

Simple Phenotypic Sweeps Hide Complex Genetic Changes in Populations

Ram P. Maharjan^{1,†}, Bin Liu^{2,3,†}, Lu Feng^{2,3}, Thomas Ferenci¹, and Lei Wang^{2,3,4,*}

¹School of Molecular Bioscience, University of Sydney, New South Wales, Australia

²TEDA Institute of Biological Sciences and Biotechnology, Nankai University, Tianjin, People's Republic of China

³Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, Tianjin, People's Republic of China

⁴State Key Laboratory of Medicinal Chemical Biology, Nankai University, Tianjin, People's Republic of China

*Corresponding author: E-mail: wanglei@nankai.edu.cn.

†These authors contributed equally to this work.

Accepted: January 6, 2015

Abstract

Changes in allele frequencies and the fixation of beneficial mutations are central to evolution. The precise relationship between mutational and phenotypic sweeps is poorly described however, especially when multiple alleles are involved. Here, we investigate these relationships in a bacterial population over 60 days in a glucose-limited chemostat in a large population. High coverage metagenomic analysis revealed a disconnection between smooth phenotypic sweeps and the complexity of genetic changes in the population. Phenotypic adaptation was due to convergent evolution and involved soft sweeps by 7–26 highly represented alleles of several genes in different combinations. Allele combinations spread from undetectably low baselines, indicating that minor subpopulations provide the basis of most innovations. A hard sweep was also observed, involving a single combination of *rpoS*, *mgID*, *malE*, *sdhC*, and *malT* mutations sweeping to greater than 95% of the population. Other mutant genes persisted but at lower abundance, including *hfq*, consistent with its demonstrated frequency-dependent fitness under glucose limitation. Other persistent, newly identified low-frequency mutations were in the *aceF*, *galF*, *ribD* and *asm* genes, in noncoding regulatory regions, three large indels and a tandem duplication; these were less affected by fluctuations involving more dominant mutations indicating separate evolutionary paths. Our results indicate a dynamic subpopulation structure with a minimum of 42 detectable mutations maintained over 60 days. We also conclude that the massive population-level mutation supply in combination with clonal interference leads to the soft sweeps observed, but not to the exclusion of an occasional hard sweep.

Key words: experimental evolution, mutational sweeps, metagenomics, *E. coli* genome, periodic selection.

Introduction

Once upon a time, the concept of mutational sweeps was relatively simple. It was assumed that individual beneficial mutations sweep to fixation in bacterial populations before another mutation arises in the fitter mutant and in turn increases in frequency (Atwood et al. 1951a). This periodic selection notion tied in nicely with ecological principles that clones excluded others from a single niche (Armstrong and McGehee 1980), so the supposed mutational successions reflected this search for the fittest in one environment. This mode of adaptation also depends on the selective sweep originating from a single new mutation, this scenario being called a “hard” sweep (Hermisson and Pennings 2005). A further consequence of such hard sweeps is that ancestral variation that

is unlinked to the selected allele will be eliminated. An example of this early view, relevant to the experimental system we use, is the statement that “evolution in continuous cultures usually consists of a succession of dominant strains of increasing growth rate” (Kubitschek 1970; Dykhuizen and Hartl 1983).

Views on mutational adaptation in large populations have shifted in recent years and suggest that adaptation is more likely to produce “soft” selective sweeps, where multiple adaptive alleles sweep through the population at the same time, because the alleles either were already present as standing genetic variation or arose independently by recurrent de novo mutations (Pennings and

Hermisson 2006; Messer and Petrov 2013). A source of standing genetic variation comes from the concept of clonal interference (Gerrish and Lenski 1998; de Visser and Rozen 2006), which reduces the elimination of competing mutations in large populations. A combination of clonal interference and soft sweeps has been used to explain events in experiments with replicate yeast populations (Lang et al. 2013) and several other independent studies are inconsistent with successions and hard sweeps.

For example, a pure culture of a bacterium in a constant, stirred, single-resource continuous culture environment rapidly diversified (Maharjan et al. 2006), contrary to historical assumptions about successions and niche exclusion in a highly controlled environment. Diverse lineages were the reproducible outcome of *Escherichia coli* growth in glucose- or phosphate-limited continuous cultures (Kinnersley et al. 2009; Wang et al. 2010; Maharjan et al. 2012) as are glucose-limited yeast populations (Gresham et al. 2008). The divergence in single-resource chemostats is apparent in 60–100 generations and populations eventually form consortia after longer-term maintenance in a glucose-limited chemostat (Rosenzweig and Adams 1994). This result is inconsistent with hard sweeps being the sole recurring mode of adaptation in large bacterial populations. Rapid clonal divergence is evident also in nature, in clonal infections by *E. coli* (Levert et al. 2010) as well as in the diversity of phenotypes generated by the same mechanisms as are operating in chemostats (Maharjan, Nilsson, et al. 2013; Phan and Ferenci 2013).

Another set of observations inconsistent with hard sweeps in a chemostat population came from genome sequencing of individual highly represented clones (Kinnersley et al. 2009; Maharjan et al. 2010, 2012; Gaffe et al. 2011). By identifying beneficial mutations in different isolates, it was clear that multiple mutational lineages arose in chemostats directly from ancestor, so multiple, concurrent changes in allele frequency are more likely to be the norm. Equally, adaptation did not start from a single new mutation as is expected from a hard sweep; these findings are also more consistent with soft sweeps where more than a single copy of the allele contributes to an adaptive substitution (Hermisson and Pennings 2005; Pennings and Hermisson 2006). Once distinct mutational lineages are present in a population, complex trade-offs, epistasis, and frequency-dependent interactions provide the basis of coexistence of bacteria in the same environment (Maharjan and Ferenci 2013; Maharjan, McKenzie, et al. 2013).

Evidence supporting soft sweeps has recently emerged from metagenomic studies using experimental populations of various microbes (Herron and Doebeli 2013; Kvitek and Sherlock 2013; Lang et al. 2013). These are also consistent with independent lineages emerging with yeast and *E. coli*. From the above accumulating evidence, it would seem that a core feature of evolution, sweeps by fitter types within populations, is more complex than simply sweeps to complete

fixation (Barrick and Lenski 2013). Here, we study the detailed kinetics of allele frequency changes in an evolving *E. coli* chemostat population to allow a detailed understanding of events such as the extent of fixation of mutations by soft and hard sweeps and the time-scale of concurrent sweeps in fixed environments. The nature of classical, so-called periodic selection events (Atwood et al. 1951b) causing fluctuations in the proportion of neutral alleles in populations (Koch 1974; Berg 1995; Notley-McRobb and Ferenci 2000) needs to be reconciled with what is really happening in large populations. We here attempt to define such events using metagenomic analysis.

A further aim here is a comprehensive description of the mutational changes under glucose limitation because previous mutation analysis was biased by sequencing only high-abundance members of chemostat populations. In a chemostat system we use, a large population of 10^{10} bacteria is under constant strong selection for the improved utilization of the limiting nutrient (Ferenci 2008a). Three mutations have been previously identified in replicate populations and increased in representation to greater than 50% of the population by about 100 generations (Notley-McRobb et al. 2003; Ferenci 2008b). Mutations in these *rpoS*, *mgID/O*, and *malT* genes were of considerable competitive benefit under glucose limitation (Notley-McRobb and Ferenci 1999a, 1999b). More recently, *hfq* mutations have also been found in most but not all populations (Maharjan et al. 2010); these, in contrast to the three above, are present in less than 50% of population members and show negative frequency dependence in fitness assays (Maharjan, McKenzie, et al. 2013). Epistasis prevents the occurrence of *hfq* and *rpoS* mutations in the same background so the *rpoS* and *hfq* lineages persist in parallel (Maharjan and Ferenci 2013).

To avoid biases against low-abundance types and to overcome this limitation, we here used a high-coverage metagenomic analysis of a glucose-limited chemostat population over 60 days, which allows us to reveal the sweeps and persisting low-level subpopulations as well as to identify novel genes mutated in chemostats. In this study, we follow the temporal distribution of all mutations represented at and above 1% frequency at multiple time-points and analyze the genetic basis of major changes in population composition and the alleles responsible.

Advances in genomics now allow very high coverage analysis of sequences in a population. Although metagenomics is mostly applied to environmental samples rather than pure cultures, it has the potential to reveal all of the changes in populations starting from a pure clone (Barrick and Lenski 2009; Herron and Doebeli 2013; Kvitek and Sherlock 2013; Lang et al. 2013). Here, we apply metagenomics to 11 samples over 60 days of adaptation to obtain a high-resolution analysis of events in a bacterial culture evolving in a homogeneous environment. The very high sequence coverage in our analysis has now permitted the identification of many new

mutations and the observation that many mutations persist over time but in less than 5% abundance in the population. These low-proportion mutations are not directly impacted by the major sweeps so provide new evidence for coevolving subpopulations in a homogeneous environment. Subpopulations also, as shown here, provide the recurring source of innovation in a large bacterial population.

Materials and Methods

Bacterial Strain and Growth Conditions

For the experimental evolution, the *E. coli* K-12 strain BW2952 (Ferenci et al. 2009), an MC4100 derivative, was propagated at the growth (dilution) rate of 0.1 h^{-1} in a glucose-limited chemostat exactly as described in Notley-McRobb et al. (2003).

Phenotypic Changes in the Population

One phenotypic change in the evolving chemostat population was detected by using the iodine staining method as described previously in Notley and Ferenci (1995). Staining darkness is indicative of the level of expression of the general stress response in *E. coli* (Maharjan, Nilsson, et al. 2013). Briefly, 100- μl aliquots of 10^5 -fold-diluted chemostat cultures were plated on L-broth agar plates in duplicate and incubated for overnight at 37°C . Plates containing 250–300 colonies were then transferred into a fridge (4°C) and kept for 24 h before flooding with iodine solution. Colonies were then scored depending on their staining. Colonies with an *rpoS* mutation (no stress response) stay yellow in color, whereas colonies with high levels of *rpoS* stain dark (Maharjan, Nilsson, et al. 2013).

A second phenotype commonly changed in chemostats is the level of *mal* expression (Notley-McRobb and Ferenci 1999b). To measure this phenotype, we assayed the *malG-lacZ* reporter fusion β -galactosidase activity incorporated into the ancestral BW2952 using Miller Units as described in Ferenci et al. (2009). Bacteria from glycerol cultures stored from each time-point were transferred into 5 ml Luria broth (LB) and grown overnight to stationary phase. β -Galactosidase activities of cultures were then measured as described previously in Notley and Ferenci (1995).

Mutator Phenotype Assay

For qualitative mutator phenotype assay, a single colony of isolates to be tested was inoculated in 5 ml LB. After growing to stationary phase at 37°C , 100 μl of culture directly spread onto LB plates containing rifampicin at 100 $\mu\text{l}/\text{ml}$. For total viable count, 100 μl of appropriately diluted cultures in LB were plated in LB-agar plates. The plates were then incubated overnight at 37°C . The number of rifampicin-resistant colonies in each plate was counted. Isolates producing resistant colonies more than 10-fold in number relative to the ancestor were considered as mutator.

DNA Preparation for Population Level Whole-Genome Sequencing

For chromosomal DNA samples, 40 ml cultures from different time points were collected through the waste outlet of the evolving chemostat culture and harvested by centrifugation at $3,000 \times g$ for 5 min at 4°C . Cell-pellets were resuspended in 750 μl of TE buffer (10 mM Tris and 1 mM ethylenediaminetetraacetic acid, pH = 8.0) and treated with 50 μl of lysozyme (20 mg/ml) for 15 min at 37°C . An amount of 15 μl proteinase K (20 mg/ml) and 25 μl sodium dodecyl sulfate (20% wt/vol) was added into lysozyme-treated cell suspensions and incubated for further 1–3 h or until suspensions became clear. The DNA was then extracted using phenol–chloroform method as previously described (Ferenci et al. 2009). All chemicals used for DNA preparation were purchased from Sigma-Aldrich, Australia.

Genomics Details

The genome resequencing workflow started with sample preparation, which generated sequencer-ready DNA fragments of paired-end libraries. Solexa Genome Analyzer IIx (Illumina, Little Chesterford, Essex, UK) was used to sequence each sample aiming for 1,000-fold coverage.

The 101-bp Solexa paired-end FASTQ reads generated, after being trimmed by SolexaQA (v.1.12) (Cox et al. 2010), contiguous segments longer than 25 bp in which the probability value of each base greater than 0.05. Then, polymerase chain reaction (PCR) duplicates were filtered out by removing pairs with the same first 20 bases in both reads. All nonredundant read pairs were aligned to the *E. coli* BW2952 genome using BWA (Li and Durbin 2009) with the default parameter, which generated the sequence alignment map (SAM) file.

Freebayes (Garrison and Marth 2012) with `-pooled-discrete` enabled and SAMtools with default parameters (Li et al. 2009) were used to detect the SNPs (single-nucleotide polymorphisms), small indels (insertions and deletions), and small complex events (composite insertion and substitution events) for each position and the total read depth at the locus. Small genetic variants, including SNPs, single base pair or small indels (S-indels) and small complex events in each sequenced sample were called at positions covered by sequences of both the forward strand and the reverse stand, and by at least three reads supporting the same mutant allele. Allelic frequencies of each mutation were then estimated from the number of reads and the total number of coverage.

The method for detecting SNPs was verified by analyzing the resequencing data (100-fold coverage) for *E. coli* BW2952 and MG1655, which was generated by the same sequencing method as that for other samples in this study. No alternate allele due to sequencing error was found anywhere in these genomes.

Large structural variant events, such as large deletions, medium-sized insertions, inversions, and tandem duplications,

were detected from the alignment using Pindel (v0.2.4) (Ye et al. 2009), with a pattern growth algorithm to identify the breakpoints of these variants from the paired-end short reads.

Results

Genomic Changes in an Evolving Chemostat Population

In order to understand how phenotypic and genotypic sweeps emerge in evolving asexual populations, we performed genomic analysis of a whole population initially propagated from a single ancestral *E. coli*. This strain was the same as that used in our previous lab evolution studies (Notley-McRobb and Ferenci 1999a, 1999b; Ferenci et al. 2009). The glucose-limited population of approximately 1.6×10^{10} was studied over 60 days and samples cultures stored at -80°C . Whole-population DNA samples were however extracted from bacteria taken directly from the chemostat during the 60 days so that anomalies caused by freezing and reviving cultures did not change the genome composition.

The 11 samples obtained over 60 days were subjected to whole-population genomic sequencing as described in the Materials and Methods section. The average coverage of libraries in sequencing was 800-fold. After the informatics filtering for PCR artifacts and errors in sequencing, we tested the possible occurrence of false positives by subjecting genomes from two *E. coli* pure cultures (of the ancestor BW2952 and the reference strain MG1655) to the same sequencing and filtering methods. No mutations relative to published corrected genome sequences were observed. This provided confidence that the mutations identified in the population samples were not sequencing artifacts.

In total, we detected 2,729 changes from the ancestor (2,664 SNPs, 45 single base-pair indels, 4 large indels, and 1 duplication) in the 11 samples. All mutations, as well as their allelic frequencies at each time point, are listed in [supplementary table S1, Supplementary Material](#) online. The large majority of mutations (2,366 or 87%) were present only once (fig. 1A), of which 1,974 mutations were present in 1–2% of the population. This represents more than 10^8 bacteria of 1.6×10^{10} cells in the evolving population but did not become fixed in the population. Only 343 (13%) of the total 2,729 mutations were detected in multiple time samples, of which 32 (11%) mutations peaked to constitute over 10% of the population. The most abundant in frequency in the population were the recurrent beneficial mutations in particular genes with multiple alleles (fig. 1B). In contrast, singleton mutations were generally of lower abundance but large in number.

Of the 2,664 SNPs, 2,415 are located in coding regions of which 1,696 (64%) and 719 (27%) are nonsynonymous and synonymous SNPs, respectively. The remaining 249 (9%) SNPs are located in noncoding regions. All genes with nonsynonymous mutations that repeatedly occurred at more than three

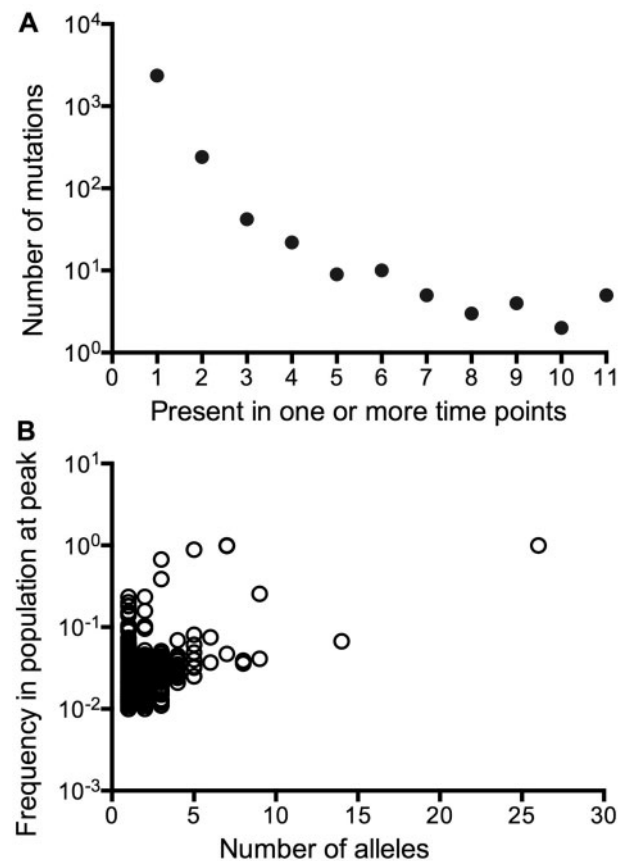


Fig. 1.—Distribution of mutations detected by whole-genome sequencing of an evolving population. (A) The number of occurrences of mutations within 11 samples. The histogram was based on a total of 2,729 mutations (1,182 SNPs, 74 small indels, 4 large indels, and 1 duplication mutation) across 11 samples collected over 60 days of continuous culture. (B) The relationship between the number of alleles found in mutated genes and their frequency at peak in the population. The histogram was based on 2,439 mutations in 1,435 genes with one or more mutations.

time points are listed in [table 1](#). These are the recurrent mutations often with multiple alleles in [figure 1B](#). The recurrence of the mutations in [table 1](#), even at low abundance, suggests that these are likely to be real population components. So below, we discuss the occurrence of mutations established enough to be repeatedly found in greater than three time points at or above 1% abundance.

Dynamics of Phenotypic and Genotypic Adaptation

We undertook phenotypic screening of the whole population for commonly acquired adaptations in our chemostat system. Phenotypes caused by mutations in *rpoS* (altered glycogen staining) and *malT* (altered *mal* gene expression) showed that the sweeps affecting *rpoS* and *mal* gene expression started before 7 and after 14 days, respectively (fig. 2).

Table 1Nonsynonymous Changes Present at Three or More Time-Points in Nonrepeat Regions of the Genome^a

Gene Name	Function	Occurrence in How Many Samples	Number of Alleles	Frequency in Population at Peak
<i>rpoS</i>	RNA polymerase, sigma 5 (sigma 38) factor	11	7	0.99
<i>malT</i>	Fused conserved protein: DNA-binding transcriptional activator/maltotriose-ATP-binding protein	10	26	0.99
<i>mgID/O (galS/O)</i>	DNA-binding transcriptional repressor	11	7	0.98
<i>sdhC</i>	Succinate dehydrogenase, membrane subunit, binds cytochrome b556	5	1	0.95
<i>malE</i>	Maltose transporter subunit	6	5	0.88
<i>mutL</i>	Methyl-directed mismatch repair protein	9	2	0.67
<i>tpiA</i>	Triosephosphate isomerase	6	1	0.62
<i>yhaM</i>	Hypothetical protein	5	2	0.58
<i>pnp</i>	Polynucleotide phosphorylase/polyadenylase	3	3	0.39
<i>asmA</i>	Putative assembly protein	7	9	0.26
<i>malS</i>	Alpha-amylase	4	2	0.16
<i>hfq</i>	HF-I, host factor for RNA phage Q beta replication	5	1	0.15
<i>rcsD</i>	Phosphotransfer intermediate protein in two-component regulatory system with RcsBC	4	1	0.14
<i>b3598/mtlA</i>	Predicted inner membrane protein	5	1	0.09
<i>murP</i>	Fused predicted PTS enzymes: IIB component/IIC component	3	5	0.08
<i>recJ</i>	ssDNA exonuclease, 5' → 3'-specific	6	6	0.08
<i>cusA</i>	Copper/silver efflux system, membrane component	3	4	0.07
<i>rhsD</i>	rhsD element protein	6	14	0.07
<i>ebgA</i>	Cryptic beta-D-galactosidase, alpha subunit	6	5	0.06
<i>eco</i>	Ecotin, a serine protease inhibitor	3	1	0.06
<i>amiD</i>	Putative amidase and lipoprotein	3	2	0.05
<i>mraY</i>	Phospho-N-acetylmuramoyl-pentapeptide transferase	3	3	0.05
<i>mrcA</i>	Fused penicillin-binding protein 1a: murein transglycosylase/murein transpeptidase	4	5	0.05
<i>glnG</i>	Fused DNA-binding response regulator in two-component regulatory system with GlnL: response regulator/sigma54 interaction protein	3	1	0.05
<i>ytfN</i>	Hypothetical protein	5	7	0.05
<i>yfhA</i>	Putative DNA-binding response regulator in two-component system	6	2	0.05
<i>fdnG</i>	Formate dehydrogenase-N, alpha subunit, nitrate-inducible	3	4	0.04
<i>gltB</i>	Glutamate synthase, large subunit	5	9	0.04
<i>sgbE</i>	L-ribulose-5-phosphate 4-epimerase	4	5	0.04
<i>methH</i>	Homocysteine-N5-methyltetrahydrofolate transmethylase, B12-dependent	4	8	0.04
<i>polB</i>	DNA polymerase II	4	3	0.04
<i>cls</i>	Cardiolipin synthase 1	3	4	0.04
<i>ybhF</i>	Fused predicted transporter subunits of ABC superfamily: ATP-binding components	3	2	0.04
<i>purL</i>	Phosphoribosylformyl-glycineamide synthetase	5	8	0.04
<i>thrA</i>	Fused aspartokinase I and homoserine dehydrogenase I	4	6	0.04
<i>frdA</i>	Fumarate reductase (anaerobic) catalytic and NAD/flavoprotein subunit	3	1	0.04
<i>recG</i>	ATP-dependent DNA helicase	6	8	0.04
<i>trxB</i>	Thioredoxin reductase, FAD/NAD(P)-binding	3	4	0.04
<i>malZ</i>	Maltodextrin glucosidase	4	4	0.03
<i>mdtC</i>	Multidrug efflux system, subunit C	3	5	0.03
<i>flu</i>	CP4-44 prophage; antigen 43 (Ag43) phase-variable biofilm formation autotransporter	3	4	0.03

(continued)

Table 1 Continued

Gene Name	Function	Occurrence in How Many Samples	Number of Alleles	Frequency in Population at Peak
<i>treB</i>	Fused trehalose(maltose)-specific PTS enzyme: IIB component/IIC component	3	2	0.03
<i>entF</i>	Enterobactin synthase multienzyme complex component, ATP-dependent	3	5	0.03
<i>fadD</i>	Acyl-CoA synthetase (long-chain-fatty-acid-CoA ligase)	3	4	0.03
<i>entS</i>	Putative transporter	3	4	0.03
<i>phnF</i>	Putative DNA-binding transcriptional regulator of phosphonate uptake and biodegradation	3	3	0.03
<i>pqiA</i>	Paraquat-inducible membrane protein A	3	4	0.03
<i>kdpB</i>	Potassium translocating ATPase, subunit B	3	4	0.03
<i>puuP</i>	Putrescine importer	3	2	0.03
<i>dnaX</i>	DNA polymerase III/DNA elongation factor III, tau and gamma subunits	4	4	0.03
<i>lhr</i>	Putative ATP-dependent helicase	3	4	0.03
<i>astD</i>	Succinylglutamic semialdehyde dehydrogenase	3	3	0.03
<i>gabP</i>	Gamma-aminobutyrate transporter	3	3	0.03
<i>ytff</i>	Putative inner membrane protein	3	1	0.03
<i>phnK</i>	Carbon-phosphorus lyase complex subunit	4	4	0.03
<i>gutQ</i>	Putative phosphosugar-binding protein	3	2	0.03
<i>acs</i>	Acetyl-CoA synthetase	5	4	0.03
<i>narZ</i>	Nitrate reductase 2 (NRZ), alpha subunit	3	4	0.03
<i>chaA</i>	Calcium/sodium:proton antiporter	3	1	0.03
<i>ligB</i>	DNA ligase, NAD(+)-dependent	3	5	0.03
<i>nfrA</i>	Bacteriophage N4 receptor, outer membrane subunit	3	5	0.03
<i>fdoG</i>	Formate dehydrogenase-O, large subunit	3	3	0.03
<i>carB</i>	Carbamoyl-phosphate synthase large subunit	4	4	0.02
<i>yfbS</i>	Putative transporter	4	3	0.02
<i>yhgF</i>	Putative transcriptional accessory protein	3	4	0.02
<i>rseP</i>	Zinc metalloproteinase	3	3	0.02
<i>ybaP</i>	Hypothetical protein	3	2	0.02
<i>lpxM</i>	Myristoyl-acyl carrier protein (ACP)-dependent acyltransferase	5	1	0.02
<i>ulaG</i>	Putative L-ascorbate 6-phosphate lactonase	3	3	0.02
<i>dam</i>	DNA adenine methylase	3	2	0.02
<i>yhiN</i>	Putative oxidoreductase with FAD/NAD(P)-binding domain	3	2	0.02
<i>garK</i>	Glycerate kinase I	3	4	0.02
<i>acnB</i>	Bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase	3	2	0.02
<i>trpD</i>	Fused glutamine amidotransferase (component II) of anthranilate synthase/anthranilate phosphoribosyl transferase	4	3	0.02
<i>uup</i>	Fused predicted transporter subunits of ABC superfamily: ATP-binding components	3	3	0.02
<i>hisB</i>	Fused histidinol-phosphatase/imidazoleglycerol-phosphate dehydratase	3	2	0.02
<i>yphD</i>	Putative sugar transporter subunit: membrane component of ABC superfamily	3	3	0.02
<i>tauB</i>	Taurine transporter subunit	3	3	0.02
<i>creC</i>	Sensory histidine kinase in two-component regulatory system with CreB or PhoB, regulator of the CreBC regulon	4	2	0.02
<i>yidR</i>	Hypothetical protein	3	3	0.02
<i>ydcR</i>	Fused predicted DNA-binding transcriptional regulator/predicted amino transferase	3	2	0.02
<i>cpsB</i>	Mannose-1-phosphate guanyltransferase	3	2	0.02

(continued)

Table 1 Continued

Gene Name	Function	Occurrence in How Many Samples	Number of Alleles	Frequency in Population at Peak
<i>eutB</i>	Ethanolamine ammonia-lyase, large subunit, heavy chain	3	3	0.02
<i>proA</i>	Gamma-glutamylphosphate reductase	3	3	0.02
<i>sseB</i>	Rhodanase-like enzyme, sulfur transfer from thiosulfate	3	3	0.02
<i>rutF</i>	Putative oxidoreductase, flavin:NADH component	3	3	0.02
<i>aer</i>	Fused signal transducer for aerotaxis sensory component/ methyl accepting chemotaxis component	3	2	0.02
<i>mglC</i>	Methyl-galactoside transporter subunit	3	2	0.02
<i>ygjK</i>	Putative glycosyl hydrolase	3	2	0.02
<i>ybhG</i>	Putative membrane fusion protein (MFP) component of efflux pump, membrane anchor	3	2	0.02
<i>hisD</i>	Bifunctional histidinal dehydrogenase/histidinol dehydrogenase	3	1	0.02

^aThe entries include single nucleotide substitutions and 1-base insertions/deletions in coding sequences.

These results are consistent with previous chemostat studies initiated with the same ancestral strain using identical culture conditions (Notley-McRobb et al. 2003), thus this single population was typical of other previous shorter-run chemostat cultures.

In order to check whether the phenotypic sweeps match the genotypic sweeps, we plotted the total frequency of all *rpoS* and *malT* nonsynonymous alleles in the population. There are 7 and 26 alleles of *rpoS* and *malT*, respectively, in the population (table 1). As shown in figure 2, the sum contribution of all the *rpoS* and *malT* alleles combined matches the phenotypic dynamics due to mutations in these genes. So the dynamics of mutational change in the metagenome correlates well with the independent phenotypic sweeps.

To gain further insight into genotype dynamics, we plotted the frequency of mutations in all 1,717 genes containing sequence changes (fig. 3). In figure 3B, for clarity we show a subset of the data in figure 3A at a 10-fold magnified scale and show the vast number of mutations that occur transiently, particularly before day 20. The mutation numbers in each sample are summed in figure 3C, which also shows that the total number of mutations decreases after 20 days, but at least 42 different mutations remain throughout the course of the 60 days.

Individual Gene Dynamics and Allele Frequencies

As shown in figure 3A, it is evident that already at day 7, a high proportion of the population had acquired mutations in *rpoS* and *mgl* genes; such strongly beneficial mutations have appeared reproducibly in all glucose-limited populations of BW2952 in 3–5 days at this dilution rate (Notley-McRobb et al. 2003; Ferenci 2008b) and are also consistent with the phenotypic sweep shown in figure 2. Over the remaining 54 days, the metagenome continued to include mutations in *rpoS* and *mgl* in over 80% of bacterial sequences, although

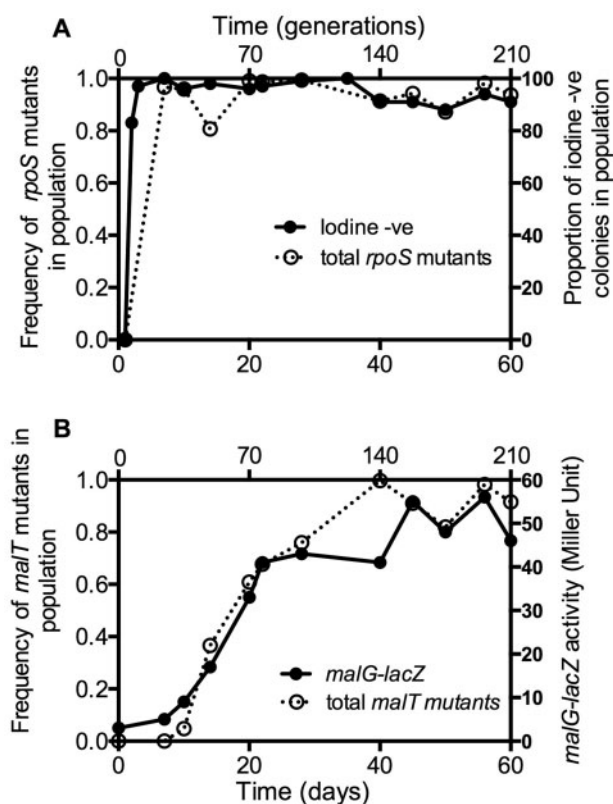


Fig. 2.—Correlation between phenotypic and genotypic sweeps. (A) The frequency of isolates from the population showing altered phenotype (iodine staining negative, ●) and the sum of all the different *rpoS* mutations (○) in the population at given time points during 60 days of glucose-limited continuous culture as shown in [supplementary table S1, Supplementary Material](#) online. (B) Changes in expression of the *mal* regulon (*malG-lacZ* activity of culture, ●) and the sum of the different *malT* mutations (○) in the population at given time points. The proportion of the population with iodine-negative phenotype and the *malG-lacZ* activity of the culture were estimated as described in the Materials and Methods section.

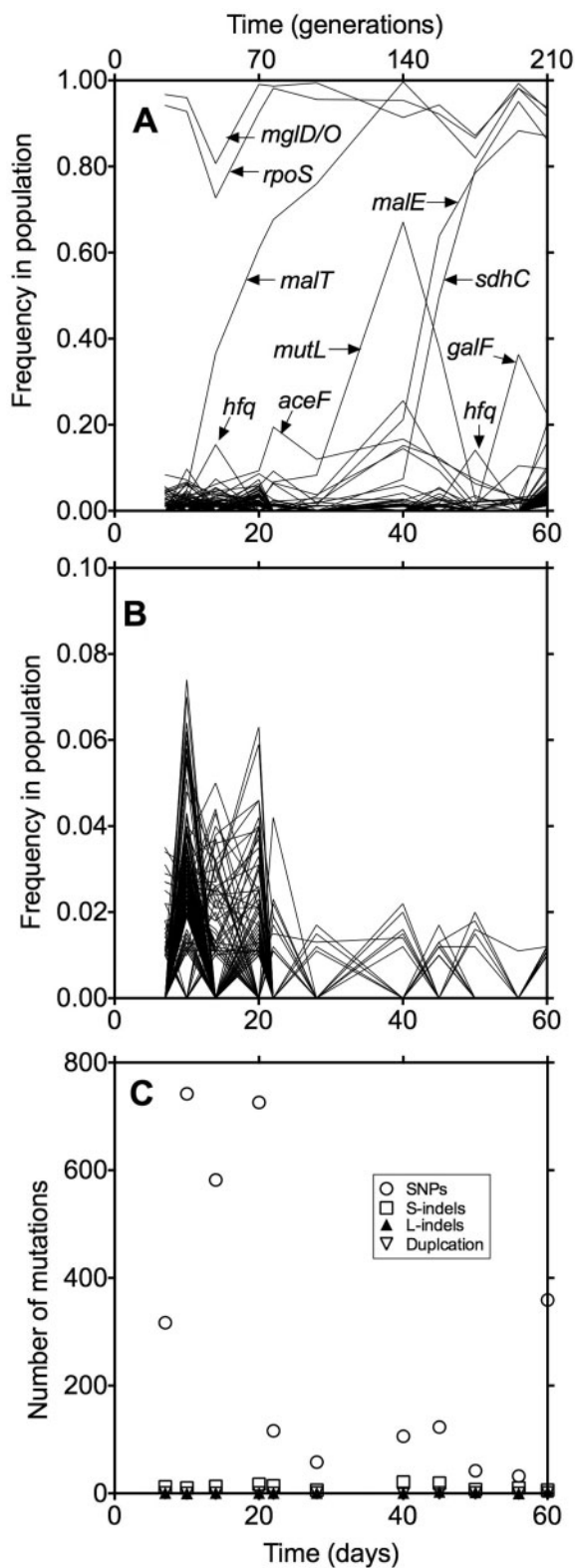


Fig. 3.—Dynamics of mutational changes in a chemostat population. (A) The total collection of the 2,709 mutations in 1,717 genes or regions identified in a glucose-limited chemostat population of *E. coli* K-12 strain BW2952. Each line represents a gene in which a mutation had occurred.

never to 100% fixation. The only time-points at which the *rpoS* and *mgI* mutations drop below 80% are at days 14 and 50, when *hfq* mutations transiently increase to greater than 10% abundance. As shown before (Maharjan and Ferenci 2013), *hfq* mutants remain *rpoS*⁺*mgI*⁺ explaining the reciprocal blips in the *rpoS/hfq* proportions (fig. 3A).

Another significant sweep was the *malT* change from day 14, which also, as previously observed (Notley-McRobb and Ferenci 1999b), became highly represented for the rest of the experiment. The *malT* sweeps were then followed by *malE* and *sdhC* mutations (figs. 3 and 4); the possible role of these mutations is discussed below. These were not previously seen in chemostats but occurred after 40 days, beyond the 28 days sampled in earlier studies.

The most unique change occurred concurrently in several genes, peaking at days 40–45 in greater than 50% of the population and included a *mutL* mutation. Mutations in this gene result in a mutator phenotype and so it is likely that the other exactly coincident peaks (in *yggE*, *tpiA*) were linked with the mutator genome when it increased in frequency. In turn, as shown below in figure 4E, the enrichment of the mutator-associated mutations is most likely due to the benefit that other mutations such as *rpoS*, *mgID*, and *malT* provided at day 40 but this lineage was outcompeted by other members of the same population with different beneficial combinations. Independent whole-genome sequencing of a clone isolated from the day 40 sample showed that the *mutL*-containing genome also has mutations in *rpoS*, *mgID*, *malT*, *yggE* and *tpiA*, which further confirms their linkage. Mutator phenotypic assay showed that this clone also had an elevated rate (>100-fold) of spontaneous mutation. Fitness and other phenotypic contributions of *yggE* and *tpiA* encoding YggE hypothetical protein and triosephosphate isomerase, respectively, are unknown. This *mutL* finding increases to three the mutator types found in chemostat populations (*mutS* and *mutY* were previously found [Notley-McRobb, Seeto, et al. 2002]) so this form of temporary enrichment of mutators is a common feature of a rapidly evolving population.

Another mutation appeared in greater than 50% of the population, in *phnN* (fig. 3B). This transient peak in *phnN* at day 28 exactly coincided with the appearance and disappearance of a *malT* mutation at position 3440466 (*malT-18*) (see fig. 3C and E) so probably was genetically hitchhiking with this

Fig. 3.—Continued

(B) An extended view of the subset of genes from (A) in which mutant sequences were present in less than 20% of the population during the 60 days. (C) The total load of mutations in the population over 60 days. The counts of SNPs, single base-pair indels (S-indels), large indels (L-indels), and duplication mutations that were detected in at least 1% of the total population across 11 different samples collected over 60 days of continuous culture are shown. Frequencies of each mutation were estimated from the number of reads and total coverage at that particular genome position.

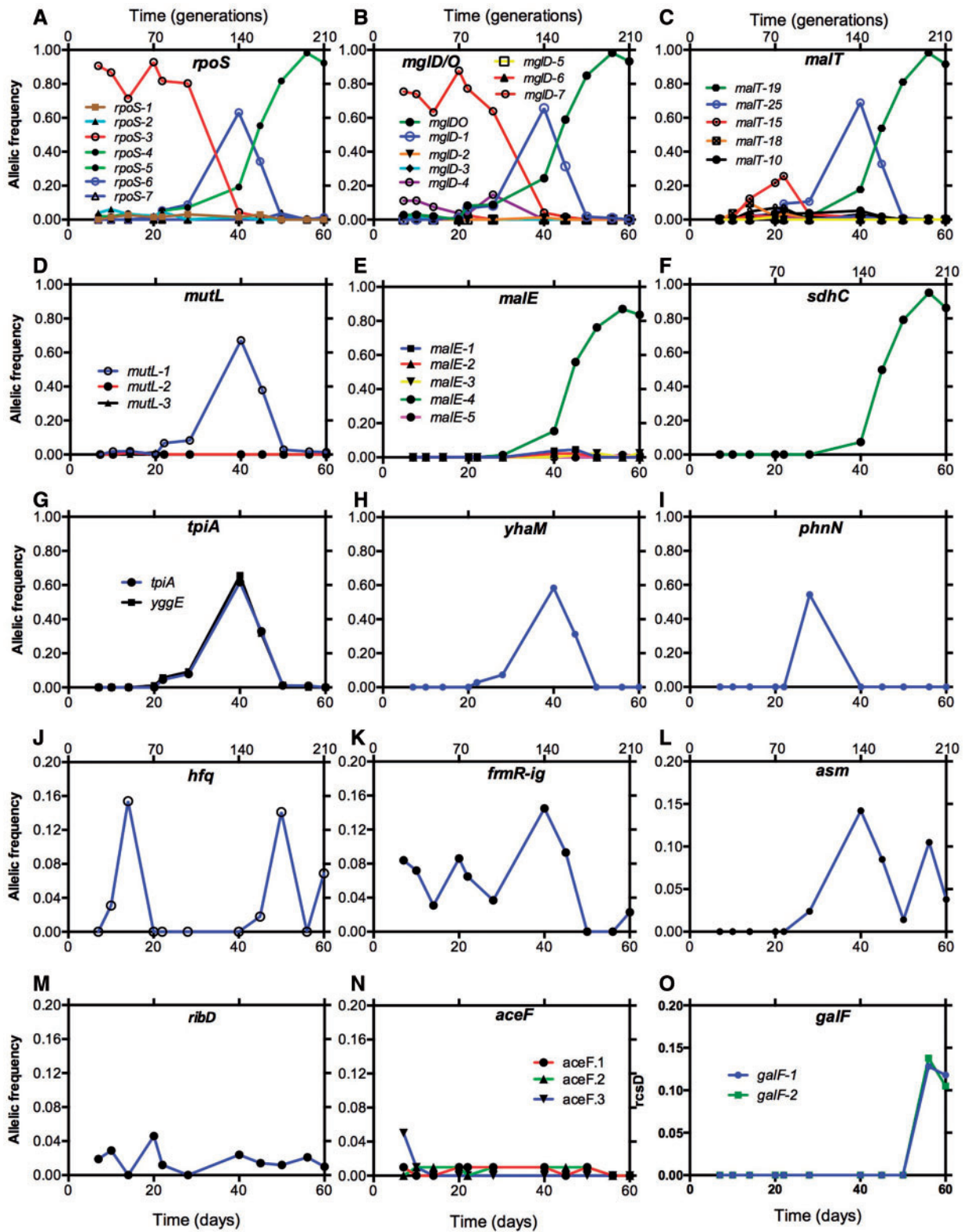


FIG. 4.—Dynamics of frequency changes of alleles within individual genes. The genomic position of each mutation in each gene is listed in the [supplementary table S1, Supplementary Material](#) online. Labels are shown for only the more dominant 5 of the 26 nonsynonymous *malT* alleles.

equally transient allele. It is not obvious why *phnN* would have a role under glucose limitation, being involved in phosphonate utilization that is cryptic in *E. coli* K-12 (Hove-Jensen et al. 2003).

Smooth Phenotypic Sweeps Are Genetically Complex

In figure 4A–C, we can observe how multiple soft sweeps involving *rpoS*, *mgl*, and *malT* mutations occur in unprecedented detail. Most interestingly, the net sweeps reducing the ancestral proportion of wild-type sequences shown in figures 2 and 3 are caused by a succession of different combinations of the mutations in *rpoS*, *mgl*, and *malT*. There were 7–26 different alleles of the genes that appear and disappear together over the 60 days even though the total population load of *rpoS**mgl**malT* triple mutants exceeded 80% over extended periods (fig. 3). The presence of multiple alleles of these genes is strongly indicative of soft sweeps in the population. In the large population studied, many independent strongly beneficial mutations can arise and changes at several sites within a gene can satisfy the fitness gain; in *malT*, *mglD* and *rpoS*, this is certainly the case (Notley-McRobb and Ferenci 1999a, 1999b; Notley-McRobb, King, et al. 2002).

The first dominant *rpoS*-3 allele was a mutation at genome position 2751003, prominent already at day 7. Minor proportions of alleles such as *rpoS*-2 were also present early on. Other minor alleles of *rpoS* such as *rpoS*-4 were barely detectable prior to day 20 but became the dominant allele by day 56 (fig. 4A). In between, the first dominant allele, *rpoS*-3, was eliminated by successive sweeps of *rpoS*-4 and then *rpoS*-6, which had a sharp, transient peak at day 40. The loss of function mutations in *rpoS* at each position does not have a significant fitness difference, as evidenced by iodine staining, so the rapid alterations in the proportion of dominant *rpoS* alleles must have been due to additional mutations; the combinations leading to the sweeps can be deduced from the concurrent frequencies of other mutations as shown below.

The *mgl* profile in figure 4B showed a near-identical pattern of allele successions to that of *rpoS*. The first dominant *mgl* mutation was at position 2132151 (*mglD*-7), already by day 7. Another *mgl* allele (*mglD*-4) present in lower proportions at early times was at position 2132020 (fig. 4B). The proportions of these alleles in combination match that of *rpoS*-3, suggesting that the strain with *rpoS*-3 diverged into *rpoS*-3*mglD*-7 (dominant) and *rpoS*-3*mglD*-4 (minor) before day 7. From day 28, another allele combination, *rpoS*-6*mglD*-1 swept to over 60% abundance, reducing the proportion of the earlier *rpoS**mglD* combinations. Further, the final dominant *mgl* allele (*mglD*O) emerges beyond day 28, with an *mglO* operator mutation in combination with yet another *rpoS* allele, *rpoS*-4. The proportion of the individual *rpoS* and *mgl* mutations contributing to the paired alleles was near identical and their trajectories were also remarkably similar, indicating the linkage

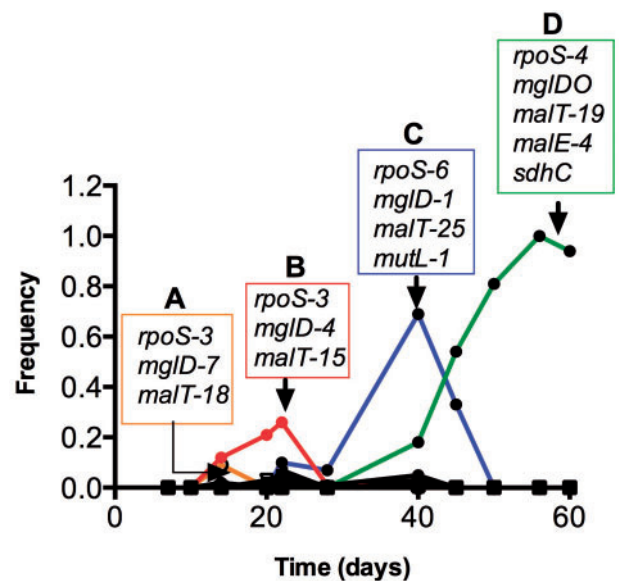


Fig. 5.—Summary of events superimposed on the trajectories of *malT* allele frequencies. The mutations peaking at particular times are here superimposed on the *malT* allele sweeps shown in figure 4C (*malT* panel) to show the combinations of mutations participating in replacements over the 60 days. The association of different alleles of *rpoS*, *mglD*, *malT*, *malE*, and *sdhC* were based on the allelic frequency dynamics shown in figure 4 and verified by sequencing of individual isolates (see text for more detail).

of the alleles in the same genome. These sweeps are indicated in a summary form shown in fig. 5.

The benefit of these *rpoS* and *mgl* combinations still does not explain the frequency changes, until one considers the additional sweeps of *malT* and other mutations in these backgrounds as summarized in figures 4 and 5. Mutations in *malT* became evident after day 10. Mutations at positions 3440466 (*malT*-18) and 3440328 (*malT*-15) reached 10–20% abundances together the coincident *rpoS*-3*mglD*-7 (group A in fig. 5) and *rpoS*-3*mglD*-4 (group B in fig. 5). At day 40, the transient peak with the *rpoS*-6*mglD*-1 combination coincided exactly with the spread of the *malT*-25 allele, both in scale and in profile (fig. 4C). Thus, it is most likely that the *rpoS*-6*mglD*-1 *malT*-25 (group C in fig. 5) combination was central to the sweep between days 28 and 40. As mentioned earlier, also coincident at day 40 was *mutL*. Thus, *mutL* and other associated genes must have hitchhiked together with *rpoS*-6*mglD*-1 *malT*-25 allelic combination.

The combination of mutations dominant at day 60 had yet another set of allele combinations in *rpoS*, *mgl*, and *malT*. The major sweep between days 40 and 60 contained the *rpoS*-4, *mglO* and *malT*-19 alleles, with *malT*-19 becoming prominent only after day 28 (fig. 5). Also spreading simultaneously with the same trajectory were *malE* with mutation at 4183156 (*malE*-4) and *sdhC* mutation at 657360. This

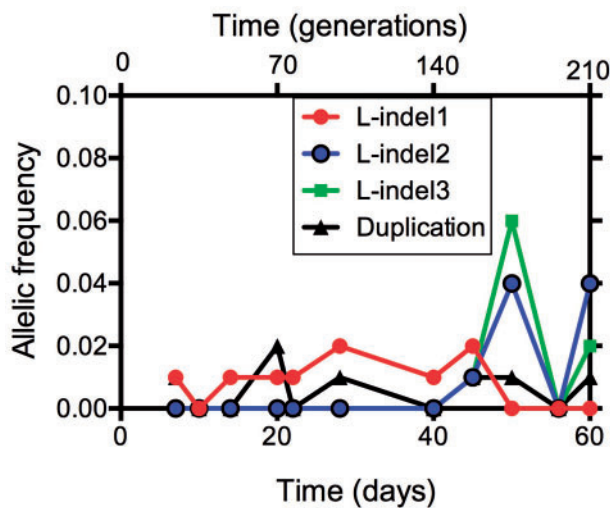


Fig. 6.—Dynamics of frequency changes of each large indel and duplication mutation found in the chemostat population. The genomic position of each mutation in each gene is listed in the [supplementary table S1, Supplementary Material](#) online.

rpoS-4mglOmalT-19malE-4sdhC (group D in fig. 5) combination resulted in the drastic reduction of all other previously appearing *rpoS*, *mgl*, and *malT* alleles in this population. The combination of the five mutations in one isolated clone from day 60 shown in figure 5 was confirmed by PCR of each gene and Sanger sequencing. The events at days 56–60 come closest to a purifying sweep in this experiment but even here the population still contained a significant proportion of *hfq* and other mutational lineages (fig. 4J–O), maintaining persisting diversity in the culture. The numerous low-frequency mutations (fig. 3B) also provided a rich source material for further evolution.

The Persistence of Low-Abundance Alleles of the Same Gene

As shown in figure 4J–O, multiple genes besides the above-mentioned *hfq* were altered at multiple time-points and in significant numbers. Several of the mutations shown in figure 4 were consistently present at or above the 1% abundance, without being eliminated by other fluctuations in the population. The persistence under glucose limitation suggests a fitness benefit of these mutations without a capacity to sweep through the whole population; this is best explained by a frequency-dependent fitness as was demonstrated for *hfq* (Maharjan, McKenzie, et al. 2013). Curiously, three alleles of *aceF* and two of *galF* were simultaneously present in the population. Only the last major change in frequency of other genes toward day 60 reduced the numbers of *aceF* mutants in the population, but the others shown in figure 4 were maintained and indeed rose in the case of *galF*.

AceF is a subunit of pyruvate dehydrogenase so mutations in *aceF* could change central metabolism; *aceF* mutants can for example have elevated pyruvate levels in cells (Tomar et al. 2003). It is not clear how this could change fitness under glucose limitation but is the kind of change that could give rise to cross-feeding consortia as seen in long-running chemostats (Rosenzweig and Adams 1994). As shown in figure 4B, some other mutations in *asm*, *galF*, *ribD*, and near *frmR* were persistently present over many samples until day 60, but never exceeded 20% of the population. The apparently frequency-dependent benefit of these mutations is so far unidentified.

The Persistence of Low-Abundance Indels in the Chemostat Population

As shown in figure 6, four larger genome changes, including four large insertions and a tandem duplication, also persisted as minority types over multiple samples. Of the five large mutations (4 L-indels and 1 duplication), four persisted as minority types over multiple time points (fig. 6 and [supplementary table S1, Supplementary Material](#) online). One of the four large insertions (L-indel1) was located inside *mglD*, so it is likely to have the same beneficial role as the other *mgl* mutations found as SNPs (fig. 4B) causing loss of MglD repressor function. L-indel 2 and L-indel 4 were located in *hyfA* and *ydeK*, respectively. These encode hydrogenase 4, 4Fe–4S subunit and a putative lipoprotein, whereas L-indel 3 was located in an intergenic region near *dnaK*, a molecular chaperone protein. The large tandem duplication comprised 23,060 bp (positions 1301981–1325042 of the ancestral BW2952 genome), which contains 25 genes. The role of the duplication of these genes under glucose-limitation is unknown and covers no obviously beneficial functions.

Discussion

We demonstrated that apparently simple phenotypic sweeps can hide complex genetic changes in populations. As most clearly seen with *rpoS* over 60 days, the constantly high abundance of the low-RpoS phenotype was due to the successive contribution of three distinct *rpoS* mutations in combination with other beneficial mutations. Each of these alleles arose directly from the ancestor, but became more dominant when additional mutations were acquired. The richness of the population in *rpoS* mutations extends beyond the three prominent alleles and at least five other alleles could be detected at lower frequencies; other alleles were probably there but below the detection limit of metagenomics. In all the alleles in the population, there was phenotypic convergence as to the benefit of *rpoS* mutations, but the combination of different alleles with different other mutations contributes to the complexity of fitness pathways in the evolving population. Likewise, 8 *mgl* mutations and 26 *malT* mutations contribute to the deceptively smooth phenotypic behavior (fig. 2) but highly lumpy mutational landscape in the evolving population

(fig. 5). Simple assumptions about 1:1 relationships between phenotypic sweeps and mutations are therefore unwarranted, especially in large populations.

Our results are consistent with the contribution of multiple alleles and soft sweeps to most, but not all, adaptations in a large bacterial population. The significant frequency changes resembling hard sweeps and originating from single alleles were associated with the *mutL* and *sdhC* mutations (fig. 4D and F, respectively). *sdhC* is the only single-allele gene rising from 1% to greater than 90% frequency in the population in a single sweep. The *mutL* case is more misleading however, because mutator mutations are not themselves beneficial and thus hitchhiked with other beneficial mutation(s) at day 40. It is remarkable that a mutator mutation, thought to be important when evolution is limited by mutation supply (de Visser et al. 1999) actually becomes common even in a large population where every single nucleotide mutation will occur nearly ten times every generation.

One of the most fascinating aspects of the multiple alleles involving *malT*, *mgl*, or *rpoS* was that each mutation appeared from an undetectably low base frequency even late in the population and not from other, more frequent types. This is consistent with the notion that there is a significant pool of bacteria with mutations at a low frequency in the population from which most of these sweeps emerged. In the culture studied, it is worth recalling that 1% frequency is still a subpopulation of 10^8 cells, large enough to be a source of further mutational innovations. Our results do not resemble any of the proposed models recently discussed (e.g., fig. 2 in Barrick and Lenski [2013] or fig. 1 in Herron and Doebeli [2013]), all of which assume that beneficial mutations repeatedly occur consecutively in the majority types in a population. In our study, all the successful later lineages emerged from the heterogeneity in the population present at less than 1% abundance in the population.

Periodic selection, as seen in the fluctuation in the proportion of neutral mutations in a population, is a long-observed feature of evolving bacterial populations (Atwood et al. 1951b). Because periodic selection events are operationally described by the fluctuation in the proportion of a neutral marker (such as T5 resistance) in the population, they describe population changes due to all mutational sweeps, even those with mutations in the same gene. The fluctuations in neutral mutations in chemostats have been described in detail (Atwood et al. 1951a; Kubitschek 1970). A characteristic of these periodic events is that they are frequent and only partial in reducing the proportion of neutral mutations. As an example, our own published experiments in following periodic selection events (Notley-McRobb and Ferenci 2000; Notley-McRobb et al. 2003) demonstrated, on reanalysis of the data, that there was approximately 1 sweep/generation at 0.1 h^{-1} dilution rate as used in this study. We can now see how the multiple frequency changes involving many genes (figs. 3 and 4) determine frequent periodic selection events

and changes in genotype and phenotype in a population. Multiple changes in 23 genes with population frequency shifts exceeding 10% were evident from figures 3 and 4 and multiple combinations of mutations contributed to the number of fluctuations in this large population. Our results are thus consistent with classic observations that periodic selection events are numerous, show only partial sweeps and are indeed more frequent than can be explained by 1-gene:1-mutation combinations. The observation that different combinations of mutations contributed to population shifts means that it is unclear whether true clonal interference (de Visser and Rozen 2006; Barroso-Batista et al. 2014) is operating in this population. Our results do however strongly resemble recent results with yeast populations in that multiple mutations arise and move synchronously through the population as mutational “cohorts” (Lang et al. 2013).

This high-coverage metagenomics analysis identified a large number of mutations spreading in glucose-limited populations. One of the most common new mutations, in *malE*, may provide a benefit by reducing the load imposed by the *malT* mutations widespread in the population. The *malT* constitutivity mutation results in high levels of not only LamB protein (beneficial because this increases the transport of limiting glucose [Death et al. 1993]) but also other *mal* regulon proteins including MalE (which is not of benefit). MalE is a protein made in high amounts and has the cost of having to be secreted into the periplasmic space (Kellermann and Ferenci 1982). A likely benefit of its loss is thus the alleviation of the cost of producing large amounts of an unneeded protein. Another possibility is that loss of MalE alleviates competition in protein secretion, because useful proteins such as LamB need to be exported from the cytoplasm in large amounts through the same secretion pathway (Misra et al. 1991).

We do not know the physiological benefit of several of the other mutations identified here, such as in *acs* and *aceF* (table 1). Given the role of these gene products in central metabolism affecting the tricarboxylic cycle, it can be proposed that metabolic adaptations can provide a benefit under glucose limitation. Indeed, an alternative to the TCA cycle, the glyoxylate cycle, has been implicated in glucose-limited growth so a speculative explanation is a beneficial switch between these pathways (Fischer and Sauer 2003; Maharjan et al. 2005). The nature of these benefit(s) will need to be investigated by detecting changes in metabolic fluxes in the mutant strains. There are also less obvious benefits under glucose limitation among the mutations present in table 1 in multiple samples, such as in *galF* and the putative efflux protein *mdtG* as well as the indels in figure 6. Further physiological analysis is needed to identify these benefits.

The single most successful combination of mutations peaked at day 56 and analysis of individual isolates purified from the population confirmed the combination of mutations in the dominant type shown in figure 5. At that time, the population came closest to a purifying sweep with the

major type close to 100% abundance; nevertheless the minority types carrying *hfq*, *asm*, and *ribD* mutations were largely unaffected at days 56–60. Two large insertions and a tandem duplication also persisted through day 60 (fig. 6). From the metagenomics, we do not know whether any or all of these low-abundance mutations occurred in combination in the same cell to provide the fitness benefit under glucose limitation.

Crucially from the point of view of maintained diversity, the total number of mutations in the population did not continue to decrease (fig. 1) and the continued presence of greater than 40 mutations in the population over time indicates that a diverse range of mutations is maintained in the population. The mutation supply and the minority lineages in figures 4 and 5 provide the genetic potential for further evolution as well as the maintenance of heterogeneity during the later phases of this experiment.

In conclusion, high coverage metagenomic analysis has revealed in unprecedented detail the multiple soft sweeps in an evolving *E. coli* population and the way these contribute to the sum total of phenotypic changes. The results explain the early findings on periodic selection events and why incomplete sweeps were previously found. The data have identified several low-proportion mutations not previously found in evolving chemostat populations and have established that minority alleles can persist stably for hundreds of generations in independent lineages.

Supplementary Material

Supplementary table S1 is available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

This study was supported by an Australian Research Council Discovery Grant to R.M. and T.F., the National 973 Program of China Grant (2013CB733904) and the International Science & Technology Cooperation Program of China (2012DFG31680) to L.F., the National Natural Science Foundation of China Key Program Grant (31030002) to L.W., and the National Key Program for Infectious Diseases of China (2013ZX10004221) to B.L.

Literature Cited

- Armstrong RA, McGehee R. 1980. Competitive exclusion. *Am Nat.* 115: 151–170.
- Atwood K, Schneider LK, Ryan FJ. 1951a. Periodic selection in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 37:146.
- Atwood KC, Schneider LK, Ryan FJ. 1951b. Selective mechanisms in bacteria. *Cold Spring Harb Symp Quant Biol.* 16:345–355.
- 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.
- Barroso-Batista J, et al. 2014. The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps. *PLoS Genet.* 10: e1004182.
- Berg OG. 1995. Periodic selection and hitchhiking in a bacterial population. *J Theor Biol.* 173:307–320.
- Cox M, Peterson D, Biggs P. 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11:485.
- de Visser JAGM, Rozen DE. 2006. Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. *Genetics* 172: 2093–2100.
- de Visser JAGM, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE. 1999. Diminishing returns from mutation supply rate in asexual populations. *Science* 283:404–406.
- Death A, Notley L, Ferenci T. 1993. Derepression of LamB protein facilitates outer membrane permeation of carbohydrates into *Escherichia coli* under conditions of nutrient stress. *J Bacteriol.* 175:1475–1483.
- Dykhuizen DE, Hartl DE. 1983. Selection in chemostats. *Microbiol Rev.* 47: 150–168.
- Ferenci T. 2008a. Bacterial physiology, regulation and mutational adaptation in a chemostat environment. *Adv Microb Physiol.* 53:169–229.
- Ferenci T. 2008b. The spread of a beneficial mutation in experimental bacterial populations: the influence of the environment and genotype on the fixation of *rpoS* mutations. *Heredity* 100:446–452.
- Ferenci T, et al. 2009. Genomic sequencing reveals regulatory mutations and recombinational events in the widely used MC4100 lineage of *Escherichia coli* K-12. *J Bacteriol.* 191:4025–4029.
- Fischer E, Sauer U. 2003. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *J Biol Chem.* 278: 46446–46451.
- Gaffe J, et al. 2011. Insertion sequence-driven evolution of *Escherichia coli* in Chemostats. *J Mol Evol.* 72:398–412.
- Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv* 1207.3907v2 [q-bio.GN].
- Gerrish PJ, Lenski RE. 1998. The fate of competing beneficial mutations in an asexual population. *Genetica* 103:127–144.
- Gresham D, et al. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* 4:e1000303.
- Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169: 2335–2352.
- Herron MD, Doebeli M. 2013. Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. *PLoS Biol.* 11:e1001490.
- Hove-Jensen B, Rosenkrantz TJ, Haldimann A, Wanner BL. 2003. *Escherichia coli phnN*, encoding ribose 1,5-bisphosphokinase activity (phosphoribosyl diphosphate forming): dual role in phosphonate degradation and NAD biosynthesis pathways. *J Bacteriol.* 185: 2793–2801.
- Kellermann OK, Ferenci T. 1982. Maltose-binding protein from *Escherichia coli*. *Methods Enzymol.* 90:459–463.
- Kinnersley MA, Holben WE, Rosenzweig F. 2009. E Unibus Plurum: genomic analysis of an experimentally evolved polymorphism in *Escherichia coli*. *PLoS Genet.* 5:e1000713.
- Koch AL. 1974. The pertinence of the periodic selection phenomenon to prokaryote evolution. *Genetics* 77:127–142.
- Kubitschek HE. 1970. Introduction to research with continuous cultures. Englewood Cliffs (NJ): Prentice-Hall Inc.
- Kvitek DJ, Sherlock G. 2013. Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLoS Genet.* 9: e1003972–3972.
- Lang GI, et al. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500:571–574.

- Leverett M, et al. 2010. Molecular and evolutionary bases of within-patient genotypic and phenotypic diversity in *Escherichia coli* extraintestinal infections. *PLoS Pathog.* 6:e1001125.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25:1754–1760.
- Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Maharjan R, et al. 2010. Genomic identification of a novel mutation in *hfq* that provides multiple benefits in evolving glucose-limited populations of *Escherichia coli*. *J Bacteriol.* 192:4517–4521.
- Maharjan R, et al. 2012. The multiplicity of divergence mechanisms in a single evolving population. *Genome Biol.* 13:R41.
- Maharjan R, McKenzie C, Yeung A, Ferenci T. 2013. The basis of antagonistic pleiotropy in *hfq* mutations that have opposite effects on fitness at slow and fast growth rates. *Heredity* 110:10–18.
- Maharjan R, Nilsson S, et al. 2013. The form of a trade-off determines the response to competition. *Ecol Lett.* 16:1267–1276.
- Maharjan R, Seeto S, Notley-McRobb L, Ferenci T. 2006. Clonal adaptive radiation in a constant environment. *Science* 313:514–517.
- Maharjan RP, Ferenci T. 2013. Epistatic interactions determine the mutational pathways and coexistence of lineages in clonal *Escherichia coli* populations. *Evolution* 67:2762–2768.
- Maharjan RP, Yu PL, Seeto S, Ferenci T. 2005. The role of isocitrate lyase and the glyoxylate cycle in *Escherichia coli* growing under glucose limitation. *Res Microbiol.* 156:178–183.
- Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol.* 28:659–669.
- Misra R, Peterson A, Ferenci T, Silhavy TJ. 1991. A genetic approach for analyzing the pathway of LamB assembly into the outer membrane of *Escherichia coli*. *J Biol Chem.* 266:13592–13597.
- Notley L, Ferenci T. 1995. Differential expression of *mal* genes under cAMP and endogenous inducer control in nutrient stressed *Escherichia coli*. *Mol Microbiol.* 16:121–129.
- Notley-McRobb L, Ferenci T. 1999a. Adaptive *mgl*-regulatory mutations and genetic diversity evolving in glucose-limited *Escherichia coli* populations. *Environ Microbiol.* 1:33–43.
- Notley-McRobb L, Ferenci T. 1999b. The generation of multiple coexisting *mal*-regulatory mutations through polygenic evolution in glucose-limited populations of *Escherichia coli*. *Environ Microbiol.* 1:45–52.
- Notley-McRobb L, Ferenci T. 2000. Experimental analysis of molecular events during mutational periodic selections in bacterial evolution. *Genetics* 156:1493–1501.
- Notley-McRobb L, King T, Ferenci T. 2002. *rpoS* mutations and loss of general stress resistance in *Escherichia coli* populations as a consequence of conflict between competing stress responses. *J Bacteriol.* 184:806–811.
- Notley-McRobb L, Seeto S, Ferenci T. 2002. Enrichment and elimination of *mutY* mutators in *Escherichia coli* populations. *Genetics* 162:1055–1062.
- Notley-McRobb L, Seeto S, Ferenci T. 2003. The influence of cellular physiology on the initiation of mutational pathways in *Escherichia coli* populations. *Proc Natl Acad Sci U S A.* 270:843–848.
- Pennings PS, Hermisson J. 2006. Soft Sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Mol Biol Evol.* 23:1076–1084.
- Phan K, Ferenci T. 2013. A design-constraint trade-off underpins the diversity in ecologically important traits in species *Escherichia coli*. *ISME J.* 7:2034–2043.
- Rosenzweig RF, Adams J. 1994. Microbial adaptation to a changeable environment: cell-cell interactions mediate physiological and genetic differentiation. *Bioessays* 16:715–717.
- Tomar A, Eiteman MA, Altman E. 2003. The effect of acetate pathway mutations on the production of pyruvate in *Escherichia coli*. *Appl Microbiol Biotechnol.* 62:76–82.
- Wang L, et al. 2010. Divergence involving global regulatory gene mutations in an *Escherichia coli* population evolving under phosphate limitation. *Genome Biol Evol.* 2:478–487.
- Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25:2865–2871.

Associate editor: Daniel Sloan