How Archiving by Freezing Affects the Genome-Scale Diversity of *Escherichia coli* Populations

Kathleen Sprouffske[1,2], José Aguilar-Rodríguez[1,2], and Andreas Wagner[1,2,3,*]

[1]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland

[3]The Santa Fe Institute, Santa Fe, NM, USA

*Corresponding author: E-mail: andreas.wagner@ieu.uzh.ch.

Accepted: March 11, 2016

Abstract

In the experimental evolution of microbes such as *Escherichia coli*, many replicate populations are evolved from a common ancestor. Freezing a population sample supplemented with the cryoprotectant glycerol permits later analysis or restarting of an evolution experiment. Typically, each evolving population, and thus each sample archived in this way, consists of many unique genotypes and phenotypes. The effect of archiving on such a heterogeneous population is unknown. Here, we identified optimal archiving conditions for *E. coli*. We also used genome sequencing of archived samples to study the effects that archiving has on genomic population diversity. We observed no allele substitutions and mostly small changes in allele frequency. Nevertheless, principal component analysis of genome-scale allelic diversity shows that archiving affects diversity across many loci. We showed that this change in diversity is due to selection rather than drift. In addition, ~1% of rare alleles that occurred at low frequencies were lost after treatment. Our observations imply that archived populations may be used to conduct fitness or other phenotypic assays of populations, in which the loss of a rare allele may have negligible effects. However, caution is appropriate when sequencing populations restarted from glycerol stocks, as well as when using glycerol stocks to restart or replay evolution. This is because the loss of rare alleles can alter the future evolutionary trajectory of a population if the lost alleles were strongly beneficial.

Key words: experimental evolution, glycerol stocks, diversity, heterogeneity, population sequencing, *Escherichia coli*.

Introduction

In the experimental evolution of microbes, one or more replicate populations are seeded from a culture derived from a single ancestral clone. When experimental evolution is conducted in batch culture, a small subsample of each replicate population is regularly transferred into fresh medium. Aliquots of an evolving population are usually archived as frozen stocks at regular intervals to characterize the populations genotypically and phenotypically, or to replay evolution from any archived time point (Garland and Rose 2009; Kawecki et al. 2012; Barrick and Lenski 2013; Elena and Lenski 2003; Kassen 2014). In long-term evolution projects, evolving populations usually comprise many clonal subpopulations with different genotypes and phenotypes that change in frequency over time (Elena and Lenski 2003; Barrick and Lenski 2009; Garland and Rose 2009; Kawecki et al. 2012; Wielgoss et al. 2012; Barrick and Lenski 2013; Herron and Doebeli 2013; Kassen 2014; Levy et al. 2015). In an iconic experimental evolution study, Lenski and co-authors have evolved 12 replicate lineages of *Escherichia coli* for over 60,000 generations (Lenski et al. 1991; Fox and Lenski 2015), periodically archiving the evolving populations by supplementing their growth medium with 15% glycerol and storing them at −80 °C (Barrick et al. 2009).

In analogy to the paleontological fossil record, the frozen archive of an evolution experiment is subject to preservation biases (Efremov 1940; Behrensmeyer et al. 2000). Specifically, if not all subpopulations are equally able to survive the stress of archiving, the process of archiving may bias downstream experiments and analyses. Indeed, *E. coli* clones independently evolved from the same ancestor survive freezing and thawing at different rates in the absence of a cryoprotectant (Sleight et al. 2006, 2008; Sleight and Lenski 2007). In previous experiments in which two *E. coli* strains were mixed, archiving a mixed population changed the relative frequencies of the strains after freezing and thawing (Turner et al. 1996).

However, these observations were made in experiments to detect cross-feeding interactions and thus cannot shed light on how the diversity that arises during evolution from a common ancestor is affected by archiving. Here, we used whole genome sequencing to explicitly study the extent to which archiving affects the genomic diversity of an evolving population of *E. coli*, and to examine the potential influence of archiving on the population's evolutionary dynamics.

Archiving Populations Using Glycerol Stocks

Most commonly, *E. coli* cultures are archived by adding glycerol to a final concentration of 10–20%, freezing the culture, and storing it in an ultra-low-temperature freezer at −80 °C. Glycerol was discovered to be a good cryoprotectant in early experiments with bacteria (Keith 1913) by lowering the freezing temperature and thus reducing osmotic shock during freezing (Lovelock 1953; Mazur 1984; Gao and Critser 2000). Early experiments with *E. coli* found that 12% glycerol protected the highest fraction of cells and leaves ~55% of cells viable (Sharp 1983). A standard laboratory manual (Sambrook and Russell 2000) recommends using glycerol to a final concentration of 15%, though it has been suggested that each type of cell may have a unique optimal condition (Calcott 1986).

Previous studies showed that freezing and thawing themselves affect survival much more than the duration of keeping a sample frozen (Mazur 1984; Gao and Critser 2000; Sleight et al. 2006). Freezing that is either too slow or too fast can reduce the survival of eukaryotic cells (Mazur 1984; Gao and Critser 2000), while freezing *E. coli* either very quickly or very slowly leads to the highest survival (Dumont et al. 2004). However, slow freezing of *E. coli* is the optimal method feasible in a standard microbiology lab. The three freezing methods commonly used and previously described (Sambrook and Russell 2000; Tedeschi and De Paoli 2010) for *E. coli* differ in the freezing rate. Placing cells in a −80 °C freezer freezes them more slowly than flash-freezing in a dry-ice ethanol bath or in liquid nitrogen.

Here, we tested a range of glycerol concentrations and freezing methods to determine the best conditions for archiving the two common *E. coli* strains K12 MG1655 and REL606, an *E. coli* B derivative, as well as a hypermutable K12 derivative. We then used the optimal conditions to assess the effects of archiving an evolving and diverse population of the hypermutable *E. coli* strain.

Results

Optimal Glycerol Concentration and Freezing Conditions

We set out to identify the final glycerol concentration and freezing method that best protected two common *E. coli* strains, REL606 and K12 MG1655, as well as a hypermutable strain of K12 with a mutation rate that is 4,000-fold greater than wildtype. We archived replicate samples with glycerol to a final concentration between 0% and 50% and measured the fraction of cells that survived. For the REL606 and hypermutable strains, the optimal glycerol concentration was 15%. For the K12 strain, it was between 20% and 25% (fig. 1*A*). Fewer than 40% of cells survived freezing and thawing when supplemented with glycerol. Interestingly, in the absence of glycerol, 14% of the K12 cells survived, while in the other two strains <1% did. To determine the optimal freezing method, we compared the survival of replicate samples frozen with the three main methods described above. Slower freezing by placing cultures directly into the −80 °C freezer yielded more surviving cells for the REL606 and hypermutable strains. In contrast, flash-freezing with liquid nitrogen was the best method for the K12 strain (fig. 1*B*).

The Effects of Glycerol and Freezing on Genomic Diversity

Our experimental design is summarized in figure 2. To determine the effect of archiving evolving populations using glycerol stocks, we generated a large, heterogeneous culture with the hypermutable strain, which we termed the "initial" culture (see "Materials and Methods" section). We archived three glycerol stocks samples from the initial culture, using the optimal method for the hypermutable strain (15% glycerol and freezing at −80 °C). We refer to this treatment as the "glycerol + freezing" condition. To distinguish between the effects of freezing and the addition of glycerol, we also added glycerol to three replicate samples from the initial population, but did not freeze them (the "glycerol" condition). We sequenced three independent samples from the "initial" population, and each of the recovered cultures.

To quantify the allele frequency in each population, we first identified sites with sufficiently high sequencing coverage. Specifically, we identified 4,211,108 sites (91% of all sites) in the genome with at least 250-fold coverage for all sequenced samples. For sites with higher coverage, we downsampled coverage by choosing 250 reads at random for each site for further analysis, in order to prevent biasing the diversity measures of samples with different read depths by rare mutations and sequencing errors. Using these data, we computed two measures of genome-wide diversity that depend on the frequency p_i of allele (nucleotide) i at that site, which we computed as the fraction of reads with nucleotide i. The first diversity measure is Shannon's diversity index ($S = -\sum_i p_i \ln(p_i)$) (Shannon 1948) which is maximal when different alleles have the same frequency. The second measure is the frequency of the most abundant allele ($M = \max(p_i)$). The diversity, most abundant allele frequency, and minor allele frequency spectra for our samples are shown in supplementary figure S1*A*–*C*, Supplementary Material online.

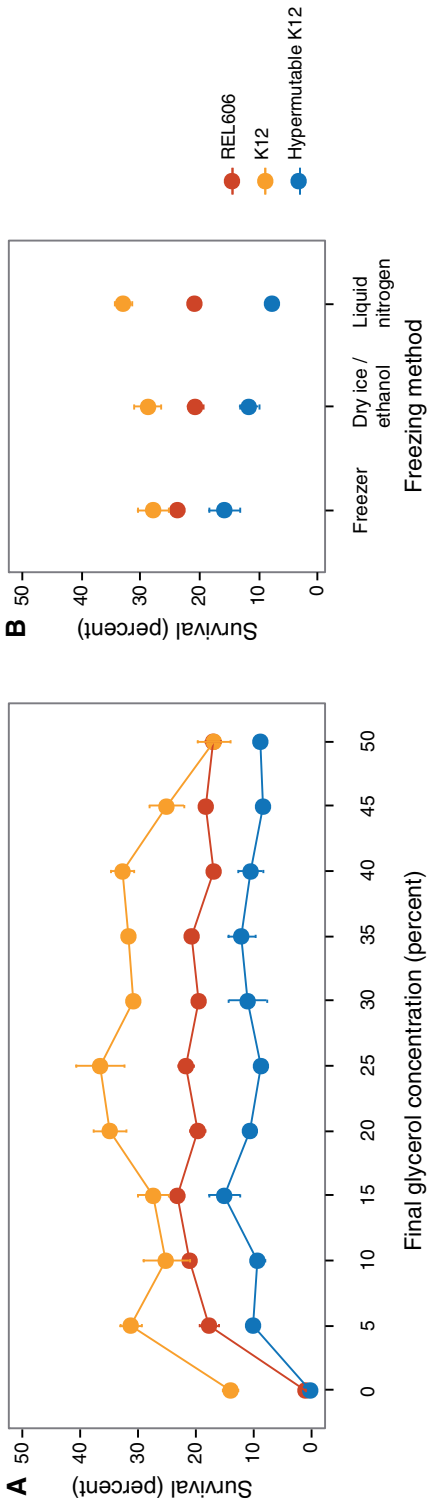

Fig. 1.—Optimal glycerol concentration and freezing method for *E. coli* strains REL606, K12, and the hypermutable K12. (*A*) A higher percentage of cells survived intermediate glycerol concentrations, while (*B*) survival of the different freezing methods was strain-dependent. For all experiments, each condition was replicated 5 times. Lengths of vertical bars correspond to the standard error of the mean (s.e.m.).

The allele frequency at most sites across the genome changed little after treatment with glycerol or glycerol + freezing. Specifically, the frequency of the most abundant allele changed, on average, by 0.089% (S.E.M. = 2.168×10^{-7}), and diversity changed by 0.0083 (S.E.M. = 1.68×10^{-6}). The changes in the spectra were significantly different (supplementary table S1, Supplementary Material online). The small changes in major allele frequency and diversity between the initial and treatment subsamples are likely due to selection, rather than drift (fig. 3, supplementary figure S2, supplementary text, Supplementary Material online). The treatment cultures show more genetic diversity at low-diversity sites than the initial culture (fig. 3, supplementary text, Supplementary Material online). In most cases, changes in allele frequency this small would have little detectable effect on the evolution of a population. However, at some sites, the changes were substantially larger, that is, up to 7 percentage points in the frequency of the most abundant allele, and up to 0.26 in diversity (supplementary fig. S3, Supplementary Material online). Such changes can lead to the irreversible loss of a rare allele. If we define the loss of a rare allele as occurring when the most abundant allele increases in frequency from at least 98.8% to 100% (equivalent to a loss of three reads with the alternative allele out of 250 total reads), and if we require that all three sequenced subsamples from a single recovered culture reflect the same loss, then such a loss occurred in 0.9% percent (3/346) of comparable sites. These three lost alleles occur in the *phoB*, *gsiA*, and *recQ* genes. Thus, even relatively small changes in allele frequency through archiving can lead to the loss of rare alleles.

A worst-case scenario for the effects of archiving is that it could cause a massive loss of genomic diversity through rapid selective sweeps at one or few loci, which have alleles that are strongly advantageous because they confer freezing, thawing, or glycerol tolerance. However, we found no such hard sweeps when we compared the sequence data for initial and treated cultures because no allele substitutions were evident. We thus focused on smaller changes in allele frequency or diversity.

To do so, we selected the sites for the diversity and most abundant allele frequency metrics that varied the most between samples (see "Materials and Methods" section) and used these to conduct a principal components analysis (PCA) (Pearson 1901; Hotelling 1933; Jolliffe 2002), where the allelic diversity at each site constituted a separate dimension. PCA is a multivariate analysis technique that projects high-dimensional data onto lower dimensional subspaces (the principal components) that retain most of the variation in the data. The first principal component maximizes the variation between our sequenced samples, the second principal component maximizes the remaining variation between samples, and so on (Jolliffe 2002). For Shannon's diversity index and the most abundant allele measure, we selected 927 and 305 sites, respectively. Figure 4 shows the first two principal

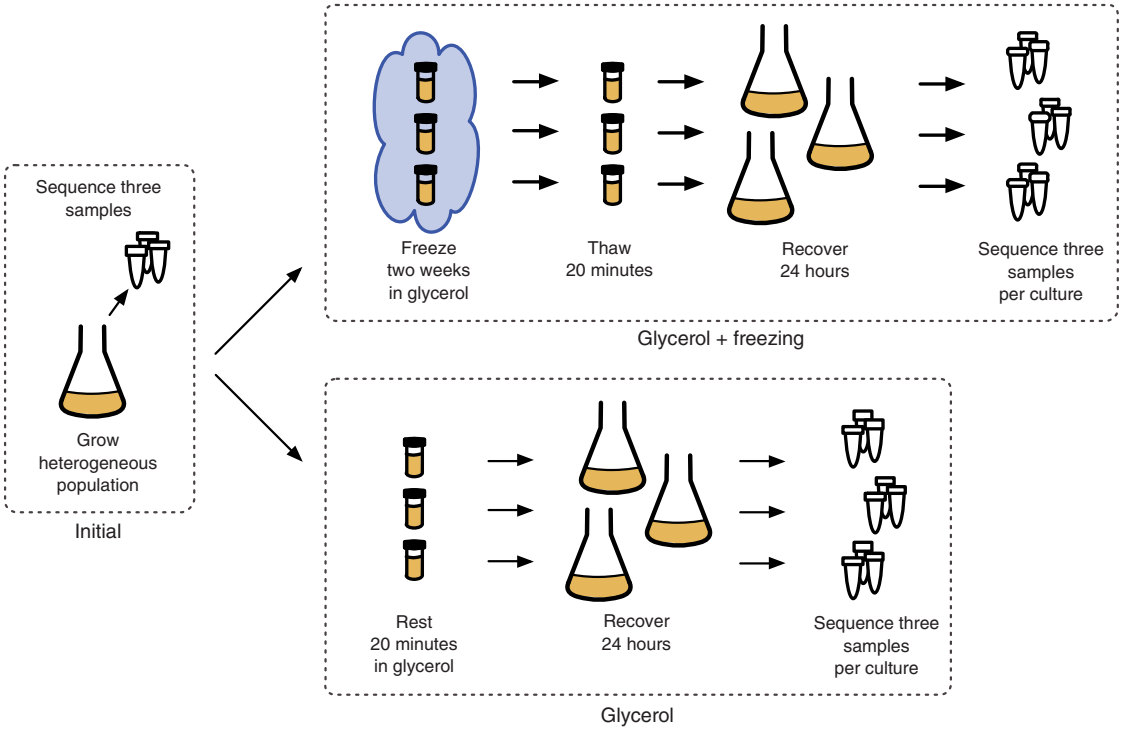

Fig. 2.—Experimental design to test the effect of archiving a heterogeneous population. We made six glycerol stocks from an initial, heterogeneous population. We split these into two treatments: the "glycerol + freezing" treatment, and the "glycerol" treatment. The "glycerol + freezing" stocks were frozen for two weeks, thawed, and recovered. The "glycerol" stocks were incubated at room temperature for 20 min and recovered. We sequenced three samples from the initial culture and each of the recovered cultures.

components, which together account for 18% and 22% of the variation in Shannon's diversity and the most abundant allele frequency, respectively. As is evident from the figure, the initial population (gray) and the treated populations (green and blue) are clearly separated in the PC1–PC2 plane, meaning that the respective populations differ in their diversity at a combination of sites. More than that, the glycerol (green) and freezing + glycerol (blue) samples are also well separated, implying that glycerol treatment alone also affects diversity, and does so differently from freezing (fig. 4).

For both diversity measures, we identified the ten sites with the highest contribution to PC1 or PC2, according to the largest absolute value of loadings (supplementary fig. S4, Supplementary Material online) (Jolliffe 2002). Three of these sites appeared in the top ten sites of both the diversity and most abundant allele PCA (the probability that these sites were important to both analyses by chance alone is $P = 7.68 \times 10^{-15}$). These sites occurred in the genes *tpx*, *yeeR*, and *yejH*. The most abundant allele in them was the reference allele, and the primary alternative allele encoded a single non-synonymous (amino acid) change (Phe89Cys, Leu172Ser, and Val140Ala, respectively). The diversity measures of the initial samples for these three loci were always between that of the glycerol treatment and the glycerol + freezing treatment; for example, the diversity index at *tpx* was 0.69, 0.79, and 0.81 for the glycerol, initial, and glycerol + freezing samples, respectively. Diversity values and most abundant allele frequencies for the remaining most important sites are shown in supplementary figure S5, Supplementary Material online.

The *tpx* gene encodes a thiol peroxidase, which is the main antioxidant under anaerobic conditions (Cha et al. 2004). The *yeeR* gene encodes a predicted inner membrane protein of prophage origin (EcoCyc, Keseler et al. 2012), while the *yejH (radD)* encodes a protein involved in the repair of double strand DNA breaks (Byrne et al. 2014; Chen et al. 2015). *E. coli* strains with mutations in other genes involved in repairing double strand DNA breaks (e.g., *recA, recB, recC*) (Wigley 2012) show reduced survival to freezing and thawing (Calcott and Thomas 1981). To our knowledge, the *tpx*, *yeeR*, and *yejH* genes have not been implicated in tolerance to freezing, thawing, or glycerol before.

Discussion

We showed that different *E. coli* strains survived archiving at drastically different rates. For example, nearly twice as many cells from the K12 MG1655 strain survived freezing as the K12 hypermutable strain, both in 15% glycerol. The reduced viability of the mutator strain is consistent with previous studies showing that a hypermutable strain loses function in tens of

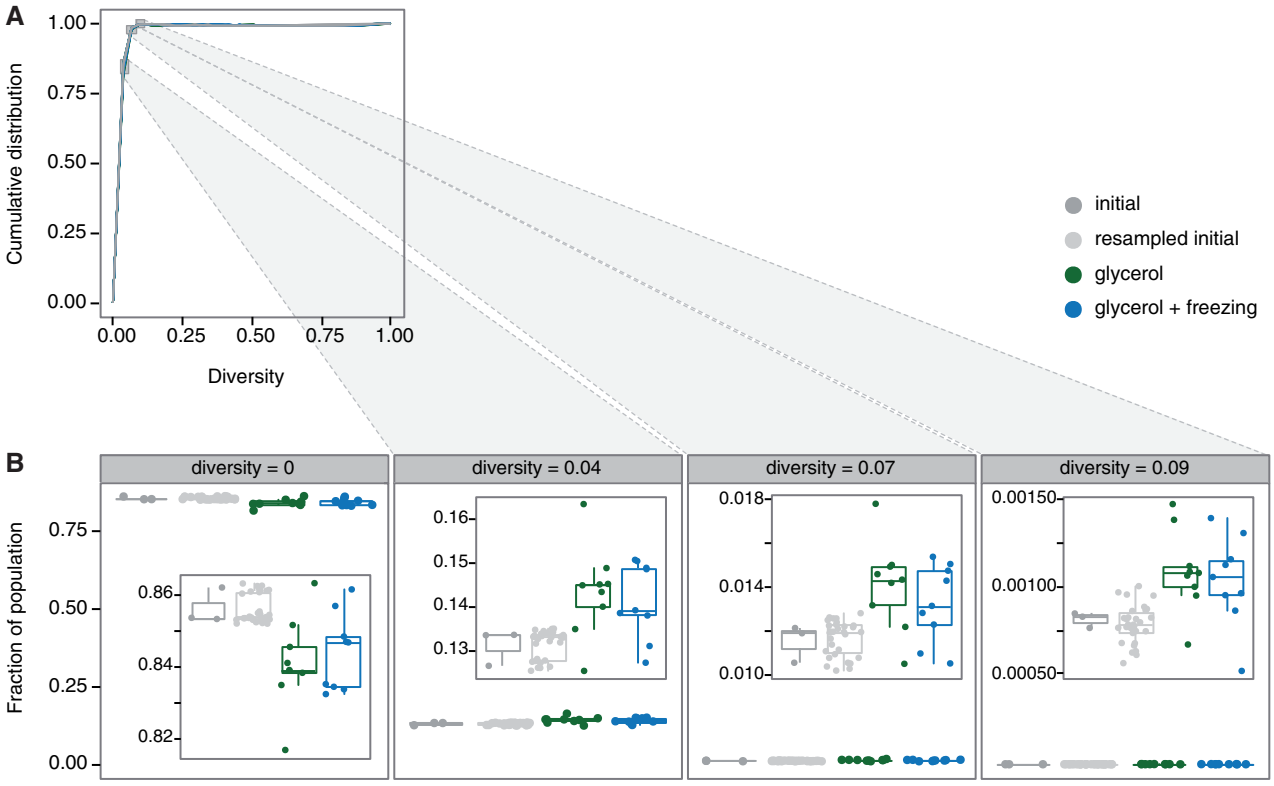

Fig. 3.—Comparison of the experimental spectra of diversity metrics to the spectra of diversity metrics expected under drift alone. We resampled the allele frequency of the initial culture ("initial") 50 times to generate a null distribution that reflects how subsampling alone affects the diversity measure ("resampled initial"). (*A*) Each point in the curve corresponds to the fraction of a sample (*y*-axis) that has at least a given diversity value (*x*-axis). It is difficult at this scale to see whether the initial sample and its resamples (gray and light gray) overlap those of the treatments (green and blue). (*B*) Thus, to further compare the diversity spectra, we plotted the fraction of sites in the genome that have a given diversity metric for each sample. The diversity values ranging from 0 (no diversity at all) to 0.09 are each depicted in the 4-panel plot with shared *y*-axes, and are labeled in the gray bars at the top of each panel. The insets in each panel depict the same information but have a smaller range of *y*-axis values. The initial resamples (light gray) depict the expected fraction of the population having a given diversity value under pure drift. Because the treatment populations (blue and green) differ from the resamples, we can conclude that drift alone is insufficient to generate the diversity spectra. The full methods are described in the supplementary text, Supplementary Material online.

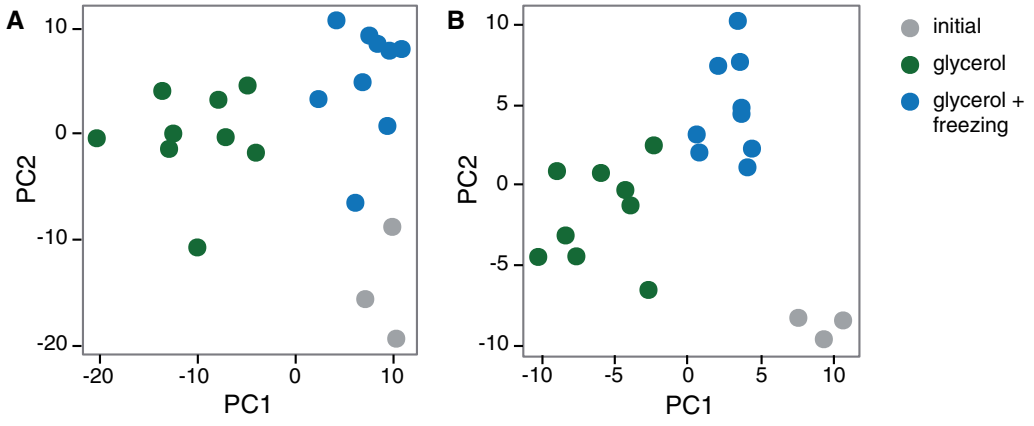

Fig. 4.—Projection of (*A*) Shannon's diversity index and (*B*) the frequency of the most abundant allele onto the first two principal components (PC1, PC2) of a PCA. Note that the experimental treatments (colors) are clustered together, meaning that the respective samples are genetically differentiated.

genes during ~1,600 generations of growth with a very small effective populations size (Funchain et al. 2000). While not much is known about the genes that confer resistance to freezing, the survival differences between the tested *E. coli* strains can likely be explained by differences in their genome sequences, or even gene content. The optimal freezing rate differed between strains; therefore, we recommend testing the optimal freezing strategy when working with new *E. coli* strains.

We showed that archiving affects diversity across many loci, and that this change in diversity is due to selection rather than drift. The treatment cultures showed more genetic diversity than the initial culture at low-diversity sites, and more diversity than expected by sampling from the initial culture. The increased diversity at these sites suggests that one or more low-frequency beneficial alleles (accompanied by their completely-linked genomes) increased in frequency during the glycerol or glycerol + freezing treatment.

We further showed that ~1% of rare alleles that occurred at low frequencies (<3.2%) were lost after treatment with glycerol or glycerol + freezing. The loss can be important during the expansion of a new rare allele in a fairly homogenous background (e.g., within a population that has recently undergone a selective sweep or has a low mutation rate). The loss can also be important during ongoing clonal interference in adapting populations. Rare beneficial alleles that occur in a subclone within the population may also be lost during archiving, thereby altering the evolutionary trajectory of the population.

The implications of our findings for the use of glycerol stocks in laboratory evolution studies are clear. Archived populations can be safely used to conduct phenotypic assays of populations in which the loss of a rare allele is not important, or where the effect of a loss is smaller than that of experimental noise, for example, in typical fitness assays or in many phenotypic assays. However, caution is appropriate when using glycerol stocks to reconstitute an entire population, for example, to restart or replay evolution. Because rare alleles can be lost, and because the loss of any allele is permanent, archiving can alter the future evolutionary trajectory of a population, especially if the lost alleles were strongly beneficial. Caution is also appropriate when interpreting studies that sequence evolving populations restarted from glycerol stock, because the diversity of the population can be affected by archiving.

In practice, experimental studies should be designed to minimize the number of freeze-thaw cycles. Sequencing populations restarted from glycerol stocks can also miss minor alleles that were found in the original population, though the extent of this problem depends on the sequencing depth. One strategy to ensure that sequenced populations reflect the evolving population is to pellet samples for sequencing during the evolution experiment rather than reconstitute populations from glycerol stock.

Materials and Methods

Bacterial Strains

We obtained strains REL606, an *E. coli* B derivative, and K12 MG1655 from Yale University's Coli Genetic Stock Center (CGSG number 12354 and 7740, respectively). An *E. coli* K12 hypermutable strain was provided by the Sniegowski lab. It has defects in mismatch repair (*mutL13*) and DNA Polymerase III proofreading (*dnaQ905*); its mutation rate is ~4,000-fold higher than that of an isogenic wildtype strain (Gentile et al. 2011).

Optimal Archiving Conditions

We tested a range of freezing methods (dry ice/ethanol bath, liquid nitrogen, and simply placing in a −80 °C freezer), as well as glycerol concentrations ranging from 0% to 50% for our three strains (K12, REL606, and the hypermutable strain). For each strain, we started a 5-ml preculture from glycerol stock in LB (Sambrook and Russell 2000) and incubated this culture for 9 h at 37 °C with shaking at 100 rpm (Infors HT Multitron Standard). We then added the 5- ml preculture to 95 ml of LB in 500 ml Erlenmeyer flasks, and incubated overnight at 37 °C with shaking at 100 rpm. We took five independent samples and plated dilutions of each onto LB agar plates for each strain and incubated the plates at 37 °C overnight to obtain the initial cell density of each culture. To identify the optimal glycerol concentration, we first prepared glycerol (Sigma) solutions in water ranging from 0% to 100% glycerol at 10% intervals, and then added 1 ml of culture to 1 ml of glycerol solution for a final glycerol concentration ranging from 0% to 50% at 5% intervals. We froze five replicates in TPP cyrotubes (89020, Switzerland) for each glycerol concentration by placing them in a −80 °C freezer (REVCO UxF70086V). To determine the best freezing method, we froze samples by (1) placing them in the −80 °C freezer, (2) flash-freezing them in a dry-ice/ethanol bath, or (3) flash-freezing them in liquid nitrogen (Sambrook and Russell 2000; Tedeschi and De Paoli 2010). After 1 week at −80 °C, we thawed the glycerol stocks at room temperature for 20 min, and plated a dilution series for each thawed stock on LB agar plates immediately after thawing. We counted colonies on plates containing between 50 and 500 colonies after overnight growth at 37 °C. For each strain, we calculated survival as the ratio $2d_{gly}/d_{init}$, where d_{gly} is the average glycerol stock cell density, d_{init} is the average initial culture cell density, and the factor two accounts for the dilution involved in adding glycerol to the glycerol stocks.

Evolution Experiment

Generating a Large, Heterogeneous Population

In order to test the effect of glycerol and freezing on an evolving population, we aimed to use a large and diverse

population in which multiple alleles segregate at different frequencies. To this end, we used the previously described *E. coli* hypermutable strain. Previously, we had evolved this strain for 2,907 generations [at 37 °C with shaking in 2-ml Davis minimal media (Carlton and Brown 1981) supplemented with 1 g/L of glucose] by diluting a culture 100,000-fold every 24 h into fresh media to obtain almost 17 generations of growth each day. We had archived the population at generation 2907 by adding 400 μl of 50% glycerol to 800 μl of culture, for a final glycerol concentration of 17%. For our current experiment, we used 10 μl of this glycerol stock to start a 10-ml culture in LB. We evolved this culture for another 207 generations at 37 °C with shaking at 100 rpm by diluting the culture 100-fold into fresh media every 24 h for 31 days. At the end of this period, in order to have enough cells for the experiments, we expanded the population from 10 to 40 ml during an additional 6 h (two generations) of growth.

Archiving the Heterogeneous Population

We collected cell pellets from ten independent 1 ml samples of our final 40 ml culture, and made 12 glycerol stocks from this culture. For these glycerol stocks, we added 333 μl of 60% glycerol to 1 ml of culture to arrive at a final glycerol concentration of 15%. We archived six of the stocks by freezing them at −80 °C for 2 weeks after which we allowed them to thaw for 20 min at room temperature. We exposed the other six stocks to room temperature for 20 min. We added each room temperature stock to 8.667 ml of LB, and incubated it at 37 °C with shaking for exactly 24 h. For each recovered culture, we obtained nine samples of 1 ml each and collected cell pellets from these samples.

Sequencing

We sequenced three independent cell pellets from the initial culture, from three of the recovered non-frozen glycerol cultures, and from three of the recovered frozen cultures (fig. 2). We isolated genomic DNA using Qiagen's DNeasy Blood and Tissue Kit (cat. no. 69582) from the cell pellets. We followed the manufacturer's instructions, except we eluted the DNA by incubating the spin column's filter at room temperature for 2 min with 200 μl of Qiagen's Buffer AE diluted 10-fold with water preheated to 56 °C to increase the yield.

We used Illumina's TruSeq DNA PCR-Free kit (cat. no. FC-121-3002) to prepare and barcode libraries for paired-end sequencing, following the manufacturer's instructions for obtaining 350 bp fragments. Importantly, we used no PCR steps in the preparation of the libraries. We used qPCR with Roche's FastStart Essential DNA Green Master kit (cat. no. 06402712001) to quantify the libraries and they were mixed in equimolar amounts. Paired-end, 125 bp sequencing was completed on an Illumina HiSeq 2500 v2. The data from this project has been deposited at the NCBI Sequence Read Archive under the accession SRP063330.

We used bowtie2 v2.1.0 (Langmead and Salzberg 2012) to align the Illumina reads to the *E. coli* K12 genome (Blattner et al. 1997). We downloaded the complete reference genome and its annotated genes from the National Center for Biotechnology Information (NCBI; reference sequence NC_000913.3) (Pruitt et al. 2002; Hayashi et al. 2006; Riley et al. 2006;Tatusova et al. 2013) on January 6, 2015. We kept only those reads where both paired-end reads properly aligned. We used Samtools v0.1.19 (Li et al. 2009) to sort the bam files, Picard v1.127 (http://broadinstitute.github.io/picard, last accessed March 21, 2016) to assign the readgroup information, and GATK v.3.3.0 (DePristo et al. 2011) to realign reads locally around indels. We obtained the allele frequency at every site in the genome for each sample using bam readcount 0.7.4 (https://github.com/genome/bam-readcount, last accessed March 21, 2016) retaining only alleles with mapping and base quality scores of at least 20 for further analysis. We truncated the coverage at each site to 250 reads by selecting 250 reads per site at random, and included just those sites with at least 250-fold coverage for all samples in all further analyses. We considered a given allele to be substituted when it reached at least 90% of the reads at a site after treatment. The gene or intergenic region, and the putative effect, of each mutation was annotated using SnpEff v.4.1, build 2015-01-07 (Cingolani et al. 2012) and the previously mentioned gene annotations from NCBI.

Data Analysis

For all sites in the genome with 250-fold coverage, we computed two measures of diversity that depend on the frequency p_i of allele i at that site (determined as the fraction of reads with allele i), namely Shannon's diversity index ($S = -\sum_i p_i \ln(p_i)$), and the frequency of the most abundant allele ($M = \max(p_i)$). For both measures, we selected the sites to use for the PCA (Pearson 1901; Hotelling 1933; Jolliffe 2002) by choosing the sites that vary the most between samples according to a measure of the total variation between samples. We used $\sum_{n<m}|S_n - S_m|$, where is S_n Shannon's diversity index for sample n at a given site, or $\sum_{n<m}|M_n - M_n|$, where M_m is the frequency of the most abundant allele at a given site to quantify the total variation between samples for each site. We then selected the sites within the top 0.0001 percentile for the PCA. We performed the PCA in R v.3.1.1 using prcomp, which used singular value decomposition for the PCA (R Development Core Team 2015). We centered our diversity measures at each site on zero and scaled them to have unit variance, thereby preventing the PCA from returning principal components biased by the inherent variability of the allele diversity measures across sites (Jolliffe 2002).

We defined an allele to be lost when (1) the most abundant allele increased in frequency from 247/250 to 250/250 reads (98.8–100%), and (2) this frequency increase occurred in all three sequenced samples from a recovered culture. We

considered sites for this analysis that ranged from an initial major allele abundance of 242/250 to 247/250 (96.8–98.8%).

Supplementary Material

Supplementary figures S1–S5 and table S1 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Jelena Kühn-Georgijevic for laboratory assistance, and Sinisa Bratulic, Heidi Lischer, and Macarena Toll-Riera for advice. K.S. and J.A. acknowledge support through the Forschungskredit program of the University of Zurich, grants K-74301-03-0 and FK-14-076. A.W. acknowledges support through Swiss National Science Foundation grant 31003A_146137, as well as through the University Priority Research Program in Evolutionary Biology at the University of Zurich.

Literature Cited

Barrick JE, et al. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243–1247.
Barrick JE, Lenski RE. 2009. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. Cold Spring Harb. Symp. Quant. Biol. 74:119–129.
Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution. Nat Rev Genet. 14:827–839.
Behrensmeyer AK, Kidwell SM, Gastaldo RA. 2000. Taphonomy and paleobiology. Paleobiology 26:103–147.
Blattner FR, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1462.
Byrne RT, Chen SH, Wood EA, Cabot EL, Cox MM. 2014. *Escherichia coli* genes and pathways involved in surviving extreme exposure to ionizing radiation. J Bacteriol. 196:3534–3545.
Calcott PH. 1986. Cryopreservation of microorganisms. Crit Rev Biotechnol. 4:279–297.
Calcott PH, Thomas M. 1981. Sensitivity of DNA repair deficient mutants of *Escherichia coli* to freezing and thawing. FEMS Microbiol Lett. 12:117–120.
Carlton BC, Brown BJ. 1981. Gene Mutation, in Manual of Methods for General Bacteriology (ed. Gerhardt, P). American Society for Microbiology, pp 222–142.
Cha MK, Kim WC, Lim CJ, Kim K, Kim IH. 2004. *Escherichia coli* periplasmic thiol peroxidase acts as lipid hydroperoxide peroxidase and the principal antioxidative function during anaerobic growth. J Biol Chem. 279:8769–8778.
Chen SH, Byrne RT, Wood EA, Cox MM. 2015. *Escherichia coli radD (yejH)* gene: a novel function involved in radiation resistance and double-strand break repair. Mol Microbiol. 95:754–768.
Cingolani P, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6:80–92.
DePristo MA, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43:491–498.
Dumont F, Marechal PA, Gervais P. 2004. Cell size and water permeability as determining factors for cell viability after freezing at different cooling rates. Appl Environ Microbiol. 70:268–272.
Efremov IA. 1940. Taphonomy: new branch of paleontology. Pan-Am Geol. 74:81–93.
Elena SF, Lenski RE. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet. 4:457–469.
Fox JW, Lenski RE. 2015. From here to eternity—the theory and practice of a really long experiment. PLoS Biol. 13:e1002185.
Funchain P, et al. 2000. The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. Genetics 154:959–970.
Gao D, Critser Jk. 2000. Mechanisms of cryoinjury in living cells. ILAR J. 41:187–196.
Garland T, Rose MR. 2009. Experimental evolution: concepts, methods, and applications of selection experiments. Berkeley (MA): University of California Press.
Gentile CF, Yu S-C, Serrano SA, Gerrish PJ, Sniegowski PD. 2011. Competition between high- and higher-mutating strains of *Escherichia coli*. Biol Lett. 7:422–424.
Hayashi K, et al. 2006. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol Syst Biol. 2:2006.0007.
Herron MD, Doebeli M. 2013. Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. PLoS Biol. 11:e1001490.
Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 24:417–441.
Jolliffe IT. 2002. Principal component analysis, 2nd ed. New York: Springer.
Kassen R. 2014. Experimental evolution and the nature of biodiversity. Greenwood Village: Roberts and Company Publishers.
Kawecki TJ, et al. 2012. Experimental evolution. Trends Ecol Evol. 27:547–560.
Keith SC. 1913. Factors influencing the survival of bacteria at temperatures in the vicinity of the freezing point of water. Science 37:877–879.
Keseler IM, et al. 2012. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res. 41:D605–D612.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359.
Lenski RE, Rose M, Simpson S, Tadler S. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am Nat. 138:1315.
Levy SF, et al. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature 519:181–186.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.
Lovelock JE. 1953. The mechanism of the protective action of glycerol against haemolysis by freezing and thawing. Biochimica et biophysica acta 11:28–36.
Mazur P. 1984. Freezing of living cells: mechanisms and implications. Am. J Physiol. 247:C125–C142.
Pearson K. 1901. On lines and planes of closest fit to systems of points in space. Philos Mag. 2:559–572.
Pruitt K, Brown G, Tatusova T, Maglott D. 2002. The Reference Sequence (RefSeq) database. The NCBI Handbook. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information: pp. 1–24.
Riley M, et al. 2006. *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. Nucleic Acids Res. 34:1–9.
R Development Core Team. 2015. R: A language and environment for statistical computing. Available from: http://www.R-project.org.
Sambrook J, Russell D. 2000. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory.
Shannon CE. 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27:379–423.
Sharp RJ. 1983. The preservation of genetically unstable microorganisms and the cryopreservation of fermentation seed cultures. Adv Biotechnol Processes. 3:81–109.

Sleight SC Lenski RE. 2007 Evolutionary adaptation to freeze-thaw-growth cycles in *Escherichia coli*. Physiol Biochem Zool. 80:370–385.
Sleight SC, Orlic C, Schneider D, Lenski RE. 2008. Genetic basis of evolutionary adaptation by *Escherichia coli* to stressful cycles of freezing, thawing and growth. Genetics 180:431–443.
Sleight SC, Wigginton NS, Lenski RE. 2006. Increased susceptibility to repeated freeze-thaw cycles in *Escherichia coli* following long-term evolution in a benign environment. BMC Evol Biol. 6:104.
Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. 2013. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res. 42:D553–D559.
Tedeschi R, De Paoli P. 2010. Collection and preservation of frozen microorganisms. Methods Mol Biol. 675: 313–326.
Turner P, Souza V, Lenski RE. 1996. Tests of ecological mechanisms promoting the stable coexistence of two bacterial genotypes. Ecology 77:2119–2129.
Wielgoss S, et al. 2012. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load Proc Natl Acad Sci U S A. 110:222–227.
Wigley DB. 2012. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. Nat Rev Microbiol. 11:9–13.

Associate editor: Charles Baer