneficial mutation, but also has a bad mutation hitch-hiking along which attenuates the advantage as compared to a sibling with neither mutation. In our case, maybe not a disadvantage as it keeps variety in the lagoon. Multiple genotypes in the lagoon at sample time would be informative. Does weak selection correspond to a flatter fitness landscape? I'm making an assumption here that M-selection corresponds to what Drake and others refer to as 'weak selection'.

5. We get 'genetic drift' for free, or at the cost of defective interference, anyway. To quote Patent #9023549 (Havard/PACE):

   A major problem with traditional directed evolution, whether in vitro or in vivo, is that libraries that do not contain any functional variants will be entirely lost, and the effort wasted–even if functionality lies only a couple of mutations away. The present invention overcomes this problem by allowing "switchable" genetic drift. This can be achieved by, for example, providing "free" propagation components to all library members, such as by inducible expression from an anhydrotetracycline-regulated titratable promoter, enabling all variants to infect host cells. By providing enough "free" propagation components, but less than the optimal level, any functional variants that arise will produce slightly more of the missing propagation components, produce more infectious progeny, and take over the population.

6. Adding a negative selection component may be simpler than in PACE: Just need something to re-up-modulate the mutation rate as a result of off-target binding. A suppressor supressor?

The nominal (high) mutation rate can be chosen to guarantee that there will be a minimum number of non-mutant copies produced for every infection. This percentage (of clones) will then increase proportunally as the mutation rate is reduced.

One downside to all this is defective interference from all the propagating mutants with none of the desired activity. So they are infecting host cells and crowding out the 'nice clones' as we wait for their lineages to undergo Error Catastrophe. It would be good to know how many uninfected host are making it through the lagoon. If we have an excess of uninfected host, this would reduce defective interference.

Figure 1 shows the combined mutagenic/selection mechanism replacing the Mutagenesis Plasmid(MP) and Accessory Plasmid (AP) in PACE.

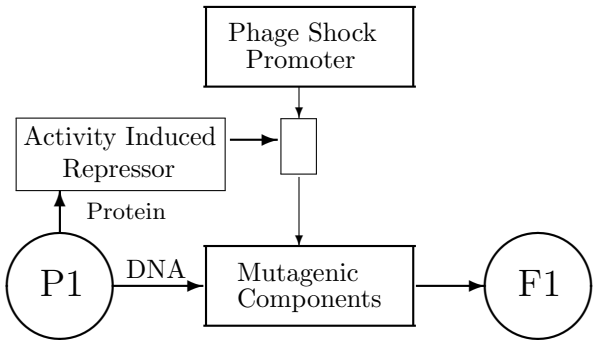


Fig 1. M-selection Plasmid

3 Comparison with Natural Selection

Selection pressure for an optimal mutation rate has its own natural history [**?**] although for settled phyla we see consistent [**?**], low mutation rates. Bacteriophage $6.4X10^{-}4$, and multicellular eukaryote mutation rates are higher than bacteria $4X10^{-}11$ and single cell eukaryotes $3X10^{-}10$.

A selection mechanism for lowered mutation rate would not seem to correspond to anything in the natural world. Natural selection is largely about responses to the external environment, which we control closely in laboratory evolution. Perhaps the closest parallel is found in species which can reproduce by asexual or sexual reproduction and change the rates between the alternatives as a way of accessing more diversity through sexual reproduction or producing more clones via asexual reproduction.

Comparison with Sexual/Asexual Reproduction

Sexual reproduction is not the same as an elevated mutation rate, but it shares the characteristic that even a perfectly fit individual will not be producing faithful copies of its genome. For organisms with access to both forms, the percentage of offspring produced by asexual reproduction would be comparable to our percentage of true copies.

Yeast

Although yeast has access to both sexual and asexual reproduction, the time and energy cost of sexual reproduction is high and it is generally induced by stress, indicating a need for more genetic diversity.

Daphnia

In Daphnia, longevity and healthy survival to the end of the season result in so-called "winter eggs" which reproduce asexually, providing a percentage of offspring which are clones of the parent. The cost differential between sexual and asexual reproduction is low and each generation represents a mix of both.

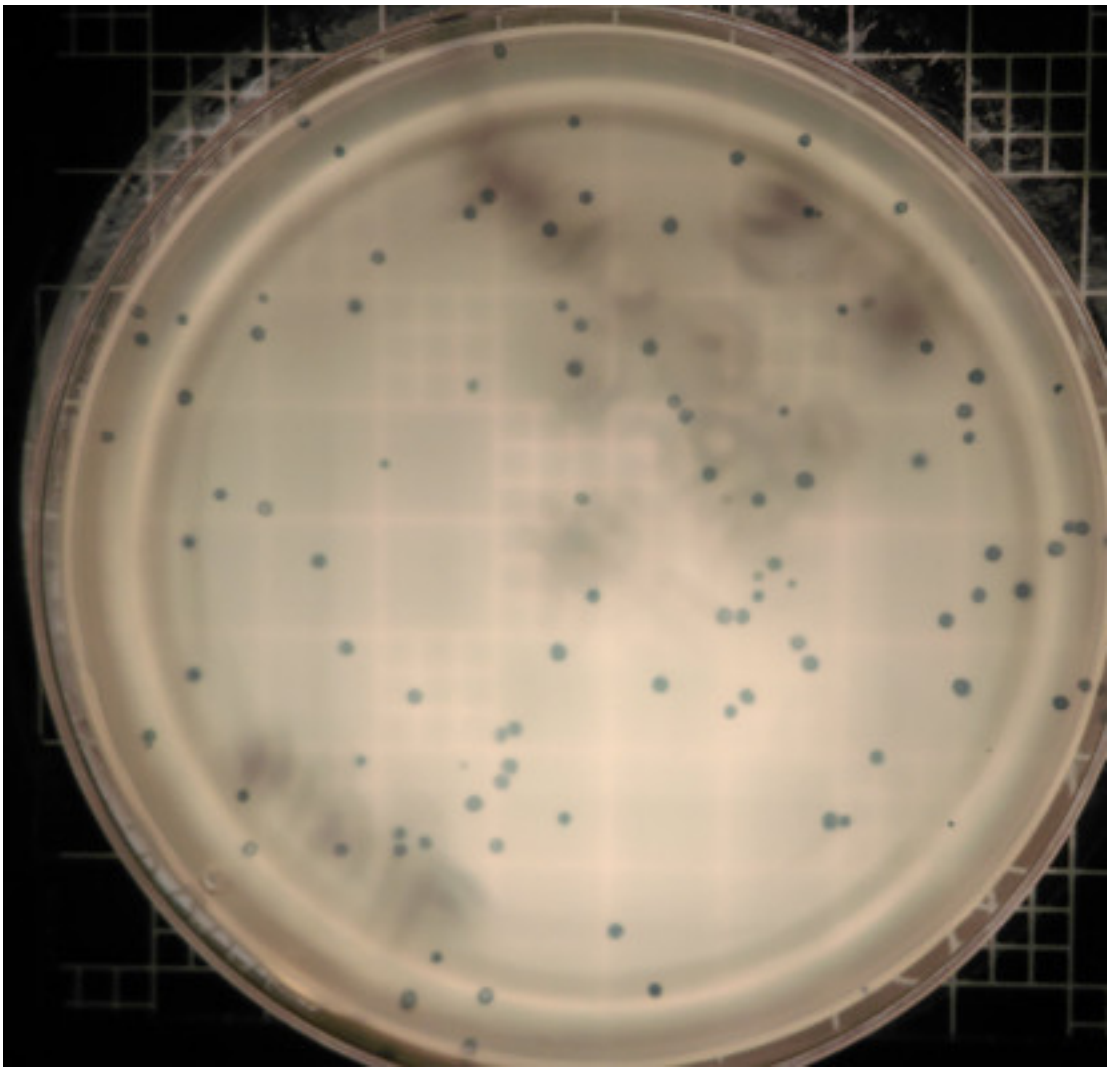
Materials and Methods

Population Dyamics of Bacteriophage (Under Construction)

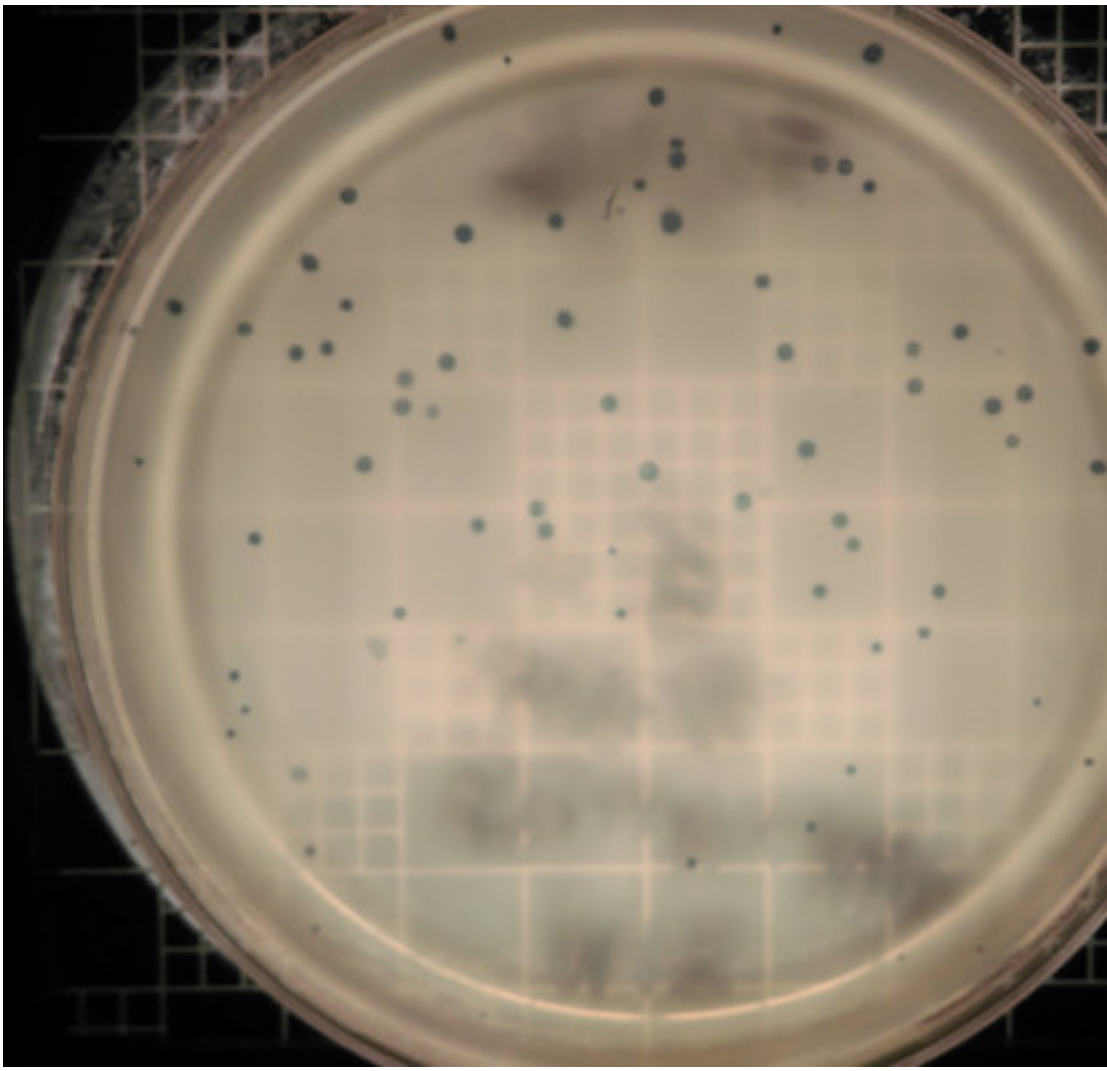
Clones versus Mutants

Although it may sounds like the title of a bad sci-fi movie, this precisely states the selection mechanism of mutation.
Percentage of activated genes due to Phage Shock promoter.
Number of polIII-(mutagenic polymerase) transcripts times error rate due to mutagenic polymerase.
Error rate of normal polII, polIII?
Number of phage particles produced np(0-6 min) = 0 (six minutes before any phage appears), then production between 75-200 per hour.
Some initial single-strand (virion) DNA production will occur in non-mutagenic environment, then mutants appear, until suppression of mutagenesis due to binding (e.g. the desired activity).

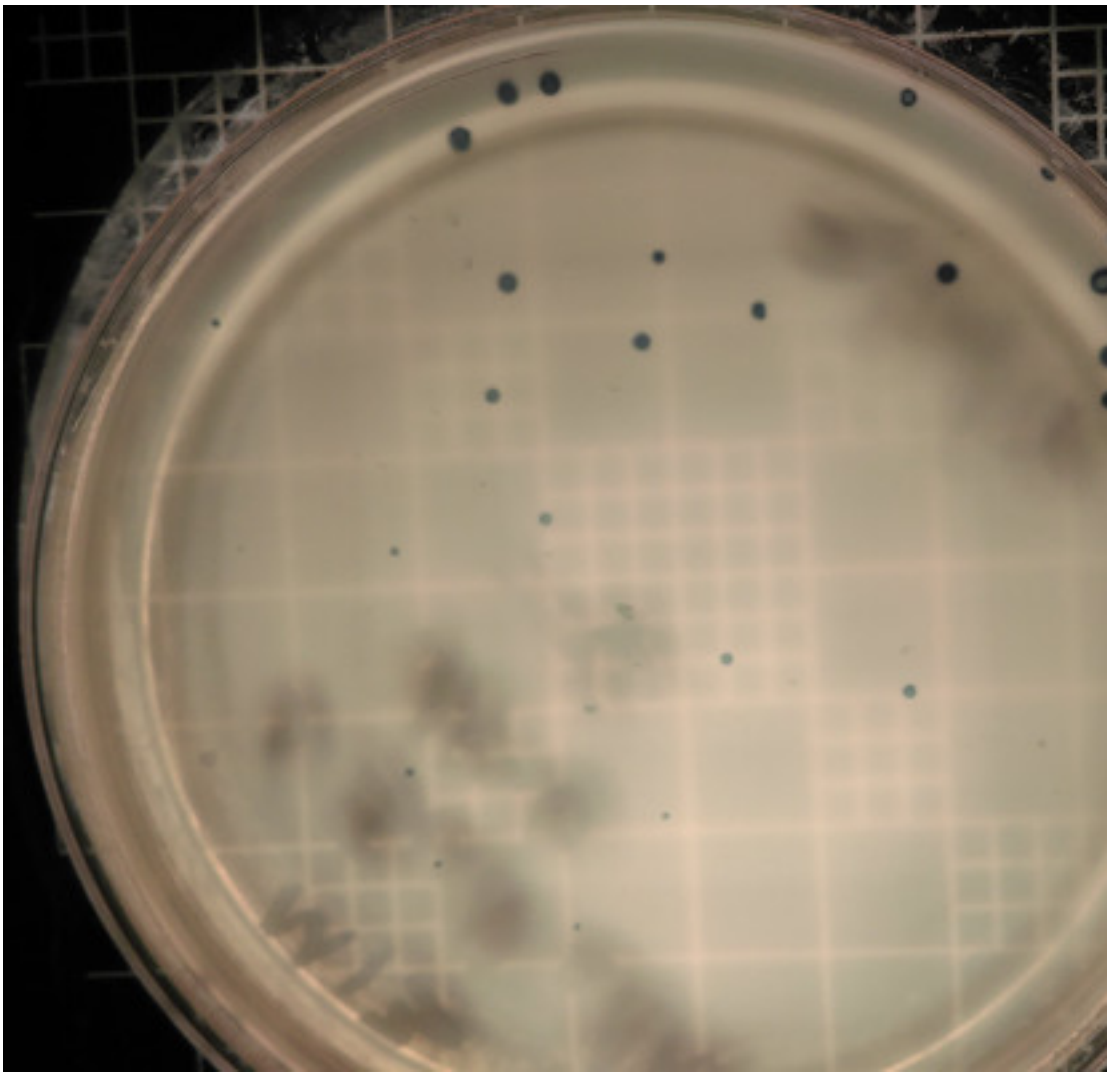
Results



M13 with 0.0 cAMP + MP6 $10^{-}7$



M13 with 5 milliMolar cAMP + MP6 $10^{-}5$



M13 with 50 milliMolar cAMP + MP6 $10^{-}5$

3.1 Methods for Lowering the Mutation Rate

RNA Anti-sense

The components for the broad-spectrum, high-level mutagenisis described in Badran and Liu [**?**] are variations (often single mutation variants) of the functional domains of high-fidelity DNA replication, and as such are not suitable targets for RNA anti-sense interference. Any such anti-sense interference would be likely to affect their normal counterparts which are needed for faithful reproduction.

RNA anti-sense of gIV could be utilized to block the phage shock promotor which is the primary stimulus that activates mutagenesis.

Other Methods

- Suppress pIV production or otherwise block phage shock promotor.
- Selectively block downstream components of mutagenesis.
- Enhance production of wild-type high-fidelity replication components.

References

1. Esvelt K. M., Carlson J. C. & Liu D. R. *A system for the continuous directed evolution of biomolecules.* Nature 472, 499–503 (2011).
2. Badran AH, Liu DR. *Development of potent in vivo mutagenesis plasmids with broad mutational spectra.* Nature Communications. 2015;6:8425. doi:10.1038/ncomms9425.
3. Barrick, Jeffrey E., Lenski, Richard E. *Genome Dynamics during Experimental Evolution* Nat Rev Genet, 2013 December, 14(12), pp. 827-839, doi:10.1038/nrg3564
4. Lynch M. *Evolution of the Mutation Rate.* Trends in Genetics, 2010 August; 26(8): 345–352. doi:10.1016/j.tig.2010.05.003.
5. Badran A. H., Guzov V.M., Huai Q., Kemp M. M.,Vishwanath P.,Kain W., Evdokimov A., Moshiri F., Turner K.H.P., Malvar T., Liu D.R. *Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance.* Nature 533,58-63 (2016) doi:10.1038/nature17938.
6. Drake J. W., *A constant rate of spontaneous mutation in DNA-based microbes*, Proc Natl Acad Sci U S A. 1991 Aug 15;88(16):7160-4.

Bacteriophage M13. The genome size is 6407 bp (4). The mutational target is 258 bp of an inserted Escherichia coli lacZa sequence (5). The spontaneous mutant frequency is $6.4x10-4$ per genome or $7.2x10-7$ per base pair *Drake, 1991*