**Ex Uno Plures: Clonal Reinforcement Drives Evolution of a Simple Microbial Community**

Margie A. Kinnersley  
*University of Montana - Missoula*

Jared Wenger  
*Stanford University*

Evgueny Kroll  
*University of Montana - Missoula, evgueny.kroll@mso.umt.edu*

Julian Adams  
*University of Michigan - Ann Arbor*

Gavin Sherlock  
*Stanford University*

*See next page for additional authors*

Follow this and additional works at: [https://scholarworks.umt.edu/biosci_pubs](https://scholarworks.umt.edu/biosci_pubs)

Part of the Biology Commons, and the Population Biology Commons

Let us know how access to this document benefits you.

**Recommended Citation**


This Article is brought to you for free and open access by the Biological Sciences at ScholarWorks at University of Montana. It has been accepted for inclusion in Biological Sciences Faculty Publications by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
Authors
Margie A. Kinnersley, Jared Wenger, Evgueny Kroll, Julian Adams, Gavin Sherlock, and Frank Rosenzweig
Ex Uno Plures: Clonal Reinforcement Drives Evolution of a Simple Microbial Community

Margie Kinnersley1, Jared Wenger2, Evgueny Kroll1, Julian Adams3, Gavin Sherlock2, Frank Rosenzweig1*

1 Division of Biological Sciences, The University of Montana, Missoula, Montana, United States of America, 2 Department of Genetics, Stanford University School of Medicine, Stanford, California, United States of America, 3 Departments of Molecular, Cellular and Developmental Biology and Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, United States of America

Abstract

A major goal of genetics is to define the relationship between phenotype and genotype, while a major goal of ecology is to identify the rules that govern community assembly. Achieving these goals by analyzing natural systems can be difficult, as selective pressures create dynamic fitness landscapes that vary in both space and time. Laboratory experimental evolution offers the benefit of controlling variables that shape fitness landscapes, helping to achieve both goals. We previously showed that a clonal population of E. coli experimentally evolved under continuous glucose limitation gives rise to a genetically diverse community consisting of one clone, CV103, that best scavenges but incompletely utilizes the limiting resource, and others, CV101 and CV116, that consume its overflow metabolites. Because this community can be disassembled and reassembled, and involves cooperative interactions that are stable over time, its genetic diversity is sustained by clonal reinforcement rather than by clonal interference. To understand the genetic factors that produce this outcome, and to illuminate the community’s underlying physiology, we sequenced the genomes of ancestral and evolved clones. We identified ancestral mutations in intermediary metabolism that may have predisposed the evolution of metabolic interdependence. Phylogenetic reconstruction indicates that the lineages that gave rise to this community diverged early, as CV103 shares only one Single Nucleotide Polymorphism with the other evolved clones. Underlying CV103’s phenotype we identified a set of mutations that likely enhance glucose scavenging and maintain redox balance, but may do so at the expense of carbon excreted in overflow metabolites. Because these overflow metabolites serve as growth substrates that are differentially accessible to the other community members, and because the scavenging lineage shares only one SNP with these other clones, we conclude that this lineage likely served as an “engine” generating diversity by creating new metabolic niches, but not the occupants themselves.

Introduction

It is interesting to contemplate a tangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent upon each other in so complex a manner, have all been produced by laws acting around us… Darwin, 1859

Illuminating the laws that produce Darwin’s “tangled bank” remains one of the great challenges of biology, one that requires understanding how differences among forms are selected for, and how interdependence among forms is generated. In natural environments, meeting this challenge is complicated by the fact that selection pressures often vary widely over both space and time. Laboratory evolution experiments, in particular those using microbes, offer an attractive alternative by which to study, under controlled conditions, both the dynamic interplay between genotype and phenotype as well as the interactions among phenotypes in simple communities. Classical models of large asexually evolving populations led to the expectation that in simple environments complexity should be transient and limited in scope [1,2]. Experimental evidence now suggests otherwise. Multiple genotypes that arise from a single ancestral clone can coexist over evolutionary time; in other words, ex uno plures (out of one many). This phenomenon has been documented in spatially and temporally unstructured chemostats [3,4], in temporally-structured batch cultures [5–9] and in spatially-structured microcosms [10]. In each setting, the emergence and persistence of polymorphism in the absence of sexual recombination requires that cohabitants opportunistically exploit unoccupied niche space, and/or accept trade-offs between being a specialist and a generalist [11–13]. In serial dilution batch culture, multiple growth parameters can come under selection [14]: different clones may arise having reduced lag time, increased maximum specific growth rate, or enhanced capacity to survive at high cell densities.


Editor: Jianzhi Zhang, University of Michigan, United States of America

Received November 21, 2013; Accepted April 24, 2014; Published June 26, 2014

Copyright: © 2014 Kinnersley et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by NASA NNX07AJ28G to FR, NIH R01 HG003328 to GS, and the NIH-NIGMS Genetics and Developmental Biology Training Program (NIH GM007790), which supported JW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: Frank.Rosenzweig@mso.umt.edu
Experimental Evolution of Cross-Feeding in E. coli

Author Summary

The variability of natural systems makes it difficult to deduce how organisms’ genotypes manifest as phenotypes, and how communities of interacting organisms arise. Using laboratory experimental evolution we can control this variation. We previously showed that a population of E. coli that originated from a single clone and was cultured in the presence of a single limiting resource, evolves into a stable, three-membered community, wherein one clone excretes metabolites that the others utilize as carbon sources. To discern the genetic factors at work in producing this outcome and to illuminate the community’s physiology, we sequenced the genomes of the ancestral and evolved clones. We identified in the ancestor mutations that may have predisposed evolution of cross-feeding. We found that the lineages which gave rise to the community diverged early on, and that the numerically dominant lineage that best scavenges limiting glucose does so as a result of adaptive mutations that enhance glucose uptake but favor fermentative over respiratory pathways, resulting in overflow metabolites. Because this clone produces secondary resources that sustain other community members, and because it shares with them only one mutation, we conclude that it is an “engine” generating diversity by creating new niches, but not the occupants themselves.

in the presence of low nutrients [15]. Periodic changes in population density and nutrient levels may bring balancing selection to bear on these different phenotypes (e.g., [8]), especially if mutations are antagonistically pleiotropic [16,17]. In spatially structured environments selection may favor mutants better adapted to particular regions or better able to colonize microhabitats formed at the boundaries between such regions. By contrast, in continuous nutrient-limited environments, theory predicts that selection will favor clones better able to scavenge the limiting resource or more efficiently convert that resource to progeny [18,19]. Ultimately, the outcome of the ‘evolutionary play’ in any of these ‘ecological theaters’ (sensu [20]) will depend on founder genotype, mutation rate, the complexity of genetic pathways leading to different adaptive strategies, as well as pleiotropy [21] and epistatic interactions [22].

Increasing evidence points to the possibility of not one, but three potential outcomes when axenic microbes evolve in simple environments: clonal succession, where a population is successively swept by clones of higher fitness arising in the dominant lineage [23,24]; clonal interference, where fixation of a single fittest clone is deferred because independent beneficial mutations arise in multiple, independent clones that compete with one another and reduce each other’s fitness [25–27], and what we propose to call clonal reinforcement, where the emergence of one genotype favors the emergence and persistence of other genotypes via cooperative interactions. Because cooperation is now recognized to be at least as important as competition in driving biological innovation and in structuring communities [28,29], we investigated the genetic and environmental factors that foster interdependence in an evolving lab population.

Previously [30], we examined the transcriptomes of an experimentally evolved E. coli community in which three different strains that arose from a common ancestor stably coexisted in a simple, unstructured environment through metabolic cross-feeding [3,31]. This system is unusual in that it clearly involves symbiotic interactions among ecotypes that co-evolved in the absence of spatial or temporal variation. Remarkably, the community’s expression profile did not proportionately represent the sum of each strain’s expression profile grown singly under evolutionary conditions, indicating that the whole was not the sum of its parts. Instead, expression data suggested that consumption of one ecotype’s overflow metabolites by the others relieved feedback inhibition, adding a layer of complexity to the evolved clones’ interactions. Using a candidate gene approach we identified two mutations in the ancestral strain that may have predisposed clonal reinforcement [30]. To better understand how physiological performance and expression profile map onto genotype, and to identify the molecular basis of community interactions we sequenced the genomes of each of the three community members and their common ancestor. We found unexpectedly high levels of genetic variation in the three-membered community as a whole and the numerically dominant strain in particular, as well as a strong mutational bias due to specific lesions in DNA repair. The dominant community member is a hypermutator that excels at acquiring glucose, but does so at the expense of carbon excreted as overflow metabolites to which it has limited access; this trade-off effectively opens up new niches for other genotypes. We discovered a set of adaptive mutations in this clone that have not been previously reported to co-occur in other E. coli evolution experiments, specifically those that enhance glucose acquisition and may serve to maintain redox balance, but may also increase net flux to overflow metabolites under aerobic conditions. Because this community member releases growth substrates that are differentially accessible to other community members, and because it shares only one SNP mutation with these strains, we suggest that it acts as an engine generating biodiversity, creating new metabolic niches, but not necessarily the occupants themselves.

Results and Discussion

Description of the community

In previous publications [3,30,31] we described certain features of a stable polymorphism that arose from a single clone during the course of 765 generations of aerobic, glucose-limited culture at 30°C (D=0.20 hr⁻¹) (Table 1). The four clones isolated from this population (CV101, CV103, CV115 and CV116) were originally distinguished by their differences in antibiotic resistance and colony morphology, and later shown to exhibit strain-specific rates of glucose uptake, residual metabolite concentrations and expression profiles. In reconstruction experiments performed under evolutionary conditions CV101, CV103 and CV116 were shown to stably coexist at frequencies of approximately 0.10, 0.65 and 0.025, respectively. Because CV115 was not stably maintained, the interactions among CV101, CV103 and CV116 were studied more intensively. Based on differences in glucose uptake and residual substrate concentrations, and the fact that strains’ steady state frequencies could be predictably altered by exogenously increasing the concentrations of residual metabolites, we concluded that this community was sustained by positive density-dependent interactions in the form of cross-feeding, in which the numerically dominant strain CV103 best takes up the limiting resource glucose, but excretes acetate and glycerol (and/or a closely-related compound, glycerol 3-phosphate) (J. Adams unpublished results). These two overflow metabolites are then scavenged by CV101 and CV116, respectively. Transcriptional profiling of each clone in monoculture relative to the common ancestor, JAL22, revealed gene expression differences among the evolved isolates related to carbon metabolism and expression differences specific to CV103 consistent with activation of stress response pathways and loss of motility.

Table 1
<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant Characteristics&lt;sup&gt;2&lt;/sup&gt;</th>
<th>Mutation rate (/cell/gen)</th>
<th>Specific growth rate (hr&lt;sup&gt;-1&lt;/sup&gt;)&lt;sup&gt;3&lt;/sup&gt;</th>
<th>Relative growth yield&lt;sup&gt;3&lt;/sup&gt;</th>
<th>Rate of glucose uptake (μmol aMG/min/gm)&lt;sup&gt;3&lt;/sup&gt;</th>
<th>Equilibrium [glucose] (nmol/mL)&lt;sup&gt;3&lt;/sup&gt;</th>
<th>Equilibrium [acetate] (nmol/mL)&lt;sup&gt;3&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>K12 MG1655</td>
<td>CGSC 6300  &lt;br&gt; F- mph-1 λ-</td>
<td>3.60 × 10&lt;sup&gt;-7&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RH201&lt;sup&gt;1&lt;/sup&gt;</td>
<td>CGSC 5346  &lt;br&gt; F- thi 1 lacY1  &lt;br&gt; tonA21  &lt;br&gt; supE44 hsdR1 glpR200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JA104</td>
<td>Deriv. of RH 201 F- thi 1 lacY1  &lt;br&gt; araD139gdh  &lt;br&gt; supE44 hsdR1; lysogenic for λ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JA122</td>
<td>As JA104 but contains plasmid  &lt;br&gt; pBR322.A5</td>
<td>1.00 × 10&lt;sup&gt;-7&lt;/sup&gt;  &lt;br&gt; 0.44 ± 0.01  &lt;br&gt; 1.14 ± 0.02  &lt;br&gt; 1.19 ± 0.09  &lt;br&gt; 1.84 ± 0.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CV101</td>
<td>Derivative of JA122; isolated  &lt;br&gt; after 773 generations; Amp&lt;sup&gt;3&lt;/sup&gt;</td>
<td>1.33 × 10&lt;sup&gt;-7&lt;/sup&gt;  &lt;br&gt; 0.50 ± 0.02  &lt;br&gt; 1.11 ± 0.02  &lt;br&gt; 1.66 ± 0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CV103</td>
<td>As CV101 but independent isolate,  &lt;br&gt; forms small colonies on TA agar, Amp&lt;sup&gt;3&lt;/sup&gt;</td>
<td>9.20 × 10&lt;sup&gt;-7&lt;/sup&gt;  &lt;br&gt; 0.40 ± 0.01  &lt;br&gt; 0.81 ± 0.04  &lt;br&gt; 2.46 ± 0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CV115</td>
<td>Derivative of JA122, isolated after  &lt;br&gt; 773 generations, lacks plasmid</td>
<td>1.14 × 10&lt;sup&gt;-7&lt;/sup&gt;  &lt;br&gt; 0.55 ± 0.02  &lt;br&gt; 1.11 ± 0.02  &lt;br&gt; ND</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CV116</td>
<td>As CV115 but forms small colonies on  &lt;br&gt; TA agar</td>
<td>1.10 × 10&lt;sup&gt;-7&lt;/sup&gt;  &lt;br&gt; 0.60 ± 0.01  &lt;br&gt; 1.20 ± 0.03  &lt;br&gt; 1.61 ± 0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>1</sup>Ref. [149].  
<sup>2</sup>Data from [3], Table 1.  
<sup>3</sup>Data from [31], Table 2.  

doi:10.1371/journal.pgen.1004430.t001
Whole genome re-sequencing

As summarized in Table 2, 584 mutations (580 SNPs, two insertions and two deletions) were identified among the four evolved clones. This number is substantially higher than has been previously reported for laboratory evolution experiments of similar duration conducted under similar conditions [32]. A majority of these mutations, 502 (86.6%), is in coding regions and 374 (64%) are non-synonymous, while 428 (73%), are unique to CV103. Almost all (99.7% (578)) SNPs are GC→TA transversions, suggesting defects in DNA repair among all of the evolved isolates. Strong transversion bias has been noted in other evolution experiments, and is likely due to defective repair of oxidatively damaged G:C base-pairs [32,33].

The community consists of lineages that diverged early in the history of the experiment

Phylogenetic analyses of the whole genome sequences [30] show that the dominant clone CV103 is part of a highly divergent lineage, while CV115 and CV116 are very closely related to one another, and share a common ancestor with CV101 (Fig. 1). Indeed, as CV115 and CV116 differ by only two SNPs, one a synonymous substitution (gutQ, A221A), the other C-terminal (gcaO G546V), and because only CV101, CV103 and CV116 could be stably maintained at steady state, we restrict most of our discussion to these three clones. Remarkably, the evolved clones all share only one SNP. Thus, it is likely that the CV103 and CV101/ CV115/CV116 lineages diverged from one another early in the experiment and that stable co-existence of more than one lineage was an early feature of the population. In fact, previous work documented a variant similar to CV103 (small colony, ampicillin-resistant) had risen to appreciable frequency by generation 340 [3]. This result is not unexpected, as subsequent to the observations of Helling et al. 1987, a number of other reports appeared showing that multiple genotypes can coexist in nutrient-limited chemostats [34-38]. An unanticipated finding in our experiments was the large number of SNPs that accumulated along the CV103 branch.

Mutation rate varies among evolved strains and their common ancestor

Fluctuation analyses performed on the ancestral strain, JA122, the canonical E. coli strain K12 MG1655, and the four co-evolved clones revealed that the founder strain, JA122, is a mutator whose mutation rate is approximately thirty-fold higher than E. coli K12 MG1655 (Table 1). This observation can be explained by a nonsense mutation in JA122 affecting the adenine glycosylase mismatch repair enzyme MutY (L299*). Defects in MutY are known to cause a GC→TA transversion bias [33,39] and have been observed by others in glucose limited chemostat experiments [40,41]. Evolved strains CV101, CV115 and CV116 all had mutation rates similar to JA122, while CV103’s mutation rate was almost 10-fold higher again (Table 1). This difference can be attributed to a second mutation resulting in an amino acid substitution (A56D) in the base-excision repair pathway DNA glycosylase, MutM. Increased mutation rate has been shown to be selected for in microbial evolution experiments performed under a variety of conditions [33,41–45] where it can strongly influence evolutionary dynamics [46]. The fitness of mutators need not decrease over time; Maharjan et al. [47] recently showed that mutators arising under glucose limitation gain fitness relative to their ancestor as a result of adaptive mutations in, for example, rpoS and mglD. Mutator lineages persisted because most of the other mutations accumulating in their backgrounds were neutral in their effect on fitness, deleterious mutations being purged as the lineages evolved. A similar mechanism may be at work in our experiment, with an appreciable number of mutations in CV103 being neutral.

Certain transcriptional units are repeatedly mutated

Given the large number of SNPs, we sought to determine whether certain genes or transcription units carried multiple

---

Table 2. Types of Single Nucleotide Polymorphisms (SNPs) and their distribution among evolved strains.

<table>
<thead>
<tr>
<th>Total mutations</th>
<th>584</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertions/deletions</td>
<td>4</td>
</tr>
<tr>
<td>SNPs</td>
<td>580</td>
</tr>
<tr>
<td>SNPs in coding regions</td>
<td>502 (86.6%)</td>
</tr>
<tr>
<td>SNPs causing missense mutations</td>
<td>335 (57.8%)</td>
</tr>
<tr>
<td>SNPs causing nonsense mutations</td>
<td>39 (6.7%)</td>
</tr>
<tr>
<td>G to A transitions</td>
<td>1 (0.2%)</td>
</tr>
<tr>
<td>C to T transitions</td>
<td>1 (0.2%)</td>
</tr>
<tr>
<td>G to T transversions</td>
<td>293 (50.5%)</td>
</tr>
<tr>
<td>C to A transversions</td>
<td>285 (49.1%)</td>
</tr>
</tbody>
</table>

---

Figure 1. Phylogeny of strains based on whole-genome sequence. 580 SNPs across all four genomes were used to reconstruct evolutionary relationships between consortium members using maximum likelihood. The number of SNPs that distinguish each lineage are shown above each branch in green.

doi:10.1371/journal.pgen.1004430.g001
RNAs that interact with the global RNA chaperone, Hfq [49,50], unknown, but it has been repeatedly identified in screens for small encodes a small RNA, RNA0-361. The function of this sRNA is contribute to the superior glucose uptake kinetics of this clone.

Mutations that alter glucose transport: The outer membrane

**Fig. 2** In earlier work we showed that all evolved isolates were better able to scavenge limiting glucose than their common ancestor [31], and that strain CV103 had a significantly higher rate of glucose uptake than the other evolved isolates [3]. We later attributed these observations to increased expression of LamB glycoporin in all four strains relative to JA122 [51], and significantly higher transcription of lamB in CV103 relative to CV101, CV115 and CV116 [30]. In other studies of adaptation under glucose limitation elevated lamB expression has repeatedly been tied to mutations that affect the regulation of the Mal operon, specifically mutations that affect expression of the lamB regulators Mlc and/or MalT [35,51,52]. While we previously found that CV101, CV115 and CV116 share a mutation in malT that could up-regulate lamB expression, this mutation did not occur in CV103 [30], the strain that best scavenge limiting glucose.

**Whole genome sequencing uncovered two other mutations in CV103 that help explain higher transcript levels of lamB mRNA and, by extension, its superior ability to scavenge limiting glucose.**

The first mutation affects the gene for MalK, a negative regulator of MalT [53]. Null mutations in malK promote constitutive mal/lamB gene expression [54,55], presumably by disrupting the MalK-MalT interaction. The malK SNP in CV103 causes an amino acid change (D297E) in the C-terminal portion of the protein that contains an important part of the MalT interaction domain. Replacement of aa 297 is known to diminish the ability of MalK to inhibit MalT [56], and thus the D297E SNP likely leads to increased lamB transcription during glucose-limited growth [57–60].

The second CV103 mutation that may influence lamB expression occurs in the gene for the RNA chaperone Hfq, Hfq is a global regulator that facilitates binding of small regulatory RNAs (sRNAs) to mRNAs and by doing so affects translation and/or degradation of those transcripts [61,62]. Details are scant as to how most of these sRNAs regulate their targets and how Hfq enhances or attenuates their activity, but the profound effect these interactions can have on key cellular processes is being increasingly recognized [61,63]. Hfq interacts with the mRNA of the *E. coli* stationary phase sigma factor, RpoS, and is required for efficient translation of RpoS mRNA [64]. Reduced RpoS expression has been extensively studied as a key adaptation to continuous glucose-limitation [40,65,66], and Hfq has recently been identified as an alternative mutational target under glucose limitation in *sps* strains [67,68]. In these isolates, a missense mutation in Hfq enhances glucose uptake via PsG, increases levels of LamB, and apparently reduces the amount of functional RpoS, resulting in lower biomass yield. The *hfq* mutation in CV103, which results in a Q52H substitution, could have similar effects: CV103 has increased uptake of the glucose analog 2-MG (Table 1), which occurs exclusively via PsG, increased LamB gene expression, and decreased biomass yield relative to the other evolved clones [3,30]. Interestingly, the genomic context of the Hfq mutation in CV103 differs from those characterized by others in which defects in Hfq have evolved in an *sps* deficient background [67,68]. CV103 and our other evolved isolates share an ancestral *sps* (Am) mutation which likely allows some RpoS translation.

Mutations that alter glucose transport: The inner membrane

**Fig. 2** After entry into the periplasm, glucose is actively transported across the inner membrane using the glucose-specific sugar phosphotransferase system (PTS). Mutations that upregulate

![Image](https://via.placeholder.com/150)

Certain SNPs are associated with strain-specific differences in transcription and/or translation

Because we previously observed significant differences in gene expression among evolved isolates grown in monoculture [30], we sought to identify SNPs that could explain these differences. Because the great majority of these transcriptional differences were manifest in CV103, we focused on genes that could directly or indirectly regulate loci differentially expressed in this strain and on the 428 genes associated with a SNP in CV103. The intersection of these two gene lists yielded three global regulators that could explain the majority of transcriptional differences between CV103 and the other evolved isolates: tpoD, csrA and sdiA. TpoD, the housekeeping sigma factor, controls transcription of over 2,300 genes. The ancestor of the four evolved isolates, JA122, has an amber nonsense mutation (E26*, GAG→TAG) in tpoD that likely leads to reduced translation of TpoD, but not necessarily its complete absence, as JA122 also carries the supE44 amber suppressor. CV103 has an additional silent mutation in *tpoD* (T459T, ACC→ACA) that does not affect protein sequence but may affect translation, as the mutant codon ACA is 30% less common than the wild-type codon ACC [48]. The regulatory protein CsrA favors glucoseogenesis and glycogen synthesis over glycolysis. While the CV103 mutation in *csrA* is also silent (G27G, GGC→GGA), the resulting codon change is from one that is common (f=0.40) to one that is rare (f=0.09). Finally, *sdiA* encodes an N-acylhomoserine-L-lactone receptor that functions in quorum sensing. In CV103, a SNP is located 27 base-pairs upstream of the *sdiA* transcriptional start site in a region that encodes a small RNA, RNA0-361. The function of this sRNA is unknown, but it has been repeatedly identified in screens for small RNAs that interact with the global RNA chaperone, Hfq [49,50], in which CV103 has a missense mutation (see below).

Mutations that alter glucose transport: The inner membrane

Because we previously observed significant differences in gene expression among evolved isolates grown in monoculture [30], we sought to identify SNPs that could explain these differences. Because the great majority of these transcriptional differences were manifest in CV103, we focused on genes that could directly or indirectly regulate loci differentially expressed in this strain and on the 428 genes associated with a SNP in CV103. The intersection of these two gene lists yielded three global regulators that could explain the majority of transcriptional differences between CV103 and the other evolved isolates: tpoD, csrA and sdiA. TpoD, the housekeeping sigma factor, controls transcription of over 2,300 genes. The ancestor of the four evolved isolates, JA122, has an amber nonsense mutation (E26*, GAG→TAG) in tpoD that likely leads to reduced translation of TpoD, but not necessarily its complete absence, as JA122 also carries the supE44 amber suppressor. CV103 has an additional silent mutation in *tpoD* (T459T, ACC→ACA) that does not affect protein sequence but may affect translation, as the mutant codon ACA is 30% less common than the wild-type codon ACC [48]. The regulatory protein CsrA favors glucoseogenesis and glycogen synthesis over glycolysis. While the CV103 mutation in *csrA* is also silent (G27G, GGC→GGA), the resulting codon change is from one that is common (f=0.40) to one that is rare (f=0.09). Finally, *sdiA* encodes an N-acylhomoserine-L-lactone receptor that functions in quorum sensing. In CV103, a SNP is located 27 base-pairs upstream of the *sdiA* transcriptional start site in a region that encodes a small RNA, RNA0-361. The function of this sRNA is unknown, but it has been repeatedly identified in screens for small RNAs that interact with the global RNA chaperone, Hfq [49,50], in which CV103 has a missense mutation (see below).
an alternative high-affinity glucose transporter, the galactose transporter MglBAC, are frequently observed following prolonged growth under glucose limitation [35,51,52]; in this regard, our system is no exception. In fact, the only SNP shared by all of the evolved isolates is in the operator sequence for the MglD repressor (mglO, C→A, +3 bp relative to the end of mglD), consistent with increased expression of the mglBAC transcription unit [30,35].

Because the rate of glucose uptake in CV103 exceeds that of the other clones, we looked for mutations that might affect expression or activity of other inner-membrane glucose transporters. CV103 has two mutations in ptsI, which encodes enzyme I of the E. coli sugar phosphotransferase system. This protein is active in its dimeric form and participates in glucose uptake by accepting a phosphate group from phosphoenolpyruvate (PEP) and passing it via phosphocarrier protein Hpr to sugar-specific enzymes such as PtsG, which then use phosphate to “charge” incoming sugars (Fig. 2 and 3) [69,70]. The mutations in ptsI result in two amino acid substitutions (A328S, M518I), both of which occur in the C-terminal region responsible for binding PEP and dimerization. These may be reasonably expected to alter PtsI activity [71,72]; aa 328 is close to two residues that are part of the PEP binding site (R332 and D335), while aa 518 is close to C502, which is required for phosphotransfer [73,74]. When ~PO₄ is not passed along by PtsI, there can be regulatory consequences: unphosphorylated PtsG can bind the transcriptional regulator Mic and prevent it from negatively regulating expression of its targets which include ptsG itself, manXYZ and malT [75–77]. Thus, these mutations could also explain the up-regulation of manXYZ in CV103, and lead to higher expression of the LamB glycoporin though derepression of malT transcription [78]. PtsI has also been shown to interact directly with other phospho-enzymes, notably acetate kinase (ackA), thus mutations in this protein may also affect acetate excretion [79].

Mutations in central metabolism may contribute to CV103’s superior performance and propensity to create niches

Although glucose transport is increased in all evolved isolates relative to their common ancestor, and especially so in CV103, expression of many glycolytic genes is decreased (Fig. 4). This trend has also been observed in yeast evolution experiments carried out under glucose limitation, and may reflect the selective advantage of an energy conservation strategy under low nutrient conditions [80]. Also, the hypermutator lineage that gave rise to CV103 accumulated a number of mutations predicted to impact conversions in glycolysis, fermentation, and the TCA cycle. Certain of these mutations likely underlie CV103’s superior glucose uptake kinetics, but some may also favor excretion of and restricted access to overflow metabolites, which opens up new niches for other genotypes.

Interconversion of fructose-6-phosphate and fructose-1,6-bisphosphate

In CV103, transcript levels of pfkB, which encodes PfkII, a secondary enzyme that converts fructose-6-phosphate into
fructose-1,6-bisphosphate, is more highly expressed in CV103 than in the other strains; CV103 also carries a missense mutation in this gene (Q201H) (Fig. 4). It is unclear whether this mutation is beneficial, as PfkII is thought to be responsible for less than 5% of the phosphofructokinase activity in E. coli [81]. However, PfkII can also use tagatose-6-phosphate as a substrate [82]. CV103 has a surprising number of differences compared to the other clones in the expression and sequence of other genes in the galactitol/tagatose-6-P glycolytic pathway, including downregulation of *gatZ* (encoding tagatose-1,6-bisphosphate aldolase 2) and *gatABC* (galactitol PTS permease) as well as a mutation in the gene for GatY (G49*) (Table S2). As in a number of other K12-derived strains, the *gat* operon is likely constitutively expressed in JA122 due to IS3E element insertion in the galactitol regulator (GatR) gene [83]. Interestingly, increased expression of *gat* genes has been observed in experiments where *E. coli* has been evolved under lactulose and/or methyl-galactoside limitation [12,84,85].

Enolase, responsible for the conversion of 2-phosphoglycerate into phosphoenolpyruvate (PEP), has two mutations in CV103, one silent and one missense (L60L, A37S) (Fig. 4 and Table S2). Aside from the regulatory role played by its product, PEP, enolase participates in the degradation of certain RNAs as part of the degradosome [86]. In particular, enolase is needed to degrade *ptsG*

---

**Figure 3. Gene expression and SNPs among loci that mediate glycerol and acetate uptake/metabolism.** As in Figure 2, a green color indicates that the gene has a lower transcription level in monoculture chemostats relative to the ancestor JA122. A red color denotes elevated monoculture transcript levels while grey denotes no change in transcript level. Expression levels of loci that vary significantly in a strain-specific manner are shown in blue; for these genes, the normalized log, expression ratios of the evolved strain relative to JA122 are shown in the inset table (for details see Materials and Methods). Positive values indicate increased expression in the evolved isolates while negative values denote decreased expression. Transcript ratios for these genes, relative to the ancestor JA122 grown under identical conditions, are presented in the inset table. Stars and corresponding text denote the location and type of particular SNPs. Consistent with strain representations in Figure 1, light blue stars indicate ancestral mutations present in JA122, purple indicates SNPs present in CV101, yellow indicates SNPs in CV103, green indicates those in CV115/116, brown indicates SNPs shared by CV101/CV103/CV115/CV116 and pink denotes SNPs shared by CV101 and CV115/116.

doi:10.1371/journal.pgen.1004430.g003
mRNA when intracellular levels of G-6-P are high (i.e., during phosphosugar stress) [86,87]. This interaction also involves the sRNA SgrS, Hfq and Pnp (polynucleotide phosphorylase), in which CV103 also has a substitution (P104Q) [50,87–89].

Table S2). ptsG mRNA is not degraded when Hfq is mutated [90], thus in CV103 the combined action of mutations in enolase, Pnp and Hfq may increase longevity of ptsG transcripts.
From pyruvate to acetyl CoA and/or acetate

Pyruvate is a major metabolite that sits at the branch point between glycolysis, the TCA cycle and fermentation (Fig. 4). Measurements of intracellular pyruvate indicated that all of the evolved isolates had significantly less intracellular pyruvate than their common ancestor under glucose-limited conditions, although no significant differences in pyruvate concentration could be detected among the evolved strains (Table S3).

The primary route for oxidation of pyruvate during aerobic growth is pyruvate dehydrogenase (PDH), a three-enzyme complex that catalyzes the conversion of pyruvate into acetyl-CoA and contributes to the redox burden by transferring electrons to NAD+. (Fig. 4 and 5). Protein profiling of CV103 suggested the presence of a mutation in one of the three PDH enzymes, ypd, severe enough to eliminate the corresponding spot on a 2D gel [51]. Whole genome sequencing confirmed a missense mutation in the CV103 ypd gene that results in an amino acid substitution (F76L) in the N-terminal portion of the translated protein (Table S2). Because this substitution occurs in the FAD binding domain, it likely affects electron transfer from the reduced co-factor FADH2 to NAD+. PDH specific activity in CV103 was previously shown to be 2–3 fold lower than that in the other strains, indicating that this mutation does indeed have a negative effect [51]. Loss of ypd activity can lead to phenotypic changes consistent with many of the unique characteristics of CV103. In a screen for E. coli knockouts with extended lifespans, ypd null mutants were identified that had extended survival compared to wild-type E. coli K-12 MG1655 [91]. This enhanced survival was accompanied by reduced growth rate, reduced stationary phase cell density, reduced oxygen consumption, reduced respiration and increased accumulation of extracellular acetate, many of which phenotypes are exhibited by CV103 (Table 1, [3]).

The CV103 rapidly consumes glucose while excreting acetate. This suggests that a mutation affecting the conversion of pyruvate into acetyl-CoA would be retained in this strain. An ypd knockout can still grow on glucose and produce acetate, but it does so more slowly by using an alternate route to acetate, pyruvate oxidase (PoxB) [92]. In an effort to determine whether CV103 might be using an alternate enzymatic pathway to convert pyruvate into acetate, we compared gene expression and sequence data for the three alternative pyruvate oxidation pathways: pyruvate oxidase (POX), pyruvate:flavodoxin oxidoreductase (PFOR) and pyruvate formate-lyase (PFL) (Fig. 5).

The POX pathway directly converts pyruvate into acetate without concomitant ATP generation (Fig. 5). Nevertheless, this pathway is active during growth on glucose and can substitute for PDH if highly expressed [93]. In our system, expression of the gene for POX (poxB) is downregulated in all four evolved isolates relative to their common ancestor. This is not unexpected, as other glycolytic genes are downregulated. However, CV103 has a nonsense mutation at amino acid 14 of poxB (E14*) which, given its position at the extreme N-terminus, is likely to completely inactivate the protein. While the ancestral supE4 amber suppressor might allow limited poxB translation to occur, it is unlikely that the POX pathway produces an appreciable amount of acetate in CV103.

A second gene involved in acetate excretion, ydbK, is expressed under both anaerobic and aerobic (albeit at very low levels) conditions (Fig. 5), [94,95]. ydbK encodes PFOR, which catalyzes the conversion of pyruvate into acetyl-CoA with concomitant reduction of flavodoxin or ferredoxin [96]. Interestingly, both of these reduced molecules can be used to re-activate oxygen sensitive PFL (see below) [94]. CV103 carries a missense mutation of unknown effect (R539L) in ydbK; however, because this ORF was not represented on our expression array, its transcription level is unknown.

Finally, acetyl-CoA can also be produced by the cleavage of pyruvate by pyruvate formate-lyase (PFL). PFL activity is primarily associated with the pflB gene but is also encoded by lpdE, ybiW and pflD [97–99]. PFL is oxygen sensitive and thus the primary route to acetyl-CoA under anaerobic conditions, but it is transcribed, active and useful during microaerobiosis [100–103]. Moreover, production of PFL requires a smaller anabolic investment than production of PDH, and may thus be preferred under conditions of nutrient stress [104]. Functional PFL requires both transcriptional (regulated by FNR and ArcA/B) and post-translational activation by the activating enzyme PIAA or the alternate activator Yfd (103,105). Consistent with the downregulation of many glycolytic genes, transcript levels of both the primary PFL (pflB) and its activator (pflD) are also downregulated across all of the evolved strains relative to the ancestor JA122. The gene for pflB is unchanged, and pflA has a silent substitution in CV103 (1205). By contrast, transcript levels of the three alternate PFLs (ldeE, ybiW, pflD) and their alternate activator yfdA are not altered in any of the evolved strains relative to JA122, though CV103 has SNPs in the genes that encode two of the alternate PFLs: YbiW (A1469S) and PFD (G513G) (Fig. 5).

Perturbations in pyruvate transformation are known to impact a key state variable in central metabolism: cellular redox balance. NADH accumulates in cells with high glycolytic flux because it cannot be re-oxidized as fast as it is generated. This increased redox ratio (NADH/NAD+) creates a cellular response reminiscent of anaerobiosis, stimulating the cell to direct pyruvate toward overflow metabolites and leading to repression of TCA cycle genes such as isocitrate dehydrogenase (icd) and citrate synthase (gltA) [106]. Given the relatively rapid uptake of glucose and higher rate of acetate production by CV103, stress in the form of high redox ratio may have influenced the evolution of this strain. NAD+ regeneration typically occurs downstream of pyruvate either via the TCA cycle or fermentation, but could also occur via the conversion of DHAP into glycerol (see section on glycerol metabolism below).

Mutations in CV103 likely diminish flux through aerobic pathways

(Fig. 6) The TCA cycle consists of eight steps beginning with the conversion of acetyl-CoA to citrate and ending with the conversion of malate to oxaloacetate. Expression profiling of each evolved isolate showed that relative to their common ancestor, levels of transcripts for proteins involved in three of these steps (α-ketoglutarate dehydrogenase, succinate dehydrogenase and fumarase) were upregulated across all evolved strains. At two other steps, aconitase (aceA) and isocitrate dehydrogenase (icdA), transcript levels were elevated in three of four strains, but reduced in CV103. Whole-genome sequencing uncovered intergenic SNPs in CV103 that could impact flux at the icd branch point connecting the TCA cycle with the glyoxylate bypass. In the ancestral strain, JA122, the negative regulator of the glyoxylate bypass, wcrR, has a promoter mutation which likely affects wcr negative autoregulation, leading to higher expression and concomitant repression of the glyoxylate shunt genes aceA, aceB and aceK. AceK negatively impacts flux through isocitrate dehydrogenase by phosphorylation, diverting carbon through isocitrate lyase (aceA) and malate synthase (aceB) [107]. Thus, in the ancestor increased IcrR can reasonably be expected both to decrease transcription of glyoxylate bypass enzymes and to prevent inactivation of Icd by AceK. While CV101 and CV116 exhibit increased relative expression of icdA and aceB, CV103
shows decreased expression of both, perhaps owing to a C→A mutation that lies between the aceA and aceK open reading frames.

Cellular pyruvate levels have long been thought to modulate TCA cycle flux, principally by modulating isocitrate dehydrogenase activity (Fig. 6, [108,109]). The CV103-specific mutations described above may also affect the pool of pyruvate. lpd mutants have been shown to increase Entner-Doudoroff and glyoxylate shunt activities and to decrease TCA cycle activity [92], while poxB null mutations have been shown to repress citrate synthase and malate dehydrogenase, and to activate acs [110]. CV103 contains mutations in the lpd FAD binding domain, which likely accounts for the 2–3 fold reduction of its pyruvate dehydrogenase activity, as well as a nonsense mutation at amino acid 14 of poxB (E14*) that likely inactivates this protein (Table S2). Together these mutations could be expected to impede flux through the TCA cycle in CV103. Conspicuous among genes differentially expressed between evolved strains and their ancestor, and among the evolved strains themselves, were those encoding respiratory proteins that have a high H+/O coupling ratio. When each was grown by itself in glucose-limited chemostats, transcript levels for genes in the cyoABCDE operon that encode cytochrome oxidase subunits were significantly increased in CV101 and CV115/116, but significantly decreased in CV103, relative to their common ancestor JA122 [30].

Finally, CV103 also shows gene expression differences and mutations affecting the respiratory chain. Under conditions of nutrient stress, metabolic cost-benefit analyses predict E. coli will shift from the more efficient but anabolically expensive NADH:ubiquinone oxidoreductase II (ndh)/cytochrome bd oxidase (cydAB) pairing [104,111,112]. Compared with JA122, expression of nuoGHJKLMN is 1.1-fold lower in CV103 but 1.3 to 1.5-fold higher in CV101, CV116 and CV115. Similarly, transcript levels of cyoABC are 2.1-fold lower in CV103 and 1.3 to 1.9-fold higher in the other evolved strains. As noted, CV103 also has missense mutations that affect NuoM (L336F) and NuoI (R93L) as well as a silent substitution in NuoE (L14L). The effect of these differences in expression and sequence of NDH-I and cytochrome bo oxidase is unknown, but they may impact cellular redox balance.

In summary, by considering the whole genome sequencing data in the light of transcriptome [30], proteome [51], and phenotype data [31] a coherent picture emerges of the dominant clone, CV103, being highly fermentative but impaired in aerobic pathways, resulting in the production of overflow metabolites, which in the case of acetate creates a redox imbalance, but in the case of glycerol/glycerol phosphate, may provide a means to correct this imbalance. Because the other consortium members, like their common ancestor, are respiro-fermentative they are capable of exploiting the new biochemical niches created by CV103.

The genetics of clonal reinforcement (Fig. 3) We previously showed that differential production and scavenging of acetate, glycerol and/or glycerol phosphate explained stable coexistence of multiple genotypes under glucose limitation [31,36]. Concerning acetate, whose excretion by CV103 supports growth of CV101, whole genome re-sequencing confirmed previously identified mutations in CV101 that up-regulate
acetyl CoA-synthetase (Acs) and the acetate transporter ActP, as well as an ancestral mutation shared by the other clones that prevents efficient re-uptake of acetate by disrupting a CRP regulatory site in the \textit{acs} promoter [30]. No additional mutations affecting acetate excretion or uptake were found in any of the evolved strains’ genomes, so it is likely that these two alone explain the ability of CV101 to cross feed on the acetate produced by CV103.

Regarding glycerol and/or glycerol phosphate, previous experiments had shown that chemostat-grown CV116 assimilate radiolabeled glycerol 50% faster than all other strains, and that when CV116 and CV103 were co-cultured under glucose limitation, the addition of exogenous glycerol [31] or glycerol 3-phosphate caused the frequency of CV116 to increase. All of the evolved strains carry an ancestral mutation (G55A) in the DNA-binding transcriptional repressor, GlpR, that renders it constitutively inactive [113]. We also discovered a CV116-specific SNP in \textit{glpK} encoding glycerol kinase, the first step in glycerol assimilation (Table S2), though this is a silent substitution at a site (G232G; GGC→GGA) not currently known to be associated with \textit{glpK} regulation. Transcriptional profiling of GlpR regulated genes showed that the transcription unit containing glycerol kinase (\textit{glpFKX}) remained at ancestral levels, while genes for the glycerol-3-phosphate:phosphate antiporter GlpT (\textit{glpT}), and the divergently transcribed anaerobic glycerol-3-phosphate dehydrogenase genes (\textit{glpABC}) were upregulated across all isolates relative to their common ancestor [30].

The absence of differences in \textit{glp} gene expression between CV103 and CV116 led us to explore the possibility that glycerol and/or glycerol phosphate cross-feeding might be mediated by other mechanisms controlling the assimilation and production of these metabolites. For example, post-transcriptional regulation in CV103 could explain why JA122, CV101 and CV116 have three-fold higher specific activity of glycerol kinase, and two-fold higher specific activity of glycerol-3-phosphate dehydrogenase than CV103 [31]. GlpK is known to be post-transcriptionally inactivated by the unphosphorylated version of the glucose-specific PTS enzyme \textit{IIA}^{Glc} (enr), or by an excess of the effector molecule fructose-1,
As described above, IIAGlc is phosphorylated via its interaction with the sugar non-specific EI (ptsI) and HPr (ptsH) (Fig. 3). When glucose is transported through the inner membrane by the Enzyme IIGlc complex (ptsG/crr), IIAGlc (crr) transfers its phosphate to IIBCglc (ptsG), which then phosphorylates glucose to yield intracellular glucose-6-phosphate (reviewed in [114]). High levels of unphosphorylated IIAGlc signal glucose abundance, and inhibit enzymes needed for catabolism of alternate carbon sources such as glycerol and glycerol phosphate. Specifically, unphosphorylated IIAGlc is known to inhibit glycerol kinase activity [114,115,117]. We have already noted two mutations in CV103 ptsI, which, by impairing EI, could lead to excess unphosphorylated IIAGlc, which would inhibit glycerol kinase (Fig. 3), and thereby restrict CV103’s access to glycerol.

CV103 may also have an excess of another potent glycerol kinase inhibitor: fructose-1,6-bisphosphate (FBP) [116,118]. We previously noted that relative to JA122 and the other evolved isolates, CV103 has enhanced expression of pfkB, which encodes the minor FBP creating enzyme, and lower levels of fbp, which encodes the reverse enzyme fructose bisphosphatase [30]. Whole-genome sequencing revealed a missense mutation of unknown effect in pfkB (Q201H). These distinctive features of CV103, combined with its demonstrated capacity for enhanced glucose transport and assimilation, may produce elevated levels of the GlpK inhibitor, FBP, further impeding CV103’s ability to assimilate glycerol.

To understand why CV103 might release glycerol and/or glycerol phosphate as metabolic by-products, we examined the transcript levels and sequences of genes encoding proteins involved in their production (Fig. 3 and Fig. 7). Glycerol can be generated by E. coli as either (1) a by-product of phospholipid synthesis from sn-glycerol-3-phosphate, (2) an end-product of the detoxification of dihydroxyacetone phosphate/methylglyoxal, or (3) via hydrolysis of glycerol-1-phosphate [119]. CV103 has multiple mutations that affect nearly every step of phospholipid biosynthesis (Fig. 7A). While the number of SNPs suggests CV103 reaps some benefit from altering phospholipid production, these mutations should also have the effect of restricting glycerol and phospholipid formation from glycerol-3-phosphate. The second route to glycerol, detoxification of DHAP/methylglyoxal, is catalyzed by the reversible glycerol dehydrogenase GldA (Fig. 7B). This reaction is likely to be useful for a strain such as CV103 that is consuming glucose at a high rate, because it prevents the buildup of DHAP and subsequent production of toxic methylglyoxal, and, significantly, because it re-oxidizes NADH [119,120]. CV103 may also have higher amounts of intracellular DHAP and methylglyoxal as a consequence of a non-synonymous mutation (A6D) in the gene for glycerol-3-phosphate dehydrogenase GpsA, which favors the production of DHAP from G3P (Fig. 3 and 7). No SNPs affecting GldA were found, and gldA transcript levels were not up- or downregulated in any of the isolates. The third mechanism for the production of glycerol is via the activity of the enzyme YbT. YbT catalyzes the conversion of glycerol-1-phosphate into glycerol and its gene is truncated in CV103 (E22*), resulting in a defect likely expected to increase the pool of glycerol phosphate but not glycerol [121].

6-bisphosphate [114–116]. As described above, IIAGlc is phosphorylated via its interaction with the sugar non-specific EI (ptsI)/HPr (ptsH) (Fig. 3). When glucose is transported through the inner membrane by the Enzyme II(Glc complex (ptsG/crr), IIAGlc (crr) transfers its phosphate to IIBCglc (ptsG), which then phosphorylates glucose to yield intracellular glucose-6-phosphate (reviewed in [114]). High levels of unphosphorylated IIAGlc signal glucose abundance, and inhibit enzymes needed for catabolism of alternate carbon sources such as glycerol and glycerol phosphate. Specifically, unphosphorylated IIAGlc is known to inhibit glycerol kinase activity [114,115,117]. We have already noted two mutations in CV103 ptsI, which, by impairing EI, could lead to excess unphosphorylated IIAGlc, which would inhibit glycerol kinase (Fig. 3), and thereby restrict CV103’s access to glycerol.

CV103 may also have an excess of another potent glycerol kinase inhibitor: fructose-1,6-bisphosphate (FBP) [116,118]. We previously noted that relative to JA122 and the other evolved isolates, CV103 has enhanced expression of pfkB, which encodes the minor FBP creating enzyme, and lower levels of fbp, which encodes the reverse enzyme fructose bisphosphatase [30]. Whole-genome sequencing revealed a missense mutation of unknown effect in pfkB (Q201H). These distinctive features of CV103, combined with its demonstrated capacity for enhanced glucose transport and assimilation, may produce elevated levels of the GlpK inhibitor, FBP, further impeding CV103’s ability to assimilate glycerol.

To understand why CV103 might release glycerol and/or glycerol phosphate as metabolic by-products, we examined the transcript levels and sequences of genes encoding proteins involved in their production (Fig. 3 and Fig. 7). Glycerol can be generated by E. coli as either (1) a by-product of phospholipid synthesis from sn-glycerol-3-phosphate, (2) an end-product of the detoxification of dihydroxyacetone phosphate/methylglyoxal, or (3) via hydrolysis of glycerol-1-phosphate [119]. CV103 has multiple mutations that affect nearly every step of phospholipid biosynthesis (Fig. 7A). While the number of SNPs suggests CV103 reaps some benefit from altering phospholipid production, these mutations should also have the effect of restricting glycerol and phospholipid formation from glycerol-3-phosphate. The second route to glycerol, detoxification of DHAP/methylglyoxal, is catalyzed by the reversible glycerol dehydrogenase GldA (Fig. 7B). This reaction is likely to be useful for a strain such as CV103 that is consuming glucose at a high rate, because it prevents the buildup of DHAP and subsequent production of toxic methylglyoxal, and, significantly, because it re-oxidizes NADH [119,120]. CV103 may also have higher amounts of intracellular DHAP and methylglyoxal as a consequence of a non-synonymous mutation (A6D) in the gene for glycerol-3-phosphate dehydrogenase GpsA, which favors the production of DHAP from G3P (Fig. 3 and 7). No SNPs affecting GldA were found, and gldA transcript levels were not up- or downregulated in any of the isolates. The third mechanism for the production of glycerol is via the activity of the enzyme YbT. YbT catalyzes the conversion of glycerol-1-phosphate into glycerol and its gene is truncated in CV103 (E22*), resulting in a defect likely expected to increase the pool of glycerol phosphate but not glycerol [121].
In short, there are multiple, non-mutually exclusive reasons to explain why CV103 has both a diminished capacity to assimilate glycerol as well as an increased propensity to produce glycerol and/or glycerol phosphate as overflow metabolites.

**Clonal reinforcement drives the evolution of biocomplexity**

Biocomplexity, here defined as stable co-existence of multiple genotypes, can emerge in clonal populations evolving in environments that are spatially structured with respect to their physical features [10,13] or temporally structured with respect to availability of limiting nutrients [8,11,122]. Under these conditions, the emergence of complexity can be explained in terms of classical niche theory [123]. Biocomplexity can also emerge in clonal populations evolving in simple, unstructured environments where reproduction is continuously limited by a single resource [3,25,47,124]. There, population genetic complexity can be maintained by clonal interference, wherein competitive interactions preclude fixation of a fittest genotype [25], or by clonal reinforcement where one clone supports growth by others via the excretion of metabolizable substrates [3]. Our results bear out certain predictions arising from Mazancourt and Schwartz's resource ratio theory of cooperation [125]. In their model, two species initially competing for two resources can evolve towards a cooperative trading relationship one of whose emergent properties is enhanced resource utilization. Conditions that favor this outcome include low mortality, low resource levels and differential efficiency between species at depleting limiting resources. Our community evolved under similar conditions: clonal lineages persisted for many generations under chronic resource limitation and gained differential access to available resources. Two differences between their model and our system are that our population was founded by a single clone initially limited on one resource, and that the mutations which gave one of its descendants preferential access to that resource have pleiotropic effects that favor production of secondary resources on which other clones can thrive. As for why this genotype arose, we know by PCR analysis that key CV103-specific mutations at malK, mutM, ptsI, lpf and lpd are present in the earliest Helling et al. [3] population sample archived (~350 generations) (data not shown).

**Clonal reinforcement is driven by the genotype best able to acquire the limiting resource**

A category of mutations certain to prove beneficial under nutrient limitation is one that favors increased uptake of the limiting nutrient, in this case glucose. Levels of residual glucose in steady state CV103 monocultures are significantly lower than in those of the other evolved strains and their common ancestor [31], consistent with CV103's more rapid uptake of the non-metabolizable glucose analogue α-MG [3] and its higher expression of LamB glycoporin [51]. The amino acid substitution in MalK (D297E) likely diminishes its ability to deactivate MalE and, by extension, activates lamB expression. A mutation in Hfq (Q52H), that may negatively affect translation of the stress response global regulator RpoS, would enhance glucose scavenging via LamB and PtsG [126,127].

The evolution of a superior resource acquisition strategy results in trade-offs

Because glucose consumption in most organisms, including *E. coli*, is greater under anaerobic than under aerobic conditions (e.g., [128] and refs within), and because enhanced glucose consumption is an adaptive strategy selected for in glucose-limited chemostats, it is perhaps not surprising that, relative to the other evolved clones and their common ancestor, the metabolic profile of CV103 appears to be fermentative. Multiple lines of evidence support this interpretation: (i) TCA cycle genes encoding aconitase (acsA) and isocitrate dehydrogenase (icl) are downregulated (Fig. 6), (ii) expression of the *cyoB/CyD* operon encoding cytochrome oxidase is reduced [30], (iii) multiple mutations suggest CV103 is dealing with excess formate production (Table S4), and (iv) steady state chemostats of CV103 contain appreciable residual concentrations of overflow metabolites [3,31]. While we found no CV103-specific mutations that would constitutively repress aerobic pathways, missense mutations at *pdf* and *msdc*, and a nonsense mutation at *psb* could be expected to diminish TCA cycle flux. Moreover, increased glucose consumption by *E. coli* under glucose limitation has been shown to repress both respiration and the TCA cycle via changes in global regulators such redox balance and pyruvate [106]. High glucose consumption also strongly represses transcription of *acs* encoding acetyl CoA synthetase [106], which helps to explain why no activity of this acetate scavenging enzyme can be detected in CV103 monocultures [31], and why extracellular acetate is present at a dilution rate (D = 0.2 h⁻¹) where none is expected. We therefore conclude that CV103's fermentative metabolism arises as a consequence of selection for enhanced glucose consumption.

The CV103 lineage's response to selection for enhanced glucose consumption results in trade-offs that critically determine its role in niche construction. At lower dilution rates *E. coli* growing under aerobic glucose-limitation usually carries out a high-yield metabolism that converts all available glucose to CO₂ ([129] and refs therein). Indeed, under aerobic conditions the switch from respiratory to respiro-fermentative metabolism typically occurs at high growth rates and glucose concentrations, resulting in the production of overflow metabolites such as acetate. (Under anaerobic conditions, *E. coli* typically ferments all glucose to CO₂, acetate and ethanol [130].) The excretion of overflow metabolites under conditions of restricted TCA flux creates an imbalance in cellular redox, because acetate production in *E. coli*, unlike ethanol production in yeast and lactate production in animals, fails to regenerate NAD+ from the NADH formed by glycolysis. In CV103 this problem may be further exacerbated by mutations affecting pyruvate transformation, which would place an additional premium on NAD+generating processes, including reactions that lead to glycerol and glycerol-3-phosphate production. Concerning the apparent pleiotropic effects arising from constitutively high glucose consumption in the CV103 background, it is interesting to note that a recently discovered mechanism to limit the production of overflow metabolites is to overexpress the small RNA SgrS, which effectively reduces the rate of glucose consumption [131]. SgrS, complexed with its binding partner, diminishes cells’ ability to create new sugar transporters, in particular PtsG [132]. In CV103 the binding partner of SgrS, Hfq, has a missense mutation (Q52H), which may contribute to its rapid uptake, but incomplete assimilation of glucose.

The glucose scavenging adaptations seen in CV103, which contribute to its persistence and its ability to create new niches, contrasts with metabolic adaptations commonly observed in experimentally evolved yeast [21]. Yeast cultured at similar dilutions rates adapt to aerobic glucose limitation by reversing the Pasteur effect, switching from fermentative to respiratory metabolism, which results in an enormous gain in their reproductive output. While yeast evolution experiments typically begin with populations that produce appreciable amounts of overflow metabolites, mainly ethanol, strains quickly evolve a metabolic strategy that produces essentially no residual carbon,
because it has accumulated mutations in enzymes required for glycerol and glycerol phosphate assimilation.

By contrast, CV103 likely has limited access to these resources downstream effects include inhibition of glycerol kinase, the rate-scavenging, and another in \textit{glpR} retains two ancestral mutations: one in \textit{lpd}, whose reaction whose dehydrogenase steps help regenerate NAD pressure only increased with the advent of the CV103 lineage, making it likely that selection pressure for acetate scavenging existed at the outset of these experiments. This changes likely favor the formation of glycerol through side-

The cost paid by one genotype to acquire the limiting resource opens up new niches for others Superior glucose consumption by the largely fermentative strain CV103 helps to create multiple resources in an environment where initially only one was limiting, thereby providing a selective advantage to genotypes that can access those secondary resources. The mechanism by which strain CV101 gains preferential access to acetate is straightforward and relates to insertion of an IS30 element in the upstream regulatory region of \textit{acs} encoding the acetate-scavenging enzyme acetyl CoA synthetase \[36\]. The \textit{acs} locus is misregulated in the ancestral background due to a mutation in the CRP binding site. Indeed, we observed appreciable amounts of acetate in JA122 monocultures at steady state \[31\], making it likely that selection pressure for acetate scavenging existed at the outset of these experiments. This pressure only increased with the advent of the CV103 lineage, which evolved a superior mechanism for glucose consumption, but which produces even more acetate because it retains the ancestral defect, and because increased flux through fermentative pathways represses \textit{acs} transcription.

The basic CV103 phenotype is additionally supported by mutations, in particular those related to pyruvate oxidation (e.g. \textit{lpd}), that accentuate changes in redox potential typical of fermentative cells whose overflow metabolite is acetate. These changes likely favor the formation of glycerol through side-reactions whose dehydrogenase steps help regenerate NAD+. CV116 has preferential access to this other metabolite because it retains two ancestral mutations: one in \textit{acs} that impairs acetate scavenging, and another in \textit{gpkR} that constitutively derepresses enzymes required for glycerol and glycerol phosphate assimilation. By contrast, CV103 likely has limited access to these resources because it has accumulated mutations in \textit{psd} and \textit{gpkB}, whose downstream effects include inhibition of glycerol kinase, the rate-limiting step in glycerol assimilation.

Reinforcement may underlie clonal diversity in the laboratory and in nature

The evolution and persistence of multiple genotypes in a population of asexual organisms supported by a single resource seems to violate the principles of competitive exclusion \[134,135\] and periodic selection \[1,136\]. Yet this phenomenon has now been observed in a variety of simple experimental systems, in particular those that are temporally \[8,16\] or spatially structured \[10,13\]. Although continuous nutrient-limited chemostats are unstructured in both respects, a number of studies have now shown that multiple bacterial lineages, some of them mutators, can arise and coexist in a single chemostat vessel \[3,5,8,57,82,137\]. However, none of these studies have provided evidence that biocomplexity can arise and be sustained by means of the trading relationships we call clonal reinforcement. We contend that this mechanism may be quite common, as it is formally analogous to and may sometimes be a precursor to syntrophy, a ubiquitous feature of natural bacterial communities \[138\]. Clonal reinforcement may also be at work in clinically relevant environments. For example, the dominant clones in tumors are often fermentative, differing markedly from normal (ancestral) tissue in their demand for \textit{O}$_2$ and nutrients, their production of \textit{CO}$_2$ (see \[139\] and refs therein), and in their release of overflow metabolites that acidify the local environment \[140\]. Such cells may create opportunities for subpopulations to follow independent evolutionary trajectories that lead to further genetic differentiation, perhaps even to changes in their contact inhibition and drug resistance phenotypes. Chronic bacterial infections are also genetically heterogeneous \[141\], and can even be supported by syntrophic interactions \[142\]. We may therefore reasonably ask: to what extent does clonal reinforcement enable subpopulations in tumors and chronic infections to differentiate and become more resistant to chemotherapy and/or less visible to the immune system? The answers to these questions have far-reaching implications.

Materials and Methods

\textbf{Strains, media and culture conditions} \textit{Escherichia coli} JA122, CV101, CV103, CV115 and CV116 were maintained as permanent frozen stocks and stored at $\sim$80°C in 20% glycerol. Davis Minimal medium was used for all liquid cultures with 0.025% glucose added for batch cultures and 0.0125% for chemostats \[143\]. Chemostat cultures were initiated using colonies picked from Tryptone Agar (TA) plates and outgrown in Davis minimal batch medium overnight. For transcriptional profiling, total protein and metabolite assays chemostats were maintained at 30°C with a dilution rate of $\sim$0.2 h$^{-1}$ for approximately 70 hours ($\sim$15 generations). At the end of each chemostat run, three aliquots of 50 mL culture were rapidly filtered onto 0.2 \textmu m nylon membranes, flash-frozen in liquid nitrogen and stored at $\sim$80°C.

\textbf{Mutation rate determination} Strains were streaked to colonies on LB agar overnight. Twelve colonies were picked in their entirety for each strain and inoculated into 3 mL of liquid LB, then grown overnight at 37°C. Subsequently, a portion of the liquid cultures was spread on LB plates containing 300 \textmu g mL$^{-1}$ rifampicin. Colonies were counted 48 h after spreading. To determine the titer, three cultures for each strain were spread on LB aga plates at 10$^{-7}$, 10$^{-8}$ and 10$^{-9}$ dilutions. Colonies were counted and titers determined as the average of the three 10$^{-8}$ dilutions. Mutation rates per cell per generation were calculated using maximum likelihood calculator FALCOR (http://www.keshavsingh.org/protocols/FALCOR.html) \[144\].

\textbf{Pyruvate assays} Following fast filtration and disruption by sonication intracellular pyruvate was determined on chemostat-grown cells using the Pyruvate Assay kit and PicoProbe (Biovision, Milpitas, CA, K609 and K317) as directed by the manufacturer’s guidelines. Estimated values were normalized to cell protein, which was determined via the Pierce BCA Protein Assay Kit (Cat. # 23227, Thermo Scientific, Rockford, IL) using BSA as standard.

\textbf{Transcriptional profiling} Global gene profiling was described in detail in a previous publication \[30\]. Briefly, total RNA was extracted from triplicate cultures of chemostat-grown JA122, CV101, CV103 and CV116 ($D = 0.2$ h$^{-1}$, 0.0125% glucose) and hybridized to full-length open reading frame PCR products spotted onto aminosilane-coated slides. Raw data was analyzed using TIGR MIDAS and MeV software pipelines (www.tm4.org) and Significance Analysis of Microarrays (SAM) \[145\] was used to examine expression differences between strains using a multi-class comparison consisting of four groups. Similarities among strains were identified using one-class SAM and differences between the strains were
examined using a 4-class SAM. δ cutoffs were set at the 0% FDR threshold (i.e. the highest δ value that gave a median false discovery rate of 0%). Average (mean) log₂ ratios were calculated after SAM analysis using Microsoft Excel and represent the relative expression ratios of each evolved isolate compared to their common ancestor.

Nucleic acid extraction
Genomic DNA for Illumina sequencing was extracted from cells grown in batch culture using a modification of methods described by Syn and Swarup [146]. Subsequent to DNA precipitation, spun pellets were resuspended in 1XTE, 1 mM EDTA, pH 8.0 containing 50 μg/mL DNase-free RNase A and incubated at 37°C for 30 minutes. Samples were re-extracted once with phenol:chloroform (3:1), once with phenol:chloroform (1:1) and twice with chloroform and then precipitated with EtOH using standard techniques. Following re-precipitation, DNA was dissolved in TE.

Illumina sequencing
Single-end 36 bp sequencing libraries were created using the Illumina Genomic DNA Sample Prep Kit according to manufacturer’s instructions (5 μg input genomic DNA), and sequencing flow cells were prepared using the Illumina Standard Cluster Generation Kit. Samples were sequenced on the Illumina Genome Analyzer II, and image analysis and data extraction were performed using Illumina RTA 1.5.35.0. Reads (with qualities) were aligned to the K12 reference genome (gi|49175990|ref|NC_000913.2) using BWA v0.5.8 [147] with default parameters. Whole-genome pileup files were generated using SAMTools v0.1.8–18 [148] and single-nucleotide polymorphisms due to the evolution were called using custom perl scripts that compared each evolved strain with the original ancestor. Briefly, SNPs passed the filter if they were represented in at least 40% of reads in the evolved strain and at most 10% in the reference strain, with at least 5 reads covering the position in both strains. Additional heuristic filters included a confirming read from both strands, and no more than one ambiguous SNP call (“N”) or deletion (“*”) at that position. Insertions and deletions relative to the K12 reference sequence were identified using Bresaq v. 0.18 with default parameters (http://barricklab.org/bresaq; [32]).

Sanger sequencing
Thirty-seven SNPs identified by Illumina sequencing were verified by Sanger sequencing. PCR products were generated using Fermentas DreamTaq PCR Master Mix (Thermo Scientific, Waltham, MA) following manufacturer’s instructions, treated with ExoSAP-IT (Affymetrix, Santa Clara, CA) and sequenced on an ABI3730XL machine by High Throughput Genomics Center (Seattle, WA) (Table S5).

Supporting Information
Table S1 Genes and transcription units (T.U.) affected by more than one mutation. (PDF)
Table S2 All mutations. (PDF)
Table S3 Intracellular pyruvate. (PDF)
Table S4 Mutations associated with formate metabolism. (PDF)
Table S5 SNP verification primers. (PDF)

Acknowledgments
The authors gratefully acknowledge helpful discussions with Tom Ferenci and Ivana Gudej, thank Matt Herron and Karen Schmidt for critically reading the manuscript, and Jacob Childress for assisting with fluctuation tests.

Author Contributions
Conceived and designed the experiments: MK JW GS FR. Performed the experiments: MK JW EK. Analyzed the data: MK JW EK GS FR. Contributed reagents/materials/analysis tools: GS. Wrote the paper: MK JW EK JA GS FR.

References


102. Sawers G, Bock A (1989) Novel transcriptional control of the pyruvate formate-
83. Nobelmann B, Lengeler JW (1996) Molecular analysis of the gat genes from
75. Overflow metabolism in Escherichia coli during steady-state growth:
74. Zwaig N, Lin EC (1966) Feedback inhibition of glycerol kinase, a catabolic