Microbial Communities

Part 2: Evolution

Dave Ussery
DTU course 27105 - Comparative Genomics
Chapter 14 in the textbook

Thursday, 19 June, 2014
Outline

• Genome Diversity is large!

• Evolution can be very slow.

• Or evolution can be very fast!

  *e.g.*, a few hundred bp changes per million generations!

  *e.g.*, genome duplications, IS elements, HGT
Sequences as Biological Information

Organisms, the number of species present in the environment, and, despite their small size, the biomass they represent on a worldwide scale. Even inside an animal, microbes are abundant: only one out of every 10 cells in a human body is actually human, whilst the other nine cells are prokaryotic.

From an evolutionary perspective, Bacteria and Archaea have been around for more than 3 billion years; plants and animals are relatively recent 'newcomers' on the scene, arriving less than half a billion years ago. Since Bacteria and Archaea can divide rather quickly and have had much more time to evolve, their diversity by far exceeds that of eukaryotes (the members of Eucarya). Our human perception is that plants and animals are completely unlike each other, and so are, say, insects and mammals, as they are strikingly different even at first sight. The diversity of

Fig. 1.1

A phylogenetic tree displaying the genetic distances between members of the three superkingdoms of life: Bacteria, Archaea, and Eucarya. The represented bacterial genera will appear in examples throughout the book. The distance between bacterial genera is much larger than that of plants and animals, drawn on the same scale of genetic distance.
Univ ersal phylogenetic tree showing the relationships among Bacteria (e.g., most bacteria and blue-green algae), Archaea (e.g., methanogens and halophiles) and Eucarya (e.g., protists, plants, animals, and fungi).

rRNA tree

HUMANS

Corn
Figure 4.1: 23S rRNA tree with NJ method and 1000 bootstrap resamplings from ClustalW alignments. The tree is viewed and colored with MEGA5. Each phyla is collapsed when possible, except orders of Proteobacteria was collapsed instead of phyla. See Fig. 5.2 in Appendix for uncollapsed version.
Is the pan-genome also a pan-selectome? [v1; ref status: Indexed, http://f1000r.es/Vl9wKIl]

Francisco Rodriguez-Valera¹ and David W. Ussery²

¹Dependometo de Producción Vegetal y Microbiología, Universidad Miguel Hernandez, Apartado 18 San Juan de Alicante 03550, Spain
²Center for Biological Sequence Analysis, Department of Systems Biology, The Technical University of Denmark, Kgs. Lyngby, Denmark

Correspondence to David W Ussery: dave@cbs.dtu.dk

© Usage Licensed by Creative Commons CC-BY 3.0


Abstract The comparative genomics of prokaryotes has shown the presence of conserved regions containing highly similar genes (the ‘core genome’) and other regions that vary in gene content (The ‘flexible’ regions). A significant part of the latter is involved in surface structures that are phage recognition targets. Another sizeable part provides for differences in niche exploitation. Metagenomic data indicates that natural populations of prokaryotes are composed of assemblages of clonal lineages or “meta-clones” that share a core of genes but contain a high diversity by varying the flexible component. This meta-clonal diversity is maintained by a collection of phages that equalize the populations by preventing any individual clonal lineage from hoarding common resources. Thus, this polyclonal assemblage and the phages preying upon them constitute natural selection units.

http://f1000research.com/articles/is-the-pan-genome-also-a-pan-selectome/
High Frequency of Horizontal Gene Transfer in the Oceans

Lauren D. McDaniel, 1* Elizabeth Young, 1 Jennifer Delaney, 1 Fabian Ruhnau, 2
Kim B. Ritchie, 3 John H. Paul 1

“GTAs from R. nubinhibens ISM show a wide host range and interspecific gene transfer under eco-logically relevant conditions. Environmental gene transfer frequencies ranging from $6.7 \times 10^{-3}$ to $4.7 \times 10^{-1}$ (Table 1) are 1900 to 459 million times the frequency for transformation (2) and 650,000 to 31 million times the frequency of transduction previously measured in the marine environment (7). These results suggest a genomic flexibility in marine microbial populations that facilitates their adaptation to changing environmental conditions.”
Metagenomics approaches in systems microbiology

José M. Vieites¹, María-Eugenia Guazzaroni², Ana Beloqui¹, Peter N. Golyshin³,⁴,⁵ & Manuel Ferrer¹

¹CSIC, Institute of Catalysis, Madrid, Spain; ²CSIC, Estación Experimental del Zaidín, Granada, Spain; ³Environmental Microbiology Laboratory, HZI-Helmholtz Centre for Infection Research, Braunschweig, Germany; ⁴School of Biological Sciences, Bangor University, Gwynedd, UK; and ⁵Centre for Integrated Research in the Rural Environment, Aberystwyth University-Bangor University Partnership (CIRRE), Aberystwyth, UK

Fig. 1. Phylogenetic composition of a battery of environmental communities. Results are the average of a total of 700 Mbp. As can be seen, the composition differs significantly among geographically and niche-specific environments.
Marine viruses — major players in the global ecosystem

Curtis A. Suttle

Abstract | Viruses are by far the most abundant ‘lifeforms’ in the oceans and are the reservoir of most of the genetic diversity in the sea. The estimated $10^{30}$ viruses in the ocean, if stretched end to end, would span farther than the nearest 60 galaxies. Every second, approximately $10^{31}$ viral infections occur in the ocean. These infections are a major source of mortality, and cause disease in a range of organisms, from shrimp to whales. As a result, viruses influence the composition of marine communities and are a major force behind biogeochemical cycles. Each infection has the potential to introduce new genetic information into an organism or progeny virus, thereby driving the evolution of both host and viral assemblages. Probing this vast reservoir of genetic and biological diversity continues to yield exciting discoveries.

The oceans cover more than 70% of the Earth’s surface. They control the climate, provide a significant amount of the protein that is consumed globally and produce approximately half of the Earth’s oxygen. Microorganisms are a major force behind the nutrient and energy cycles in the world’s oceans and constitute more than 90% of the living biomass in the sea. It is estimated that viruses kill approximately 20% of this biomass per day. As well as being agents of mortality, viruses are one of the largest reservoirs of unexplored genetic diversity on the Earth.

The virosphere is probably inclusive of every environment on the Earth, from the atmosphere to the deep biosphere. However, nowhere is the importance of viruses more evident than in the world’s oceans. The observation that millions of virus-like particles are present in every millilitre of ocean water$,^1$, coupled with evidence that viruses are substantial agents of mortality in heterotrophic and autotrophic plankton$,^2$, has focused attention on the enormous underestimation of the effects of viral infection in the sea. It has become apparent that viruses are major players in the mortality of marine microorganisms and, consequently, affect nutrient and energy cycles as well as the structure of microbial communities.

Although the story of marine viruses is still emerging, it is becoming increasingly clear that we need to incorporate viruses and virus-mediated processes into our understanding of ocean biology and biogeochemistry. Progress in our understanding of marine viruses and their effects has been rapid and has been summarized in several comprehensive reviews$.^3–4$. However, many challenges remain. This Review examines our current knowledge of marine viruses, and highlights areas in which marine virology is advancing quickly or seems to be poised for paradigm-shifting discoveries.

The abundance of marine viruses

![Figure 1 | Relative biomass and abundances of prokaryotes, protists and viruses.](image)

Viruses are by far the most abundant biological entities in the oceans, comprising $10^{30}$ individuals. This accurate high-throughput method also allows large numbers of samples to be analysed quickly, which should begin to supply us with a synoptic picture of the biosphere. However, nowhere is the importance of viruses more evident than in the world’s oceans. The observation that millions of virus-like particles are present in every millilitre of ocean water, coupled with evidence that viruses are substantial agents of mortality in heterotrophic and autotrophic plankton, has focused attention on the enormous underestimation of the effects of viral infection in the sea. It has become apparent that viruses are major players in the mortality of marine microorganisms and, consequently, affect nutrient and energy cycles as well as the structure of microbial communities.

Although the story of marine viruses is still emerging, it is becoming increasingly clear that we need to incorporate viruses and virus-mediated processes into our understanding of ocean biology and biogeochemistry. Progress in our understanding of marine viruses and their effects has been rapid and has been summarized in several comprehensive reviews. However, many challenges remain. This Review examines our current knowledge of marine viruses, and highlights areas in which marine virology is advancing quickly or seems to be poised for paradigm-shifting discoveries.

**The oceans cover more than 70% of the Earth’s surface. They control the climate, provide a significant amount of the protein that is consumed globally and produce approximately half of the Earth’s oxygen. Microorganisms are a major force behind the nutrient and energy cycles in the world’s oceans and constitute more than 90% of the living biomass in the sea. It is estimated that viruses kill approximately 20% of this biomass per day. As well as being agents of mortality, viruses are one of the largest reservoirs of unexplored genetic diversity on the Earth.**

**The virosphere is probably inclusive of every environment on the Earth, from the atmosphere to the deep biosphere. However, nowhere is the importance of viruses more evident than in the world’s oceans. The observation that millions of virus-like particles are present in every millilitre of ocean water, coupled with evidence that viruses are substantial agents of mortality in heterotrophic and autotrophic plankton, has focused attention on the enormous underestimation of the effects of viral infection in the sea. It has become apparent that viruses are major players in the mortality of marine microorganisms and, consequently, affect nutrient and energy cycles as well as the structure of microbial communities.**

Although the story of marine viruses is still emerging, it is becoming increasingly clear that we need to incorporate viruses and virus-mediated processes into our understanding of ocean biology and biogeochemistry. Progress in our understanding of marine viruses and their effects has been rapid and has been summarized in several comprehensive reviews. However, many challenges remain. This Review examines our current knowledge of marine viruses, and highlights areas in which marine virology is advancing quickly or seems to be poised for paradigm-shifting discoveries.

**The oceans cover more than 70% of the Earth’s surface. They control the climate, provide a significant amount of the protein that is consumed globally and produce approximately half of the Earth’s oxygen. Microorganisms are a major force behind the nutrient and energy cycles in the world’s oceans and constitute more than 90% of the living biomass in the sea. It is estimated that viruses kill approximately 20% of this biomass per day. As well as being agents of mortality, viruses are one of the largest reservoirs of unexplored genetic diversity on the Earth.**

**The virosphere is probably inclusive of every environment on the Earth, from the atmosphere to the deep biosphere. However, nowhere is the importance of viruses more evident than in the world’s oceans. The observation that millions of virus-like particles are present in every millilitre of ocean water, coupled with evidence that viruses are substantial agents of mortality in heterotrophic and autotrophic plankton, has focused attention on the enormous underestimation of the effects of viral infection in the sea. It has become apparent that viruses are major players in the mortality of marine microorganisms and, consequently, affect nutrient and energy cycles as well as the structure of microbial communities.**

Although the story of marine viruses is still emerging, it is becoming increasingly clear that we need to incorporate viruses and virus-mediated processes into our understanding of ocean biology and biogeochemistry. Progress in our understanding of marine viruses and their effects has been rapid and has been summarized in several comprehensive reviews. However, many challenges remain. This Review examines our current knowledge of marine viruses, and highlights areas in which marine virology is advancing quickly or seems to be poised for paradigm-shifting discoveries.
Viruses manipulate the marine environment

Forest Rohwer¹ & Rebecca Vega Thurber¹,²

Marine viruses affect Bacteria, Archaea and eukaryotic organisms and are major components of the marine food web. Most studies have focused on their role as predators and parasites, but many of the interactions between marine viruses and their hosts are much more complicated. A series of recent studies has shown that viruses have the ability to manipulate the life histories and evolution of their hosts in remarkable ways, challenging our understanding of this almost invisible world.

Marine virology has traditionally focused on two areas: viruses as pathogens of aquatic organisms, and phage-driven dynamics of the marine microbial food web. Both of these influence global biogeochemistry and host evolution, and the former also has important economic and conservation implications. For example, two common marine viral diseases, sea-turtle fibropapillomatosis and shrimp white spot syndrome, endanger protected marine species and the financial stabil-

primary producers showed that genes involved in photosynthesis are commonly carried in phage genomes¹³. These genes include the highlight-inducible (hli) gene, as well as psbA and psbD, which encode the photosystem II (PSII) core reaction-centre proteins D1 and D2, respectively¹⁴ (Table 1). The D1 protein is of particular interest because it is the most labile protein in PSII and the most likely to be rate limiting. During the lytic cycle, most of the host’s transcription and translation is shut
"An inordinate fondness of viruses..."
Scientists following the evolution of a single strain of bacteria reported that it underwent several steps of mutation, surprising in its complexity, to acquire the ability to use a new food source.

The findings, reported Wednesday in the science magazine *Nature*, are the result of an experiment started 25 years ago by Richard Lenski of Michigan State University.

“When I started that project, I thought I would find one or two mutations and be done with it,” said Zachary Blount, a member of Lenski’s lab. “But instead, there may be dozens of mutations working together.”

“Creationists sometimes argue that even two mutations for one trait is too much complexity, yet here we see that evolution manages that with ease,” he said.

To study evolution in real time, Lenski followed the descendents of a single E. coli bacterium, a bug that normally populates our intestines. Bacteria have short life spans and in this experiment went through more than six generations a day.

Every day for 25 years — over 50,000 bacterial lifetimes — members of Lenski’s lab transferred the E. coli into a new flask with sugar solution. Every 500 generations, a part of the population was stowed in a freezer, creating a fossil record that can be brought back to life.

One day in 2003, the scientists observed something peculiar: A flask was much more densely populated than usual. At first the scientists suspected contamination. But then they found that after 30,000 generations, the bacteria had discovered how to use a different chemical as a food source. Citrate, the chemical in question, is given to the bacteria to help them absorb minerals and cannot normally be digested in the presence of oxygen.

What the researchers found was that a gene, normally responsible for letting citrate into the cell only in the absence of oxygen, had moved to a new location in the bacterium’s DNA. There it was controlled by a different switch, enabling citrate to enter even when oxygen was present. But this was only the second of three steps, the scientists found. An additional set of mutations were necessary in the beginning; the final step was multiplying the gene inside the DNA to make the bacteria much more efficient in their absorption of citrate.

The scientists conclude that these three stages may be universal evolutionary principles. “Even evolutionary changes that seem to be very sudden and dramatic may typically require a series of multiple steps drawn out over much longer periods of time than meets the eye,” Lenski said.
Genomic analysis of a key innovation in an experimental *Escherichia coli* population

Zachary D. Blount\textsuperscript{1,2}, Jeffrey E. Barrick\textsuperscript{2,3,4}, Carla J. Davidson\textsuperscript{5} & Richard E. Lenski\textsuperscript{1,2}

Evolutionary novelties have been important in the history of life, but their origins are usually difficult to examine in detail. We previously described the evolution of a novel trait, aerobic citrate utilization (Cit\textsuperscript{+}), in an experimental population of *Escherichia coli*. Here we analyse genome sequences to investigate the history and genetic basis of this trait. At least three distinct clades coexisted for more than 10,000 generations before its emergence. The Cit\textsuperscript{+} trait originated in one clade by a tandem duplication that captured an aerobically expressed promoter for the expression of a previously silent citrate transporter. The clades varied in their propensity to evolve this novel trait, although genotypes able to do so existed in all three clades, implying that multiple potentiating mutations arose during the population’s history. Our findings illustrate the importance of promoter capture and altered gene regulation in mediating the exaptation events that often underlie evolutionary innovations.
Figure 1 | Phylogeny of Ara–3 population. Symbols at branch tips mark 29 sequenced clones; labels are shown for clones mentioned in main text and figures. Shaded areas and coloured symbols identify major clades. Fractions above the tree show the number of clones belonging to the clade that yielded Cit\(^+\) mutants during replay experiments (numerator) and the corresponding total used in those experiments (denominator). Inset shows number of mutations relative to the ancestor. The solid line is the least-squares linear regression of mutations in non-mutator genomes; the dashed line is the corresponding regression for mutator genomes.
Life Is Designed to Fight Darwinism

Evolution News & Views September 26, 2012 10:03 AM | Permalink

Life knows all about Darwinism. That's why it is intelligently designed to resist it.

Neo-Darwinian theory teaches that mutations are random, then a process of "selection" chooses which ones to preserve (Darwin himself was bothered by the implicit personification in "natural selection"). Some mutations are random, for sure. No cell can anticipate where a cosmic ray will hit. How, then, can cells regulate mutations, turning on a "low mutation rate phenotype" under stress? In Current Biology, McDonald and team experimented with E. coli response to mutations. Their paper featured these 4 highlights:

The evolution of low mutation rates in mutator-founded populations • Nonmutators do not only evolve due to reduced genetic load compared to mutators; they can invade • Diploidy is found to be closely associated with reduction of the impact of the mutator phenotype • Genomes may influence mutation rates by initiating more high-fidelity replication early in S phase.

The authors believe that mutations are the source of "beneficial adaptive variation," but cannot deny they also produce "deleterious genetic load." When a cell invades a novel environment, it is able to switch on a "mutator phenotype" with a 10- to 100-fold increase in mutation rate. The fact that this "mutator allele" switches on is an indication that there's a functional purpose behind it. It's risky, because mutational load is likely to drive many of the cells extinct.

http://www.evolutionnews.org/2012/09/3_ways_cells_fi064501.html
Genome-wide Mutational Diversity in an Evolving Population of *Escherichia coli*

**J.E. Barrick and R.E. Lenski**

*Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824*

*Correspondence: lenski@msu.edu*

The level of genetic variation in a population is the result of a dynamic tension between evolutionary forces. Mutations create variation, certain frequency-dependent interactions may preserve diversity, and natural selection purges variation. New sequencing technologies offer unprecedented opportunities to discover and characterize the diversity present in evolving microbial populations on a whole-genome scale. By sequencing mixed-population samples, we have identified single-nucleotide polymorphisms (SNPs) present at various points in the history of an *Escherichia coli* population that has evolved for almost 20 years from a founding clone. With 50-fold genome coverage, we were able to catch beneficial mutations as they swept to fixation, discover contending beneficial alleles that were eliminated by clonal interference, and detect other minor variants possibly adapted to a new ecological niche. Additionally, there was a dramatic increase in genetic diversity late in the experiment after a mutator phenotype evolved. Still finer-resolution details of the structure of genetic variation and how it changes over time in microbial evolution experiments will enable new applications and quantitative tests of population genetic theory.
Figure 1. Expected dynamics in an evolving bacterial population. Lineages with new beneficial mutations are depicted as shaded wedges that originate in a previous genetic background and rise in frequency as they outcompete their ancestor and other lineages (Muller 1932). The same shading indicates that lineages have equivalent fitnesses, and the light gray curve highlights the path to the final dominant genotype containing five mutations. This figure was produced using a simulation with population size and mutation parameters meant to model the first 600 generations of the *E. coli* long-term evolution experiment (Woods 2005). Note how the level of genetic diversity changes over time.
**Figure 5.** Mutational diversity in an evolving *E. coli* population. (A) Origin and eventual fate of point mutations in the 2K to 40K mixed-population samples. New mutations that first appear as SNPs or fixed alleles are shown as asterisks along the bottom or top, respectively, with arrows leading to the corresponding pools of SNPs and fixed mutations. Transient SNPs that were lost from the population are shown by descending lines ending in closed circles. Note that we only detect SNPs when they are between ~4% and 96% frequency in the population and that we only recover ~50% of the SNPs at 5% frequency. Only the 49 SNP predictions in Table 2 were included for the 2K to 20K samples. (B) Stylized summary of the mixed-population SNP analysis. Shaded wedges represent subpopulations containing new mutations relative to the previous genetic background. Mutations are grouped to highlight their eventual fates, but we do not always have linkage information to resolve which SNPs occurred together. Labeled features are explained in the text.
Genome evolution and adaptation in a long-term experiment with *Escherichia coli*

Jeffrey E. Barrick, Dong Su Yu, Sung Ho Yoon, Haeyoung Jeong, Tae Kwang Oh, Dominique Schneider, Richard E. Lenski & Jihyun F. Kim

The relationship between rates of genomic evolution and organismal adaptation remains uncertain, despite considerable interest. The feasibility of obtaining genome sequences from experimentally evolving populations offers the opportunity to investigate this relationship with new precision. Here we sequence genomes sampled through 40,000 generations from a laboratory population of *Escherichia coli*. Although adaptation decelerated sharply, genomic evolution was nearly constant for 20,000 generations. Such clock-like regularity is usually viewed as the signature of neutral evolution, but several lines of evidence indicate that almost all of these mutations were beneficial. This same population later evolved an elevated mutation rate and accumulated hundreds of additional mutations dominated by a neutral signature. Thus, the coupling between genomic and adaptive evolution is complex and can be counterintuitive even in a constant environment. In particular, beneficial substitutions were surprisingly uniform over time, whereas neutral substitutions were highly variable.

Figure 2 | Rates of genomic evolution and fitness improvement. Blue circles show the total number of genomic changes relative to the ancestor in each sampled clone. The blue line represents a model where mutations accumulate uniformly over time. The light blue curves define the 95% confidence interval for this linear model. Green squares show the improvement of this population’s mean fitness relative to the ancestor over time, and the green curve is a hyperbolic plus linear fit of this trajectory. Each fitness estimate is the mean of three assays; most of the spread of points around the fitness trajectory reflects statistical uncertainty inherent to the assays. The inset shows the number of mutations in the 40,000-generation clone; the dashed curve approximates the change in the time course of genomic evolution after a mutator phenotype appeared at about generation 26,500.
The next four rings, from outer to inner, show mutations present in genomes sampled at 15,000, 10,000, 5,000, and 2,000 generations. The innermost circle shows the genome position and scale in megabase pairs (Mb). Mutations that are off the line of descent to a genome sampled at 40,000 generations are capped with a circle. Only one mutation (kup/insJ-5), a 1-bp insertion near an IS150 element, shows an aberrant homoplasmic distribution, being present in clones 10K and 20K but not 15K. Precise molecular details for all mutations are shown in Supplementary Tables 1 and 2.
thirty-four mutations in clone 15K occur in clones 20K and 40K. Of the six mutations in clone 2K are present in all later clones, and all present in clones from all subsequent generations. For example, four mutations in individual clones that did not become fixed in the population populations after 20,000 generations. There is substantial parallelism, in identical environments. Fourteen genes in which mutations were studied here. By contrast, selection should target the same genes in the predator. Emergence of a hypermutable phenotype.(...)

<table>
<thead>
<tr>
<th>Gene or region</th>
<th>Function</th>
<th>Parallel mutations (%)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>nadR</td>
<td>Transcriptional regulator</td>
<td>100</td>
<td>Ref. 42</td>
</tr>
<tr>
<td>pykF</td>
<td>Pyruvate kinase</td>
<td>100</td>
<td>Ref. 42</td>
</tr>
<tr>
<td>rbs operon</td>
<td>Ribose catabolism</td>
<td>100</td>
<td>Ref. 43</td>
</tr>
<tr>
<td>maiT</td>
<td>Transcriptional regulator</td>
<td>64</td>
<td>Ref. 44</td>
</tr>
<tr>
<td>spoT</td>
<td>Stringent response regulator</td>
<td>64</td>
<td>Ref. 31</td>
</tr>
<tr>
<td>mrdA</td>
<td>Cell-wall biosynthesis</td>
<td>45</td>
<td>Ref. 42</td>
</tr>
<tr>
<td>infB</td>
<td>Translation initiation factor</td>
<td>45*</td>
<td>This study</td>
</tr>
<tr>
<td>fis</td>
<td>Nucleoid-associated protein</td>
<td>27</td>
<td>E. Crozat, D.S., unpublished</td>
</tr>
<tr>
<td>topA</td>
<td>DNA topoisomerase I</td>
<td>27</td>
<td>E. Crozat, D.S., unpublished</td>
</tr>
<tr>
<td>pcnB</td>
<td>Poly(A) polymerase</td>
<td>27</td>
<td>This study</td>
</tr>
<tr>
<td>ompF</td>
<td>Outer-membrane porin</td>
<td>18*</td>
<td>This study</td>
</tr>
<tr>
<td>rpsD</td>
<td>30S ribosomal protein</td>
<td>18*</td>
<td>This study</td>
</tr>
<tr>
<td>rpsM</td>
<td>30S ribosomal protein</td>
<td>0</td>
<td>This study</td>
</tr>
<tr>
<td>glmU promoter</td>
<td>Cell-wall biosynthesis</td>
<td>0</td>
<td>M. Stanek, R.E.L., unpublished</td>
</tr>
</tbody>
</table>

*In addition to populations with substitutions, one or more others were polymorphic.

Table 2 | Tests of fitness effect in competition between isogenic constructs

<table>
<thead>
<tr>
<th>Gene or region</th>
<th>Fitness effect (%)</th>
<th>Significance</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>topA</td>
<td>13.3</td>
<td>***</td>
<td>Ref. 32</td>
</tr>
<tr>
<td>pykF*</td>
<td>11.1</td>
<td>***</td>
<td>D.S., R.E.L., unpublished</td>
</tr>
<tr>
<td>spoT</td>
<td>9.4</td>
<td>***</td>
<td>Ref. 31</td>
</tr>
<tr>
<td>nadR†</td>
<td>8.1</td>
<td>***</td>
<td>D.S., R.E.L., unpublished</td>
</tr>
<tr>
<td>glmU promoter</td>
<td>4.9</td>
<td>***</td>
<td>M. Stanek, T. Cooper, R.E.L., unpublished</td>
</tr>
<tr>
<td>fis</td>
<td>2.9</td>
<td>***</td>
<td>Ref. 32</td>
</tr>
<tr>
<td>rbs operon†</td>
<td>2.1</td>
<td>***</td>
<td>Ref. 43</td>
</tr>
<tr>
<td>malT</td>
<td>0.4</td>
<td>**</td>
<td>Ref. 44</td>
</tr>
<tr>
<td>ompF‡</td>
<td>−9.7</td>
<td>**</td>
<td>D.S., R.E.L., unpublished</td>
</tr>
</tbody>
</table>

* For this mutation, isogenic constructs were made by replacing the evolved allele with the ancestral allele in the evolved genetic background. For all other mutations, isogenic constructs were made in the ancestral background.
† In these two cases, artificial deletions of the genes were constructed in the ancestral background and the fitness effects of those deletions are reported.
‡ The deleterious effect of this mutation could indicate that it hitchhiked to high frequency. Alternatively, its fitness effect was tested only in the ancestral background, and it might be beneficial in association with one or more other evolved alleles.

** $P < 0.01$; *** $P < 0.001$. All significance levels are based on multiple independent competition assays.
Curved DNA

topological boundary

plectonemic supercoils

toroidal supercoiling

RNA polymerase

Intrinsic Curvature
dev avg
0.16 0.21

Stacking Energy
dev avg
-8.66 -7.71

Position Preference
dev avg
0.14 0.16

Annotations:
- CDS +
- CDS -
- rRNA
- tRNA

Global Direct Repeats
fix avg
5.00 7.50

Global Inverted Repeats
fix avg
5.00 7.50

GC Skew
dev avg
-0.05 0.05

Percent AT
fix avg
0.20 0.80

Resolution: 1856

E. coli K-12
isolate MG1655 4,639,221 bp

http://www.cbs.dtu.dk/
Center for Biological Sequence Analysis

GENOME ATLAS
Experimental evolution with *E. coli* in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations

Tim F Cooper1*, Richard E Lenski2

**Abstract**

**Background:** Environmental conditions affect the topology of the adaptive landscape and thus the trajectories followed by evolving populations. For example, a heterogeneous environment might lead to a more rugged adaptive landscape, making it more likely that replicate populations would evolve toward distinct adaptive peaks, relative to a uniform environment. To date, the influence of environmental variability on evolutionary dynamics has received relatively little experimental study.

**Results:** We report findings from an experiment designed to test the effects of environmental variability on the adaptation and divergence of replicate populations of *E. coli*. A total of 42 populations evolved for 2000 generations in 7 environmental regimes that differed in the number, identity, and presentation of the limiting resource. Regimes were organized in two sets, having the sugars glucose and maltose singly and in combination, or glucose and lactose singly and in combination. Combinations of sugars were presented either simultaneously or as temporally fluctuating resource regimes. This design allowed us to compare the effects of resource identity and presentation on the evolutionary trajectories followed by replicate populations. After 2000 generations, the fitness of all populations had increased relative to the common ancestor, but to different extents. Populations evolved in glucose improved the least, whereas populations evolving in maltose or lactose increased the most in their respective sets. Among-population divergence also differed across regimes, with variation higher in those groups that evolved in fluctuating environments than in those that faced constant resource regimens. This divergence under the fluctuating conditions increased between 1000 and 2000 generations, consistent with replicate populations evolving toward distinct adaptive peaks.

**Conclusions:** These results support the hypothesis that environmental heterogeneity can give rise to more rugged adaptive landscapes, which in turn promote evolutionary diversification. These results also demonstrate that this effect depends on the form of environmental heterogeneity, with greater divergence when the pairs of resources fluctuated temporally rather than being presented simultaneously.
Molecular Systems Biology 7; Article number 509; doi:10.1038/msb.2011.42
Citation: Molecular Systems Biology 7:509
© 2011 EMBO and Macmillan Publishers Limited All rights reserved 1744-4292/11
www.molecularsystemsbiology.com

REVIEW

Microbial laboratory evolution in the era of genome-scale science

Tom M Conrad, Nathan E Lewis and Bernhard Ø Palsson*

Laboratory evolution studies provide fundamental biological insight through direct observation of the evolution process. They not only enable testing of evolutionary theory and principles, but also have applications to metabolic engineering and human health. Genome-scale tools are revolutionizing studies of laboratory evolution by providing complete determination of the genetic basis of adaptation and the changes in the organism’s gene expression state. Here, we review studies centered on four central themes of laboratory evolution studies: (1) the genetic basis of adaptation; (2) the importance of mutations to genes that encode regulatory hubs; (3) the view of adaptive evolution as an optimization process; and (4) the dynamics with which laboratory populations evolve.

Molecular Systems Biology 7: 509; published online 5 July 2011; doi:10.1038/msb.2011.42

Subject Categories: metabolic and regulatory networks
Keywords: epistasis; flux-balance analysis; metabolic engineering; mutation; regulatory hub
Intragenic mutations identified in *E. coli* ALE studies. (A) Single-nucleotide substitutions, insertions, and deletions found within the open reading frames by whole-genome sequencing in multiple *E. coli* ALE studies (Herrin et al., 2006; Barrick et al., 2009; Conrad et al., 2009; Charusanti et al., 2010; Kishimoto et al., 2010; Lee and Palsson, 2010) are shown on a circular representation of the *E. coli* chromosome. (B) The set of genes displayed on the *E. coli* chromosome was subjected to enrichment analysis for Gene Ontology Slim (GOslim) categories (Camon et al., 2004). Wedges that protrude outward represent statistically enriched GOslim categories (also marked by **” in the legend). (C) Genes that were mutated in multiple studies are shown. 20K—growth on glucose minimal medium for 20,000 generations (Barrick and Lenski, 2009), 45A—adaptation to high temperature (Kishimoto et al., 2010), ETM—adaptation of ethanol tolerance (Goodarzi et al., 2009), Glyc—growth on glycerol minimal medium (Herrin et al., 2006), Lact—growth on lactate minimal media (Conrad et al., 2009), and PGI—growth on glucose minimal media following the deletion of *pgi* (Charusanti et al., 2010).
Many growth-coupled designs have been predicted to improve product yields, and the production strains become more stable. Thus, adaptation for a faster growth rate can result in higher production rates. In these growth-coupled designs, the metabolic pathways of an organism are altered such that the predicted optimal growth sequence, production strains often lose their ability to excrete the product as the cellular fitness improves. Algorithms have been developed to predict gene deletions that couple fast growth with metabolite excretion (Burgard et al., 2004; Patil et al., 2005; Lun et al., 2005; Trinh and Srienc, 2009). The concept of evolution toward optimal phenotypes has a general pattern of increased expression of genes and proteins associated with unused pathways (Becker et al., 2010). In systems that are initially designed to alter the uptake of metabolites in order to improve growth, the elucidation of the topological changes through an analysis of high-throughput data in the model context may allow for mechanistic conclusions drawn to incorporate the functions of transcriptional regulation and other non-metabolic functions. Efforts are being made to go beyond metabolism (Covert et al., 2006; Feist et al., 2010) and to also be developed to model multiple cellular processes. In addition, alternative modeling frameworks are being used and the transcription/translation machinery (Thiele et al., 2009). However, in the study of regulatory proteins still will pose a challenge regardless of the modeling framework. (Gianchandani et al., 2006; Chandrasekaran and Price, 2010). A phenotypic phase plane is a representation of how two fluxes in a metabolic network relate to each other and affect in silico-predicted optimal growth. Distinct planes are represented by several colors. Here, the line of optimality (LO, yellow) defines the ratio of glycerol uptake rate to oxygen uptake rate that leads to optimal biomass production. On glycerol, wild-type E. coli initially has a phenotype that maps to a suboptimal region of the portrait. After a growing for several hundred generations on glycerol, the E. coli phenotype migrates to the line of optimality. (D) Optimality principles can be used to design strains of bacteria in which growth at the maximal rate requires the secretion of a product of interest. When subjected to ALE, the designed strains increase both their growth rate and product secretion rate. The two colored regions indicate accessible flux states before and after evolution.

**Figure 2**  Optimality principles in adaptive evolution. (A) A smooth fitness landscape consisting of a single peak. Circles represent points on the landscape and arrows indicate the pathway of genetic change through the landscape. On a smooth landscape, there is a tendency for evolutionary convergence toward the single optimum, regardless of the starting point on the landscape. (B) A rough fitness landscape consists of multiple peaks. With multiple optima, there tends to be evolutionary divergence, sometimes even when starting from the same location on the landscape. (C) A phenotypic phase plane is a representation of how two fluxes in a metabolic network relate to each other and affect in silico-predicted optimal growth. Distinct planes are represented by several colors. Here, the line of optimality (LO, yellow) defines the ratio of glycerol uptake rate to oxygen uptake rate that leads to optimal biomass production. On glycerol, wild-type E. coli initially has a phenotype that maps to a suboptimal region of the portrait. After a growing for several hundred generations on glycerol, the E. coli phenotype migrates to the line of optimality. (D) Optimality principles can be used to design strains of bacteria in which growth at the maximal rate requires the secretion of a product of interest. When subjected to ALE, the designed strains increase both their growth rate and product secretion rate. The two colored regions indicate accessible flux states before and after evolution.
What about evolution ‘in the wild’??

Figure 7. Binomial mixture model. The pie chart to the left visualizes the binomial mixture model fitted to the *E. coli* domain sequence family distribution. There are 12 different sectors, and the color indicates the selection probabilities. The size of the each sector show its relative contribution to the pangenome. The pangenome is dominated by domain sequences who are either a very high (darker blue sectors) or very low selection probability (pink sectors). The right hand pie chart shows the expected distribution in a single *E. coli* genome, where the highly conserved families (darker blue) dominates.

http://f1000research.com/articles/a-domain-sequence-approach-to-pangenomics-applications-to-escherichia-coli/
What is a pan-genome?

E. coli pan-genome
~50,000 gene families

Escherichia coli K-12 MG1655
4144 proteins

~3000
E. coli gene families

BMC Genomics 2009, 10:385
InCoB 2012 - Bangkok, Thailand

**Human**
- 23,621 genes
- 19,568 orthologs

**Gorilla**
- 19,829 genes
- 14,080 orthologs
- 99%

**Chicken**
- 18,529 genes
- 10,673 orthologs
- 76%

**Worm**
- 19,404 genes
- 1,647 orthologs
- 55%

**Yeast**
- 5,885 genes
- 1,998 orthologs
- 28%

**E. coli**
- K-12, isolate MG1655
  - 4150 genes
  - 3843 orthologs
  - 95%

**E. coli**
- K-12, isolate W3110
  - 4226 genes
  - 3508 orthologs
  - 76%

**E. coli**
- HS
  - 4378 genes
  - 3399 orthologs
  - 60%

**E. coli**
- CFT073
  - 5339 genes
  - 3399 orthologs
  - 29%

**Y. pestis**
- Antiqua
  - 4364 genes
  - 1998 orthologs
  - 29%
MINIREVIEWS

On the Origins of a *Vibrio* Species

Tammi Vesth · Trudy M. Wassenaar · Peter F. Hallin · Lars Snipen · Karin Lagesen · David W. Ussery

Received: 3 July 2009 / Accepted: 17 September 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Thirty-two genome sequences of various *Vibrio* 
*naceae* members are compared, with emphasis on what 
makes *V. cholerae* unique. As few as 1,000 gene families 
are conserved across all the *Vibrionaceae* genomes ana-
lysed; this fraction roughly doubles for gene families 
conserved within the species *V. cholerae*. Of these, 
approximately 200 gene families that cluster on various 
locations of the genome are not found in other sequenced 
*Vibrionaceae*; these are possibly unique to the *V. cholerae* 
species. By comparing gene family content of the analysed 
genomes, the relatedness to a particular species is identified 
for two unspeciated genomes. Conversely, two genomes 
presumably belonging to the same species have suspicious-
ly dissimilar gene family content. We are able to identify a 
number of genes that are conserved in, and unique to, *V. 
cholerae*. Some of these genes may be crucial to the niche 
adaptation of this species.

Introduction

The species concept for bacteria has long been under siege 
from several angles, and now with thousands of bacterial 
genomes being sequenced, the disputes have intensified [8]. 
One frequently used definition of a bacterial species is “a
V. cholerae 01
El Tor N16961
chromosome 1
2,961,149 bp
Evolution of Microbial Communities; or, On the Origins of Bacterial Species

Outline

Evolution can be thought of as the adaptation or optimization of species to their environment. Since, at the level of microorganisms, there can be considerable differences in microenvironments, it is not hard to imagine that many bacteria have a constant need to be adaptable and ready to change to new surroundings. In this final chapter, we will take a look at the processes that drive evolution, and at the evolutionary traces that are visible in the DNA sequences of genomes. Mobile DNA elements play an important role in evolution and an example is given for insertion sequences in *Shigella flexneri*. Genome islands can be considered genetic 'building blocks' that can be added to or removed from a genome core. Finally, we will take a closer look at *Vibrio cholerae*, to see how this species differs from other *Vibrio* species, and how a relatively small set of genes can be responsible for niche adaptation (and sometimes speciation). The amount of genomic diversity within closely related bacterial populations is far greater than anyone had imagined, and the raw material for evolution is abundant in the microbial world.

Introduction

As mentioned in the first chapter, cells obey the laws of chemistry and physics, and there is no need to invoke supernatural forces to explain the physical mechanical events happening inside bacterial cells. One of the undercurrent themes of this book has been to build up a firm 'post-genomic' foundation from which to view the bacterial communities. We've now come full circle, and in this last chapter, we will have a look at the evidence for evolution within individual genomes, and how we can extrapolate such observations to bacterial populations.

In order for evolution to happen, three components are necessary: (1) a number of organisms must have a diverse set of traits that have different advantages under different conditions, (2) these traits must have the ability to change, and finally (3) selection must take place by some particular condition so that (some of) these traits become dominant in the offspring population. We can add the time factor to this as an essential component, because evolution is rarely instantaneous. Before turning to biological examples, we will first take a closer look at evolution in general.
In this final chapter, we will take a look at the processes that drive evolution, and at the evolutionary traces that are visible in the DNA sequences of genomes. Mobile DNA elements play an important role in evolution, and an example is given for insertion sequences in *Shigella flexneri*. Genome islands can be considered genetic 'building blocks' that can be added to or removed from a genome core. Finally, we will take a closer look at *Vibrio cholerae*, to see how this species differs from other *Vibrio* species, and how a relatively small set of genes can be responsible for niche adaptation (and sometimes speciation). The amount of genomic diversity within closely related bacterial populations is far greater than anyone had imagined, and the raw material for evolution is abundant in the microbial world.
Genome Islands


**Fig. 14.4** Genome Atlases of four *B. pseudomallei* strains, where genome islands are identified by arrows. The outer two lanes are BLAST Atlas lanes, as explained in the text.
Chapter 14
Evolution of Microbial Communities; or, On the Origins of Bacterial Species

Outline
Evolution can be thought of as the adaptation or optimization of species to their environment. Since, at the level of microorganisms, there can be considerable differences in microenvironments, it is not hard to imagine that many bacteria have a constant need to be adaptable and ready to change to new surroundings. In this final chapter, we will take a look at the processes that drive evolution, and at the evolutionary traces that are visible in the DNA sequences of genomes. Mobile DNA elements play an important role in evolution and an example is given for insertion sequences in *Shigella flexneri*. Genome islands can be considered genetic ‘building blocks’ that can be added to or removed from a genome core. Finally, we will take a closer look at *Vibrio cholerae*, to see how this species differs from other *Vibrio* species, and how a relatively small set of genes can be responsible for niche adaptation (and sometimes speciation). The amount of genomic diversity within closely related bacterial populations is far greater than anyone had imagined, and the raw material for evolution is abundant in the microbial world.

Introduction
As mentioned in the first chapter, cells obey the laws of chemistry and physics, and there is no need to invoke supernatural forces to explain the physical mechanical events happening inside bacterial cells. One of the undercurrent themes of this book has been to build up a firm ‘post-genomic’ foundation from which to view the bacterial communities. We’ve now come full circle, and in this last chapter, we will have a look at the evidence for evolution within individual genomes, and how we can extrapolate such observations to bacterial populations.

In order for evolution to happen, three components are necessary: (1) a number of organisms must have a diverse set of traits that have different advantages under different conditions, (2) these traits must have the ability to change, and finally (3) selection must take place by some particular condition so that (some of) these traits become dominant in the offspring population. We can add the time factor to this as an essential component, because evolution is rarely instantaneous. Before turning to biological examples, we will first take a closer look at evolution in general.

---

Figure 14.1

Questions:

● How can evolution happen, if mutations are so rare?

● What is the major source of novel genes in bacteria??

● What methods could be used to predict possible genomic islands??