

RESEARCH

Open Access

# The multiplicity of divergence mechanisms in a single evolving population

Ram P Maharjan<sup>1</sup>, Thomas Ferenci<sup>1</sup>, Peter R Reeves<sup>1</sup>, Yang Li<sup>2,3</sup>, Bin Liu<sup>2,3</sup> and Lei Wang<sup>2,3\*</sup>

## Abstract

**Background:** Evolutionary divergence is common within bacterial species and populations, even during a single bacterial infection. We use large-scale genomic and phenotypic analysis to identify the extent of diversification in controlled experimental populations and apply these data to differentiate between several potential mechanisms of evolutionary divergence.

**Results:** We defined testable differences between five proposed mechanisms and used experimental evolution studies to follow eight glucose-limited *Escherichia coli* chemostat populations at two growth rates. Simple phenotypic tests identified 11 phenotype combinations evolving under glucose limitation. Each evolved population exhibited 3 to 5 different combinations of the 11 phenotypic clusters. Genome sequencing of a representative of each phenotypic cluster from each population identified 193 mutations in 48 isolates. Only two of the 48 strains had evolved identically. Convergent paths to the same phenotype occurred, but two pleiotropic mutations were unique to slow-growing bacteria, permitting them greater phenotypic variance. Indeed, greater diversity arose in slower-growing, more stressed cultures. Mutation accumulation, hypermutator presence and fitness mechanisms varied between and within populations, with the evolved fitness considerably more uniform with fast growth cultures. Negative frequency-dependent fitness was shown by a subset of isolates.

**Conclusions:** Evolutionary diversity is unlikely to be explained by any one of the available mechanisms. For a large population as used in this study, our results suggest that multiple mechanisms contribute to the mix of phenotypes and evolved fitness types in a diversifying population. Another major conclusion is that the capacity of a population to diversify is a function of growth rate.

## Background

Heritable variation is common in evolution and certainly within bacterial species and populations [1,2]. Bacterial divergence is rapid and even evident within the course of a single bacterial infection [3]. One of the most intriguing questions in evolutionary biology is how such variation emerges and persists. It used to be believed that a population evolving in a homogeneous environment should exhibit low sympatric diversity as a result of periodic selection of a fitter isolate and the competitive niche exclusion of weaker isolates [4]. Therefore, environmental heterogeneity was considered to be the main source of diversity [5]. However, several experimental evolution studies suggest that diverse types coexist even

in an initially homogeneous culture supported by a single resource [6-8].

Several distinct mechanisms or explanations have been proposed for the maintenance of diversity in evolving populations, including frequency-dependent selection [9,10], cross-feeding and cheating between population members [11,12], metabolic and other trade-offs [13,14], mutation selection balance [15,16] and regulatory degeneracy leading to convergent evolution [7,17]. The multiplicity of mechanisms begs the question: what happens in real diversifying populations? Do individual populations predominantly diversify by one of the above mechanisms or is evolved diversity a result of a mixture of mutation-selection balance, frequency-dependent selection, trade-offs with or without mutation rate differences, niche creation or convergent evolution through degenerate pathways? The predictability and a systems-

\* Correspondence: wanglei@nankai.edu.cn

<sup>2</sup>TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, Tianjin 300457, PR China

Full list of author information is available at the end of the article

level description of evolution [18] require an answer to this question.

To approach the above questions, the signatures of individual divergence mechanisms need to be searched for in the evolved isolates. It is also necessary to obtain an accurate characterization of the extent of diversity in populations at both phenotypic and genotypic levels. Bacterial cultures, with their large populations, rapid growth and regulated mutation rates should be ideal subjects for the experimental analysis of evolution and diversification [5,18,19]. Genomic diversity can be readily identified through genome sequencing, as has been done in several experimental populations [7,17,20,21]. However, owing to the difficulty of identifying phenotypic diversity, most previous experimental studies focused on a few adaptive isolates or mutations and no experimental study has systematically characterized the extent of diversity at both phenotypic and genotypic levels in multiple parallel populations. Also on ecological grounds, studies of diversity in experimental populations have tended to use culture conditions with multiple or alternative niches [22-27]. Even the intensively studied Lenski populations evolved in a selection environment that fluctuated temporally in terms of nutrient availability through multiple cycles and coexisting types have evolved to these fluctuations [28]. Recent genomic analyses of these long-term populations have revealed multiple mutations [20]. The beneficial effects of these mutations are beginning to be revealed [29,30], but many of the mutations have unidentified benefits or fitness effects. A systematic description of diversity in these populations has not been reported, although metagenomic studies indicate diversity does exist in the Lenski populations. The detection of minor types was limited by the resolution of the metagenomic technique [20]. In contrast, we showed that extensive diversification can be detected when extensive phenotypic screening is applied to populations, and described in detail the complex population structure of a single chemostat population [6].

We report a study that relates the experimental results on eight evolving populations to the above divergence mechanisms. The phenotypic testing permitted the identification of multiple phenotypic groupings, and fitness assays were applied to test for trade-offs and frequency dependence. A detailed analysis of genomic changes revealed whether the number of mutations in each isolate is significantly different. These data are used to discuss whether mutation-selection balance [15,16] or trade-offs with differences in mutational robustness [14] or regulatory degeneracy leading to convergent evolution [7,17] are involved in the mechanism of diversification. We therefore compare the number of mutations per isolate in both sympatric and allopatric populations.

In addition, we analyzed all 320 isolates compared here for mutator activity, because hypermutation can be selected in evolving populations [31,32]. We indeed find that two of eight populations contain a subset of isolates with elevated mutation rates. Our strain comparisons also address the proposal that convergent alternative mutational pathways lead to similar levels of fitness in the same culture [17,33]. We also followed frequency dependence in fitness assays and looked for mutations known to cause altered trade-offs, such as *hfq* and *rpoS*, that reset the trade-off between stress protection and nutritional competence (SPaNC) [13,34]. The availability of 48 genome sequences further allows us to detect patterns of mutations as well as simultaneous analysis of phenotype-genotype relationships in a large number of isolates. Altogether, the numbers of mutations responsible for multiple phenotypes provide for the first time a comprehensive analysis and clues to the kind of mechanisms and properties of isolates relevant to the various evolutionary models.

## Results

### Evolution of phenotypic diversity in parallel populations

In order to test how diversity emerges and is maintained in bacterial populations, we performed replicate evolution experiments using glucose-limited chemostat cultures of *Escherichia coli* at two separate growth or dilution rates (D); four at slow with  $D = 0.1 \text{ h}^{-1}$  (generation time of approximately of 6.9 hours) and four at fast with  $D = 0.6 \text{ h}^{-1}$  (generation time of 1.15 hours). These two conditions result in a 30-fold difference in basal mutation frequencies per generation [35] and this difference is important for determining the mechanisms of diversification; the two series of experiments test whether the extent of diversity changes under these otherwise near-identical selection conditions. All chemostats were initiated with a common ancestral strain, BW2952 [36], and the population size was maintained at approximately  $1.6 \times 10^{10}$  cells by a constant concentration of glucose in the feed medium throughout the evolution experiment. The cell density did not vary appreciably during the course of the experiments.

To determine the phenotypic diversity in evolving populations, 39 to 41 isolates from each of the 8 independent chemostat cultures were analyzed. Population 1 was used previously and the properties of its 41 isolates reported [6]. The samples were obtained from the last day before a particular chemostat malfunctioned or was terminated, between days 23 and 26, except for population 4, which stopped at day 18. Isolates were randomly chosen and characterized using the triple-phenotypic-test approach (see Materials and methods). This method identifies the six phenotype clusters in our best-analyzed population and detects the major population

fluctuations that occur in the first two weeks of culture [6]. Co-existence of four or more combinations of phenotypes was evident after two to three weeks in this earlier study [6] and population diversity did not change markedly between days 15 and 37, as shown in Figure S1 in Additional file 1. So samples from the eight populations at days 21 to 27 provide a good snapshot of population heterogeneity. We may underestimate the extent of diversity in this way, but as shown in Figure 1a, significant heterogeneity was detectable in a single sample from all eight cultures. We found two to six combinations of phenotypic classes (PCs) in each population (Figure 1a). Individual characteristics of all analyzed isolates are shown in Table S1 in Additional file 1 and criteria used for classification of each PC are shown in Table S2 in Additional file 1.

Altogether, 11 different PCs were distinguishable in the eight populations (Figure 1a). The extent of the diversity and the phenotypic relationships of the 11 PCs is shown in Figure 1b. Most, but not all, of the PCs were found in both slow- and fast-dilution rate cultures, although none were universally present. PC1 (ancestor class), PC3 and PC8 were present in five, six and seven cultures, respectively, making them the most widely distributed classes (Figure 1a). Some PCs were present only with a particular dilution rate; for example, PC 5 and PCs 9 to 11 were absent in fast-dilution rate cultures while all PCs except PC6 were found in slow-dilution populations. As shown in Figure 1c, on average, there were 4.75 PCs in  $0.1 \text{ h}^{-1}$  cultures compared to 3.0 in  $0.6 \text{ h}^{-1}$  cultures, which suggests that the slow-dilution cultures became more diverse than the fast-dilution cultures, despite dividing for one-sixth of the number of bulk generations but with higher mutation rates. Simpson's diversity indices ( $D_i$ ) also indicated that populations from the slow-dilution cultures were more diverse ( $D_i = 0.0282$ ) compared to those from fast-dilution cultures ( $D_i = 0.0134$ ) (Figure 1d; two-tailed  $t$ -test  $P = 0.0329$ ). Hence, both the types of PC and the overall diversity in evolution over the same elapsed time are influenced by small differences in evolutionary selection conditions, in this case residual glucose concentration determining growth rate in the chemostat.

In addition, the proportion of individual PCs varied between slow and fast dilution populations. PCs 5 and 10 are the most abundant classes in the slow-dilution populations while PCs 8 and 6 accounted for more than 50% of the total isolates analyzed from the fast-dilution populations. Some low-abundance types were unique to one population (for example, PC11 in population 1 (Figure 1a)). Some of these differences may be due to a focus on a single time point or limited sampling. However, in the one population for which we analyzed the population structure over a greater period (over 37

days) we found slightly changing proportions beyond day 27 due to continuing population turnover, but no major sweeps (Figure S1 in Additional file 1). Overall, there appears to be no sympatric or allopatric uniformity between the populations when subject to a similar, relatively homogeneous, constrained selection condition. In common with studies on another experimental system [27], unparallel diversification is an important feature of bacterial evolution.

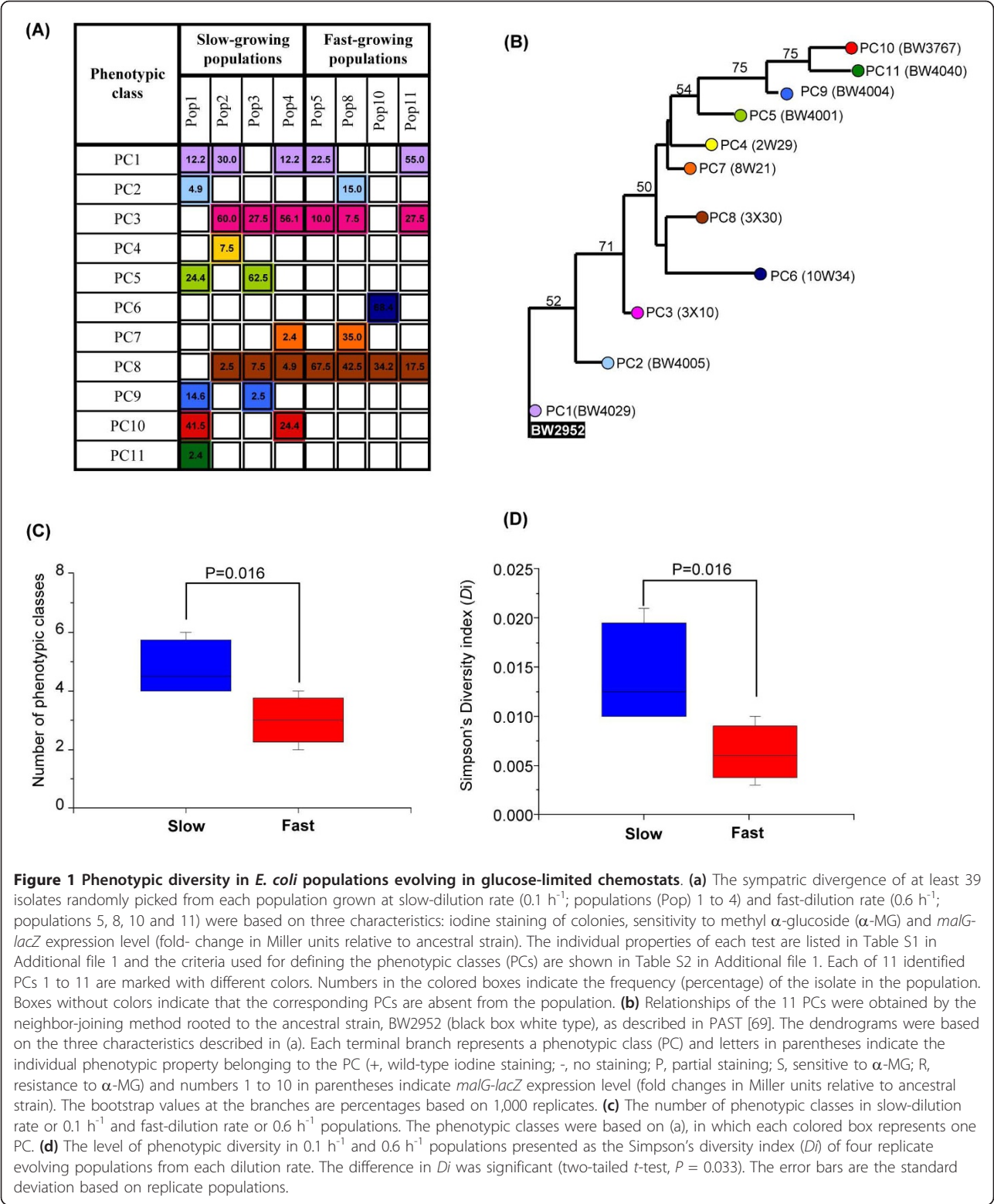
### Mutations in evolved genomes

To investigate the genetic basis of phenotypic diversity and which mechanism(s) were responsible for the emergence of diversity, we sequenced the whole genomes of at least one isolate from each PC in all eight populations. Altogether we sequenced 48 isolates, 27 from slow- and 21 from fast-dilution cultures (the strains are listed in Table S3 in Additional file 1). The fully sequenced ancestral genome [36] was used as the template for identifying the mutations in the evolved isolates.

We detected a total of 193 mutational events in the 48 evolved isolates (see Figure 2 for a graphic summary and Table S4 in Additional file 1 for a complete list of all identified mutations). There were a total of 155 SNPs, of which 140 were located in protein coding regions and the remaining 21 in intergenic regions. There were also 15 insertions and 23 deletions (indels). The majority of evolved isolates contained one to five mutations except in four isolates (BW3767, BW4004 and BW4040 from population 1 and 10W34 from population 10) that contained 28 to 46 mutations. This greatly elevated rate of genomic mutation was a result of *mutY* gene deletions as all four genomes contained a deletion of approximately 28 kb DNA including *mutY*, consistent with the demonstrated insertion sequence (IS) $\phi$ -mediated rearrangement involved in *mutY* deletions [37]. Lack of *mutY* activity causes C:G > A:T transversions [38] and almost all SNPs in these isolates showed a C:G > A:T signature. To test the overall presence of mutators, we screened all the 321 isolates for the *mutY*-related mutator phenotype by testing their ability to grow on glycolate (the deleted region in all *mutY*-mutator isolates contained the *glc* locus responsible for glycolate utilization [37]) and compared the frequency of spontaneous rifampicin-resistant mutants. We found an additional 20 and 15 isolates with *mutY*-mutator phenotypes from populations 1 and 10, respectively. No isolates with elevated mutation rates were detected in the other six populations.

### Mutational patterns in slow and fast growing cultures

Figure 2 and Additional file 2 indicate that the mutational patterns under the two selection conditions were



not identical. In the 44 non-mutator isolates, we found 27 SNPs (22 in coding and 2 in intergenic regions) and 23 SNPs (17 in coding and 6 in intergenic regions) in  $0.1\text{ h}^{-1}$  and  $0.6\text{ h}^{-1}$  populations, respectively (Table 1). All SNPs except one in a fast-growth culture (*metB* in 11Za13; Table S4 in Additional file 1) were non-

**Table 1 Quantification of mutations in slow- and fast growing cultures**

	SNPs		Indels		SNP/indel ratio	
	0.1 h <sup>-1</sup>	0.6 h <sup>-1</sup>	0.1 h <sup>-1</sup>	0.6 h <sup>-1</sup>	0.1 h <sup>-1</sup>	0.6 h <sup>-1</sup>
Coding	22	17	5	10	4.40	1.70
Intergenic	5	6	8	10	0.63	0.60
Total	27	23	13	20	2.08	1.15

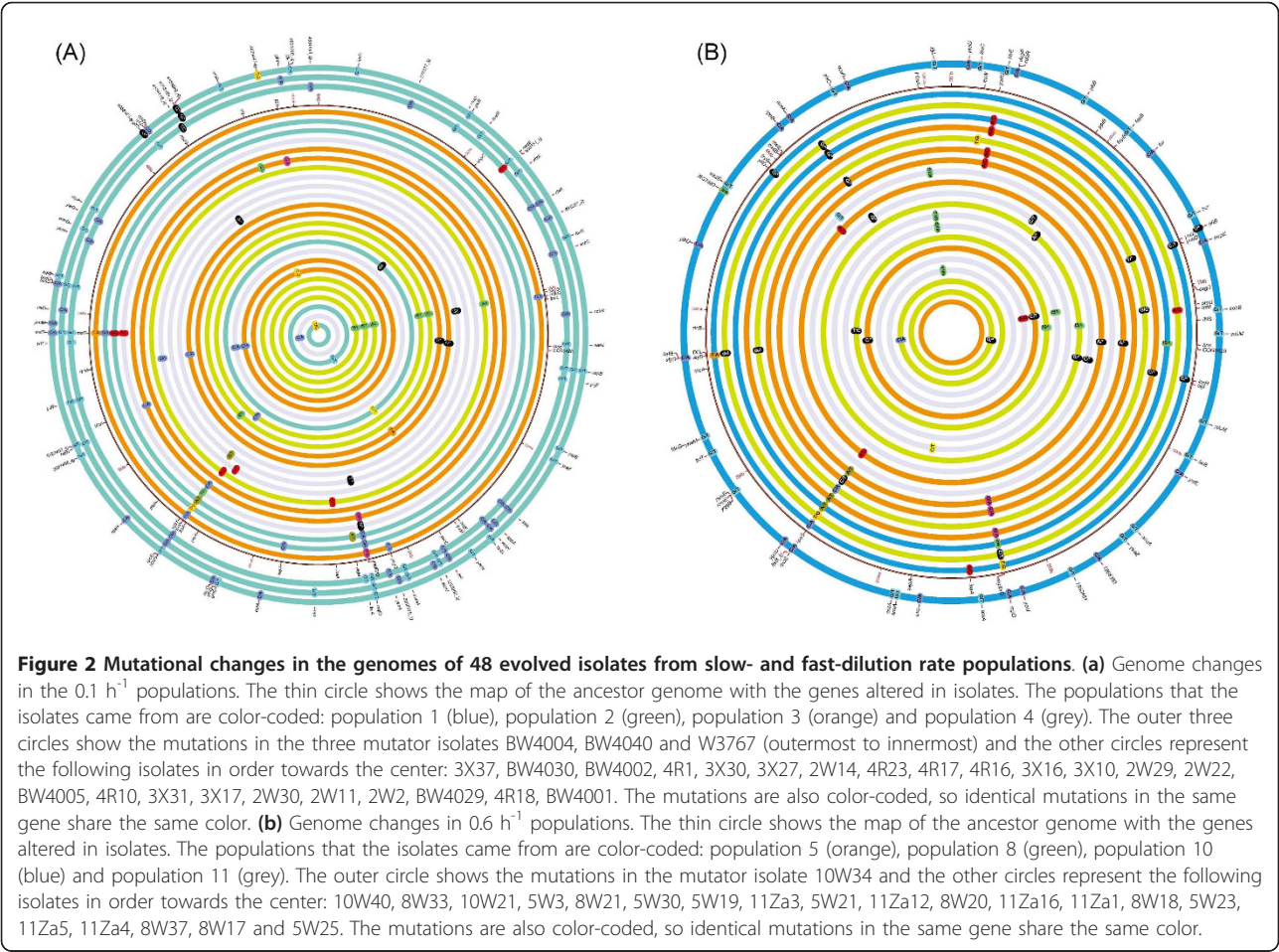
Note: only mutations from non-mutators are used for comparison.

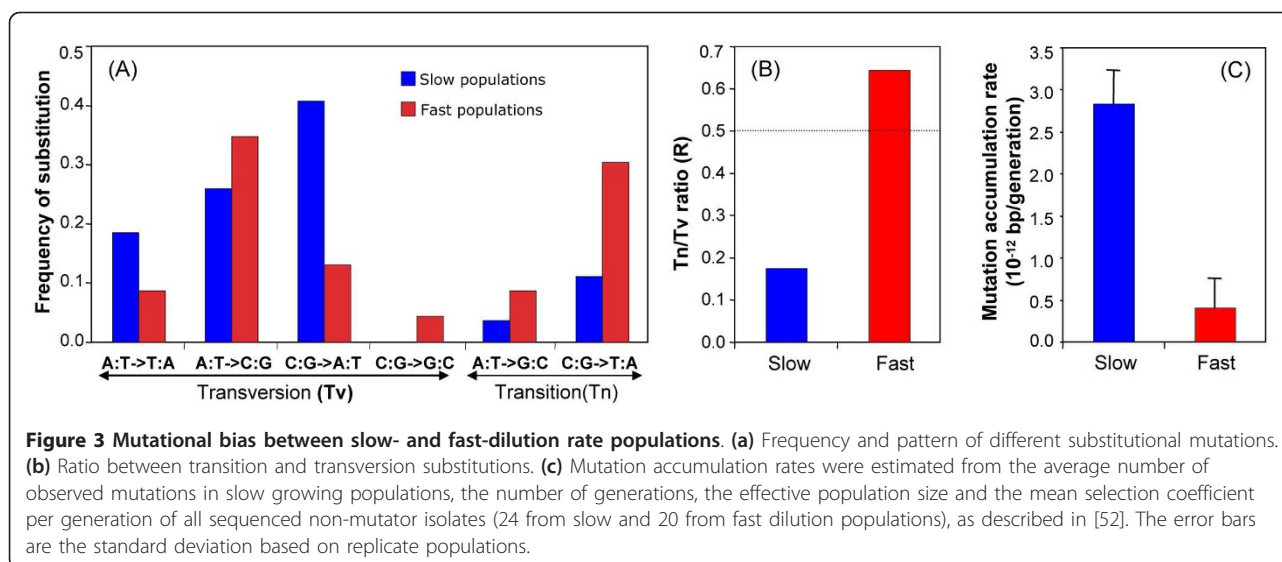
synonymous (nsSNPs), affecting polypeptide sequence. We used the SIFT algorithm [39] to evaluate the effect of nsSNPs on protein function. All of them were predicted to be non-conservative and therefore expected to exert some effect on protein function. This result is consistent with the predominance of nsSNPs in the recent studies of the *E. coli* populations evolving under different experimental settings [21,40,41], suggesting that most mutations fixed in the early phase of evolution are adaptive.

About 85% and 70% of all SNPs in the 0.1 h<sup>-1</sup> and 0.6 h<sup>-1</sup> populations, respectively, were transversions. When analyzing transitions and transversion events, the ratio

was found to be different between the 0.1 h<sup>-1</sup> and 0.6 h<sup>-1</sup> populations (Figure 3a, b). The ratio was 0.6 in the 0.6 h<sup>-1</sup> population, which was slightly higher than the cumulative ratio obtained from sequence comparison of bacterial genomes [42]. By contrast, the ratio of 0.17 in slow-growth populations was nearly three-fold less than the expected ratio of 0.5, indicating that transversion mutations are more common under slow-growing environments; the reason for this skewed profile, together with the higher mutation rates in 0.1 h<sup>-1</sup> cultures, needs investigating. However, in line with the Ochman's analysis [42], we observed only one C:G→G:C transversion mutation in all of the 48 genomes, confirming the rarity of this mutation.

We also found strikingly different patterns of indel mutations between the 0.1 h<sup>-1</sup> and 0.6 h<sup>-1</sup> populations. Among the 31 indels in non-mutator isolates, 12 were single-base while the remaining 19 were large indels. Interestingly, 10 out of 12 single-base indels were in the 0.6 h<sup>-1</sup> population. The majority of the large indels (> 1 bp) were associated with IS elements, consistent with the extensive movement of a variety of IS elements





demonstrated recently [37]. The SNP-to-indel ratio of 2.08 in 0.1 h<sup>-1</sup> populations was nearly two-fold higher than that of 1.15 in 0.6 h<sup>-1</sup> cultures, which reflects that indels are more common in 0.6 h<sup>-1</sup> populations (Table 1). However, in comparison to other bacteria, the observed SNP-to-indel ratio is significantly lower under both selection conditions, suggesting the importance of indels during constant selection under nutrient limitations. The genome-wide SNP-to-indel ratio in bacterial genome comparisons is estimated at around 19.61 [43].

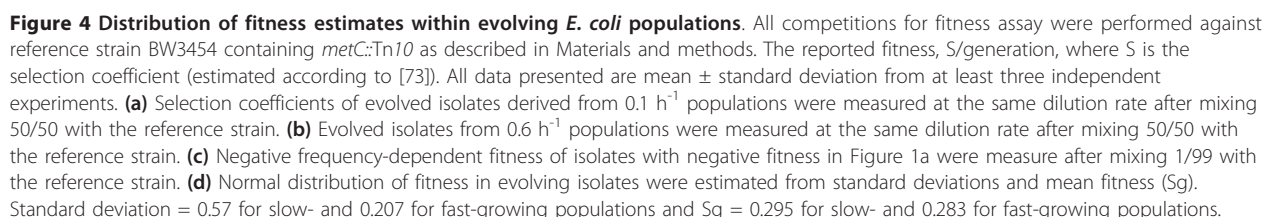
In addition to differences in mutational patterns, mutation accumulation rates (MARs) were also found to be affected by the growth rate. As shown in Figure 3c, there was approximately a four-fold higher MAR in 0.1 h<sup>-1</sup> than in 0.6 h<sup>-1</sup> populations (two tail *t*-test *P* = 0.009). The mean MAR in 0.1 h<sup>-1</sup> populations was  $1.78 \times 10^{-12}$  mutations per site per generation compared to  $0.40 \times 10^{-12}$  mutations per generation in 0.6 h<sup>-1</sup> populations (Figure 3c). These values cannot be compared to real mutation rates, which are approximately 100-fold lower [41,44,45]. Of course, MAR is affected by selection as well as mutation rate as mutation supply is probably not limiting in large populations [46].

#### Fitness properties of evolving isolates

The nature of the fitness can indicate the means of selection of particular isolates. Thus, identifying the fitness properties of evolved isolates is crucial for understanding the mechanism of adaptation. To estimate the fitness, we measured the selection coefficient of each of the 48 isolates used for genome sequencing by competing head-to-head at a 50:50 initial ratio with the ancestral isolate under the dilution conditions under which they were isolated (see Materials and methods). We showed previously that isolate-to-isolate

competition gives a similar rank order of fitness characteristics as isolate-to-ancestor competition [6]. As shown in Figure 4a, b, the majority, but not all, of the isolates from both slow and fast-dilution cultures have a significant competitive advantage against the common ancestor in their respective original selection environments.

Perhaps most interestingly, isolates with significantly reduced fitness compared to the ancestral isolate were present in each of the slow growth populations; these include BW4005 from population 1 and six newly identified isolates, 2W2 from population 2, 3X16 and 3X31 from population 3 and 4R18 and 4R10 from population 4. These isolates may have survived through adaptation to a novel niche created within the original glucose-limited environment and/or negative frequency dependence. In fact, when pair-wise competitions were started at 1% against 99% ancestral isolate, we found that these isolates exhibited a positive fitness against the ancestral clone (Figure 4c). Hence, negative frequency-dependent selection is a feature of several evolving populations, but not all isolates show this property. Interestingly, 2W2, 3X16, 3X31, 4R10, and 4R18 all belonged to PC3 while BW4005 belonged to PC2, indicating that frequency-dependence is not always associated with the same phenotype. There is also diversity in genotypes; as shown in Table S4 in Additional file 1, the frequency-dependent isolates had a variety of mutations. For example, 2W2 contained mutations associated with *csgD*, 3X16, 3X31, 4R10, and 4R18 contained a SNP in *rpoA* (RNA polymerase A) and BW4005 contained an in-frame deletion of 3 bp in *aphC* and *lpxM* and a deletion of *ogrK*-to-*yegS*, suggesting convergence of phenotype through multiple pathways. The molecular mechanism(s) of negative frequency-dependent fitness in these populations has



ancestral isolate. The measured fitness (1.29 S/g) was over two-fold higher than that of the most-fit counterparts in the 0.6 h<sup>-1</sup> populations, suggesting the fitness peak heights are very different under the two conditions.

Consistent with the phenotypic diversity, the 0.1 h<sup>-1</sup> populations exhibited a greater level of fitness diversity than the 0.6 h<sup>-1</sup> populations. The standard deviation of measured fitness amongst the 26 0.1 h<sup>-1</sup> isolates (0.53 S/g) was more than two-fold higher than the 0.21 S/g value of the 0.6 h<sup>-1</sup> populations. The mean fitness gain of 0.36 S/g in the 0.1 h<sup>-1</sup> populations was slightly higher

The data in Figure 4a, b also showed that both the magnitude of fitness gain and the extent of fitness distributions within and across the populations were markedly different between 0.1 h<sup>-1</sup> and 0.6 h<sup>-1</sup> populations. Amongst the 48 isolates, the PC11 isolate (BW4040), which contained 29 mutations, including in *rpoS*, *mgdD* and *malT*, exhibited the highest fitness relative to the

that that of 0.3 S/g in the 0.6 h<sup>-1</sup> populations. The difference, however, was not significantly different (two-tailed *t*-test *P* = 0.09). Furthermore, the normal or Gaussian distribution of fitness based on the standard deviation and the mean fitness showed that the width of fitness distributions in the 0.1 h<sup>-1</sup> populations is broader than in the 0.6 h<sup>-1</sup> populations (Figure 4d). These results are notable, particularly in the 0.1 h<sup>-1</sup> populations, because the mutation-selection balance proposal (analyzed below in detail) predicts that elimination of less fit isolates through clonal interference depends on the fitness distribution between contending lineages. However, the observed wide distribution of fitness within the 0.1 h<sup>-1</sup> populations did not lead to clonal replacement; instead we observed expansion of some less-fit subpopulations in the later samples. For example, we found that isolates from PC9 to PC11 in population 1 have the highest measured fitness but failed to replace members of the PC5 subpopulation (Figure S1 in Additional file 1). In fact, the proportion of PC5 isolates gradually increased from 5% of the total population at day 17 to over 50% at day 37 despite their fitness deficit against the isolates belonging to PC9 to PC11.

These results together with genome and phenotypic data also revealed that, in some cases, fitness convergence was achieved despite having no genotypic or phenotypic similarity. For example, BW4001 and BW4029 in population 1, 2W14 and 2W29 in population 2, 3X10 and 3X17 in population 3, 8W33 and 8W21 in population 8 and 10W21 and 10W34 in population 10 were different at both phenotypic and genotypic levels (Figure 4; Table S4 in Additional file 1) but had similar competitive fitness. These results clearly indicate that it is possible for coexisting subpopulations to achieve the same height of fitness peak through multiple mutational pathways.

#### Parallel mutations and genetic basis of the biased phenotypic diversity between 0.1 and 0.6 h<sup>-1</sup> populations

To gain further insight into the genetic basis of the phenotypic diversity and to understand which divergence mechanism is responsible for the observed diversity, we identified genes that acquired mutations in parallel populations. Genes repeatedly mutated in more than one population are more likely to contribute fitness. The 193 mutations in isolates were distributed in 165 genes or regions, of which 12 genes were found in more than one population (Table 2). Multiple alleles were found in the majority of repeatedly mutated genes and in almost all cases, the evolved alleles were different between populations. Mutations in *rpoS* and *mgdD/O* were the most common, both appearing in six out of eight populations, three each in the 0.1 and 0.6 h<sup>-1</sup> populations. Other genes mutated under both selection

conditions include *csgD-ig*, *cobB*, *inaA-ig*, *xapA*, *rpoB* and *rpoA* as well as an approximately 29 kb DNA deletion including *mutY* (Table 2).

The most striking specific parallel adaptation was a nsSNP in the *rpoA* gene encoding RNA polymerase subunit A. The same base change at the same position was present in three separate populations. The mutation affects a residue in the  $\alpha$ -CTD domain of RpoA that is involved in a multitude of regulatory interactions with DNA and transcription factors [47]. This mutation reduces RpoS levels slightly (Figure S2 in Additional file 1) so may have similar benefits to other mutations, such as in *hfq*, that reduce RpoS levels [48]. However, the effect of the *rpoA* mutation, as well as that in *rpoB*, may be on transcription more generally and needs deeper investigation.

There were also some genes that were selected in only one of the two selection conditions. For example, *malT* and *hfq* were selected only in 0.1 h<sup>-1</sup> cultures, while the *slyD* gene encoding an FKBP-type peptidyl prolyl *cis-trans* isomerase was selected in three out of four 0.6 h<sup>-1</sup> cultures but absent in slow-growth populations. The significance of *slyD* under this selection condition has not been investigated and future studies will need to investigate these and other less understood mutations, as in *inaA* (which may affect a hypothetical protein BWG-2010) and *rpoA*.

Mutations in *malT* cause elevated levels (7- to 15-fold relative to ancestor) of expression of maltose transporter genes *malG* and *lamB* [49] and, consistent with a previous study [35], mutations in *malT* were observed in three out of four populations. We did not find any isolates with high levels of *malG-lacZ* activity in the 160 isolates from 0.6 h<sup>-1</sup> cultures. This finding confirms an earlier observation that a *malT*-constitutive mutation is deleterious at high dilution rates [35]. Previously, we showed that a mutation in *hfq* was responsible for the multiple benefits under glucose limitation through changing at least five regulation targets [48]. Mutations in *hfq* exhibit a characteristic phenotype of increased sensitivity to methyl  $\alpha$ -glucoside ( $\alpha$ -MG) and a slightly elevated level of *malG-lacZ* activity [48].  $\alpha$ -MG sensitive isolates were found in six out of the eight surveyed populations, although the abundance of such isolates varied widely between populations (Table S1 in Additional file 1). To test if *hfq* mutations were responsible for  $\alpha$ -MG sensitivity, we sequenced *hfq*, including its transcription region, in all  $\alpha$ -MG sensitive isolates. We found an additional 19 *hfq* isolates, but none of them were from 0.6 h<sup>-1</sup> cultures. The  $\alpha$ -MG-sensitive mutants in fast dilution rate populations were not altered in *hfq* and were probably *mlc* or *ptsG* mutations [35]. The mutations in *hfq* were beneficial at slow growth rates but deleterious at the fast growth rate, explaining why

**Table 2 Genes or regions with mutation(s) in parallel populations**

Gene or region	Product	Total number of alleles	Selection conditions	
			0.1 h <sup>-1</sup>	0.6 h <sup>-1</sup>
<i>rpoS</i>	RNA polymerase sigma subunit	12	3 out of 4	3 out of 4
<i>mglD/O</i>	DNA-binding transcriptional regulator	9	3 out of 4	3 out of 4
<i>hfq</i>	RNA-binding protein Hfq	4	2 out of 4	None
<i>cgsD_ig</i>	DNA-binding transcriptional regulator	4	1 out of 4	2 out of 4
<i>cobB*</i>	Deacetylase of acetyl-CoA synthetase, NAD-dependent	2	1 out of 4	1 out of 4
28 to 30 kb with <i>mutY</i> deletion*	Not applicable	2	1 out of 4	1 out of 4
<i>rpoB*</i>	DNA-directed RNA polymerase subunit beta 4068948:4072976 forward	2	1 out of 4	1 out of 4
<i>malT</i>	DNA-binding transcriptional activator/maltotriose-ATP-binding protein	3	3 out of 4	None
<i>inaA_ig</i>	Hypothetical protein BWG_2010	1	1 out of 4	1 out of 4
<i>rpoA</i>	RNA polymerase, alpha subunit	1	2 out of 4	1 out of 4
<i>slyD</i>	FKBP-type peptidyl prolyl <i>cis-trans</i> isomerase (rotamase)	4	None	3 out of 4
<i>xapA</i>	Purine nucleoside phosphorylase II	2	1 out of 4	1 out of 4

\*Present only in *mutY* mutator isolates.

they were not found in 0.6 h<sup>-1</sup> cultures [50]. This bias in selection of pleiotropic *malT* and *hfq* could be a contributing factor to the observed difference in phenotypic diversity between 0.1 h<sup>-1</sup> and 0.6 h<sup>-1</sup> populations.

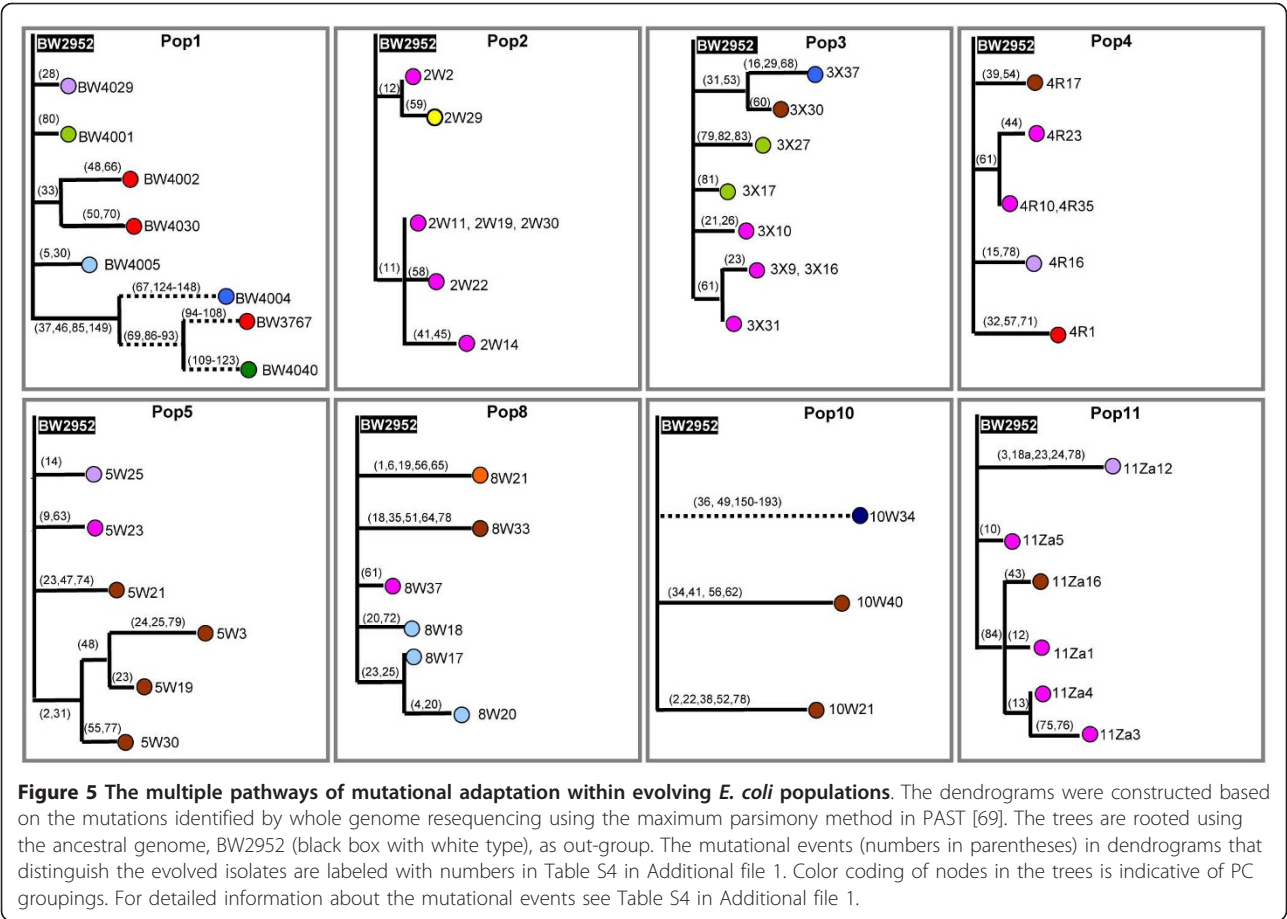
#### A mechanism of mutational divergence and adaptive convergence

Having determined the diverse phenotype, genotype and fitness properties of isolates, we next investigated whether the underlying mechanism(s) leading to the observed diversity could be identified. A recent mutation-selection balance model [16,51] suggested that adaptive mutations with small fitness effects are able to fix only in combination with those of larger effects; independent mutations with small effects were thought to be suppressed as a result of clonal interference, particularly in large populations [52]. Applying this multiple mutation model, one would expect that later-arising isolates should contain beneficial mutations on top of previously accumulated beneficial mutations. By contrast, the convergent evolution mechanism proposed that multiple mutational pathways impacting on the same physiology could generate and maintain diversity within the evolving populations [7,17].

We tested these notions in two ways. Firstly, if mutation-selection balance is important, all the evolved isolates within the same population should have originated from the most-fit initial clone. As found earlier, frequently the most-fit isolates harbored mutations in *rpoS*, which according to the previous studies using the same population, initially were present in all eight populations [35]. Therefore, we expected all evolved isolates to contain an *rpoS* mutation. However, contrary to the model,

our maximum likelihood phylogenetic analysis showed that multiple lineages emerged independently in all populations and coexisted with the *rpoS* subpopulations (Figure 5). For example, mutations in *rpoA*, *hfq* and *rssB* occurred independently in the ancestor in population 3, separately from the *rpoS* isolates. In all eight populations, multiple beneficial mutations routinely accumulated in competing lineages, rarely resulting in the complete clonal replacement of subpopulations. These results suggest clonal interference is not strong enough to prevent eventual accumulation of competitors, as was also found in evolving asexual yeast populations [53].

Secondly, we investigated the function of mutated genes within each population to test whether our results fit with the convergent evolution proposal [7,17]. As shown in Figure 6, most of the 85 mutations in Table S4 in Additional file 1 (about 55% in 0.1 and 30% in 0.6 h<sup>-1</sup> populations) were located in genes with regulatory functions according to COG (Cluster of Orthologous Groups) functional categories. These results confirm the earlier studies by Conrad *et al.* [21] and Barrick *et al.* [41], that regulatory mutations provide a main route of mutational adaptation. Mutations in *rpoA*, *rpoB*, *csgD*, *rssB* and *fhlA* as well as previously identified mutations in regulatory genes *hfq*, *rpoS*, *mglD* (*galS*) and *malT* were the most frequent (Table 2; Table S4 in Additional file 1). In addition to controlling glucose transport, several of these genes are involved in resetting the trade-off between stress protection and nutritional competence (abbreviated as SPaNC [34]). Mutations in *csgD*, *rssB*, *cyaA*, *hfq* and *rpoS*, as well as possibly *rpoA* and *rpoB*, affect regulation of the general stress response directly or indirectly. Mutations in *rpoS* were the most common,



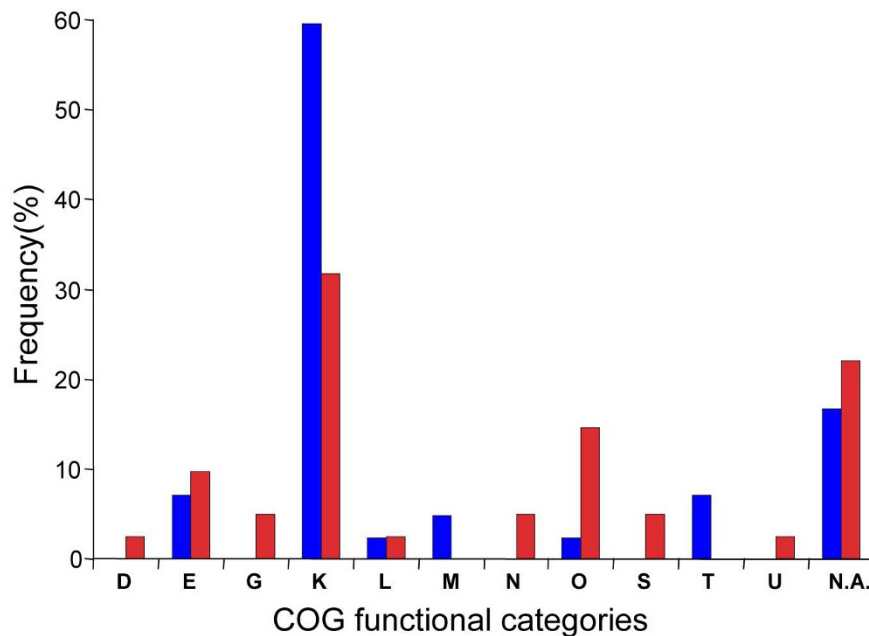
appearing in six out of eight populations. This gene directly, and the others indirectly, influence the SPaNC balance [13,34]. Hence, an altered SPaNC trade-off, which can allow co-existence of multiple types [13,34], is a common event in the eight populations.

Newly found mutations affecting the SPaNC balance include two distinct mutations, a SNP and an *IS1* insertion, in the transcribed message of *csgD* in isolates from three populations. The *csgD* gene encodes a DNA-binding transcription activator that regulates several genes involved in biofilm formation and also influences  $\sigma^S$  [54,55]. Also, in addition to a previously identified mutation in *hfq* [48], we found two further alleles of *hfq* at slow growth rates affecting translation of  $\sigma^S$ . We also found one change in *rssB* relative to the ancestor; RssB facilitates and regulates degradation of  $\sigma^S$  by the protease ClpXP under nutrient-sufficient conditions [56]. RssB can also inhibit  $\sigma^S$  function even in the absence of proteolysis, possibly via stoichiometric binding [56]. An immuno-blot analysis with an anti- $\sigma^S$  serum also indicated that these *csgD*, *rssB* and *hfq* isolates indeed contain lowered  $\sigma^S$  levels (Figure S2 in Additional file 1), suggesting that retuning of the SPaNC trade-off may be

the common consequence of the different regulatory mutations. The convergent evolution through diverged mutational pathways has also been identified in *E. coli* populations during thermal adaptation [57].

**Discussion**

A combination of phenotypic and genotypic analysis has revealed the considerable extent of diversity that develops in eight glucose-limited chemostat populations. A simple phenotypic screen revealed 11 PC groups with 10 sharing similar properties across more than one population. The genomic diversity was even greater, because several mutational means of achieving the same phenotype were possible within PC groups. Only 2 out of 48 isolates had the same combination of mutations, revealing the extent of diversity present in glucose-limited *E. coli* populations. Even strains from the same PC group and the same population generally contained many mutational differences; our limited sampling means we have not covered all possibilities. We have identified many new mutations in addition to those that have been identified in earlier studies of glucose-limited populations (*rpoS*, *ptsG*, *mglD*, *malT*, *hfq*) [48,58].



**Figure 6 Functional changes in 85 mutations identified in 44 non-mutator isolates from 4 slow-growing cultures and 4 fast growing cultures.** Groupings of genes were based on COGs (Clusters of Orthologous Groups) [74]. COG codes: D, cell cycle control and mitosis; E, amino acid metabolism and transport; G, carbohydrate metabolism and transport; K, transcription; L, replication and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, post-translational modification, protein turnover, chaperone functions; S, function unknown; T, signal transduction; V, intracellular trafficking and secretion; N.A., not analyzed.

Especially intriguing are potential global regulatory mutations in *rpoA*, *csgD*, and *rssB* as well as mutations in *lpx* genes, *fimA*, *frvB* and *ptsP* affecting membranes and surface functions. These mutations will be individually studied for fitness and physiological effects in future studies. The main emphasis in this discussion though is to discuss the impact of these results on the previously identified diversification mechanisms and then the novel differences in evolutionary paths between bacteria growing at two growth rates, especially to explain the greater phenotypic diversity at the slower growth rates.

#### How the new data fit with mechanisms explaining diversity and evolutionary coexistence

Some mechanisms of diversification involve ecological interactions exhibited as frequency-dependence [11]. This includes niche creation [11,12] and frequency-dependent coexistence due to cross-feeding, where the cross-feeder organism is maintained in balance with the producer [59], or through cannibalism of one or the other coexisting type [28]. Another frequency-dependent scenario is when two or more resources in a system result in specialization of different bacteria on different resources in the same environment [11,60]. None of these ecological scenarios are readily extended to the current studies. Cross-feeding and cannibalism do not

appear to have evolved in the chemostats, probably because the time-scale is relatively short. None of the sequenced mutations hint at cross-feeding, such as the *acs* mutation noted in [8]. There is no evidence for multiple resources, multiple niches or mutations enhancing uptake or regulation of alternative carbon sources besides glucose. Nevertheless, isolates that are less fit than the ancestor at a 50:50 population ratio do show negative frequency dependence, such as BW4005. This PC class is a minority in the 0.1 h<sup>-1</sup> populations and absent from the 0.6 h<sup>-1</sup> populations, so frequency-dependent selection is present. At the moment we have no ecological or mechanistic explanation of the frequency dependence or the mutations in BW4005. Likewise, the mutations in 2W2, 3X16, 3X31, 4R10, and 4R18 do not indicate the frequency-dependence mechanism responsible. Perhaps ecological interactions are not essential for frequency dependence, which may be a manifestation of internal metabolic or regulatory differences. Recent studies suggest the appearance of frequency-dependence can be due to complex fitness-growth rate relationships determined by altered patterns of gene expression [50]. Further investigation is needed to determine whether the mutations in the frequency-dependent isolates show non-linear fitness-growth rate relationships.

The diversity in an evolving population is dependent on the rate of appearance of mutations and the elimination of mutations through selection of the fittest types. Mutation-selection balance is therefore a potential contributor to the coexistence of many types [15,16,51]. The genomics indicates that in a large population such as ours with  $10^{10}$  bacteria, multiple mutational lineages arise directly from the ancestor; *hfq*, *rpoA*, *rpoS* mutations and other co-existing mutations in various populations do not appear on a sequential basis and clonal inference is not sufficient to eliminate the contending lineages. This is in agreement with the model for events in a large population [53,61]. If mutation-selection balance drives diversity, the mutation rate difference between slow- and fast-growing populations should contribute to greater diversity at  $0.1 \text{ h}^{-1}$ . A difference in diversity was indeed seen, but the problem is that the fitness properties of the isolates are also different. An even more dramatic case of altered mutation supply is with the *mutY* mutator sub-populations in two populations, with ten times as many mutations as the rest of the isolates. Contrary to the expectation that these high mutation rate lineages should diversify more readily by mutation selection balance, we do not see this result. Indeed, in time points subsequent to the one studied here, the mutator sub-populations do not diversify or take over populations (Figure S1 in Additional file 1). So overall, firm conclusions on mutation-selection balancing as a factor in chemostat populations are not possible, but it is unlikely to be the dominant contributor to the observed divergence.

Trade-offs in metabolism [33,62], rate-yield [63,64], and stress-nutrition [13,34] have all been postulated to lead to alternative evolutionary solutions, even in uniform environments. More recently, combinations of trade-offs have been modeled [10] to explain the level of diversity seen in chemostat populations. Several trade-offs indeed change in chemostat evolution and result in isolates with altered stress resistance-nutrition properties as well as changes in growth rate and yield [33,65]. Even more recently, trade-offs with differences in mutational robustness rate have also been postulated to lead to diversity in evolving populations [14]. Since all of the proposed ingredients (trade-offs, mutation rate differences) are present in chemostat populations as demonstrated in this study, our results are not inconsistent with these proposals. A mathematical model of the SPaNC trade-off, which is being altered by mutation in these populations, does suggest that coexistence of types with distinct settings of the SPaNC balance can be demonstrated [13]. The various *hfq*, *rssB*, *rpoS* and possibly *rpoA* mutants do offer alternative settings of the SPaNC balance, so may contribute to heterogeneity by providing alternative SPaNC settings. The

characterization of the settings and experimental coexistence studies require further work, but trade-offs do offer possibilities for heterogeneity in the eight populations. This is not a blanket explanation, however, because there are isolates in all populations that do not have changes to RpoS levels and SPaNC balances. Here again, this is a partial explanation of diversity.

Finally, a simple explanation of coexistence is the convergent evolution of genetically distinct mutants that have achieved a comparable level of fitness. A particular level of fitness (as measured against ancestor) could be reached in strains having entirely different genomic changes (Figure 4). Bacteria exhibiting alternative regulatory and metabolic pathways and mutational degeneracy in these pathways can lead to co-existing organisms in nutrient-limited chemostat populations [7,17,33,58]. As already noted previously, mutations in *hfq* and *rpoS* can both reset the stress-nutrition trade-off towards better nutrient uptake [17]. Furthermore, in line with a recent study [57], the presence of mutations in *rpoA*, *csgD*, *rssB* and *fimA*, each in different members of the same PC group, indicates that there are degenerate means of evolving similar properties within a complex regulatory network. These changes can also overlap with the trade-offs discussed above, so alternative SPaNC settings can be arrived at in degenerate ways.

#### The effects of a difference in growth rate in the selection environment

It is instructive to compare adaptational events at two chemostat dilution rates but with otherwise identical ancestors, conditions and sole resources in the shape of glucose. As evident in PCs and with mutations, an adaptive advantage at one growth rate does not necessarily provide an advantage under both growth conditions. There was already evidence that *malT*-con mutations are deleterious at faster growth rates [35], and this study confirms this, as no such mutations were selected in any of the fast-growth populations. It is not clear why *malT* mutations are deleterious at faster growth rates, but the finding of *malT*-negative mutations in batch culture selection with glucose [30] suggests that there is a negative effect of MalT on fitness when growing fast on glucose.

Another example of a gene with mutations beneficial and selected at  $0.1 \text{ h}^{-1}$  but not at  $0.6 \text{ h}^{-1}$  is *hfq*. All three evolved alleles of this gene provided a benefit at  $0.1 \text{ h}^{-1}$ , which occurs through multiple pleiotropic effects [48]. The mutations were found to be deleterious at  $0.6 \text{ h}^{-1}$  [50], explaining their absence from  $0.6 \text{ h}^{-1}$  cultures. This fitness difference is likely to be because global gene expression alters with different dilution rates [66] and the role of Hfq changes. We postulate that *hfq* mutations, through numerous epistatic effects, are beneficial

when cells are more highly stressed (at  $0.1 \text{ h}^{-1}$ ) than when growing faster (at  $0.6 \text{ h}^{-1}$ ).

Fitness in evolved isolates is by no means uniform in coexisting types or at different growth rates. The fitness distributions as well as the diversity differ between slow- and fast-growth populations. The high proportion of regulatory mutations observed in the  $0.1 \text{ h}^{-1}$  but not the  $0.6 \text{ h}^{-1}$  population (Figure 6) is consistent with a different global-regulatory background and demonstrates differences in gene expression between the two conditions [66]. The greater phenotypic diversity in the  $0.1 \text{ h}^{-1}$  cultures may well reflect a greater range of regulatory mutations observed at the slower growth rate.

## Conclusions

*E. coli* has at its disposal a redundant set of fitness strategies under glucose limitation involving various combinations of mutations. These permit concurrent evolution of distinct lineages within the same population. The simplest explanation is that fitness arose through multiple lineages of beneficial mutations that did not fully eliminate competitors and invariably led to diversification within the eight clonal cultures. The kind of mutants and level of diversity were influenced by the growth rate of cultures. In total, our results suggest that large bacterial populations are a diverse collection of clonal lineages in most settings, complex or otherwise. Therefore, it is likely that in a natural habitat with a large population size, bacteria develop enormous sympatric variation. This is likely to be of considerable importance in clinical populations of bacteria that exhibit diversification within the lifetime of an infection [3].

## Materials and methods

### Bacterial strains and growth conditions

For the experimental evolution, the *E. coli* K-12 strain BW2952, an MC4100 derivative, was propagated at growth rates (dilution rates) of  $0.1 \text{ h}^{-1}$  and  $0.6 \text{ h}^{-1}$  in a glucose-limited chemostat exactly as described in [35]. Individuals from the culture were isolated and propagated as in [6]. At least 39 randomly picked isolates (colonies) from days specified in the text were obtained by streaking chemostat samples on nutrient agar.

### Phenotypic assays, classification of phenotypic properties, cluster analysis

The changes in chemostat isolates resulting in *rpoS* mutations were detected by iodine staining as previously described [67] and confirmed by growth tests on acetate plates [65]. The method for assaying *malG-lacZ* fusion  $\beta$ -galactosidase activity for *mal* regulation was described in [67]. The growth conditions for the sensitivity assay for methyl  $\alpha$ -glucoside ( $\alpha$ -MG) were as described in

[68]. Sensitive isolates had an inhibition zone of over 2.5 cm compared to the resistant ancestral strain BW2952.

We then used these triple-phenotypic test results to classify evolved isolates using the criteria shown in Table S2 in Additional file 1. We related differences in the phenotypic properties of *E. coli* chemostat isolates by using the neighbor-joining method of tree construction based on distance estimated by using PAST [69].

### Genome re-sequencing

Chromosomal DNA was prepared as previously described [36]. Libraries for Illumina paired-end sequencing were prepared using the protocols provided by Illumina. A Solexa Genome Analyzer IIX (Illumina, Little Chesterford, Essex, UK) was used to sequence each isolate to a depth of between 96- and 159-fold coverage. The 101 bp Solexa reads generated, after removal of the duplication, were mapped to the *E. coli* BW2952 genome [GenBank:NC\_012759] to generate the assembly using BWA [70] with the default parameter, which allows 4% mismatches. The Smith-Waterman algorithm was used to rescue the unmapped end. All the reads with extremely large or small insert size ( $< 20\%$  or  $> 200\%$  of normal) were mapped again using BLASTn with an e-value of 0.00001 and the -F F flag. Only the read pairs mapped to a proper insert length and with at least one end of the pairs mapped to non-repeat regions were taken into account. For positions in doubt, due to sequence coverage reduction and in repeat regions, Sanger sequencing was used to resolve the sequence. SAMtools [71] was used to calculate the per-position coverage and base calls for each position.

SNPs in the sequenced strain were called where a position was covered by at least ten reads with at least 80% of the covering reads showing the same mismatch. Other suspected positions were sequenced using Sanger methodology.

Small indels were identified from the alignment between reads and the reference genome. Cases where an indel occurred within homopolymer tracts were manually examined, as BWA and BLASTn alignment can place such indels at different positions along the homopolymer. Indels in homopolymers are not subject to elevated error or bias in Solexa sequencing [72]. Therefore, small indels could be reliably resolved by the set of individual Solexa reads spanning the homopolymer region.

Large insertion, duplication or deletion events were also detected from the alignment between reads and the reference genome. Deletions were determined as regions that were not covered by any single read and flanked by read pairs with extremely large insert size. The positions of the large insertions or duplications, such as IS

movement or copy number variation in the sequenced genome relative to the reference genome, could be detected as read pairs aligned around the insert positions that have one of their ends unaligned or aligned to other regions that represent the beginning or end of the inserted sequence.

#### Estimation of diversity index

Simpson's diversity index ( $D$ ) was determined using the following equation:

$$D_i = \sum (n/N)^2$$

where 'n' is the total number of isolates of a particular phenotype and 'N' is the total number of isolates of all phenotypes.

#### Pair-wise competition experiments and estimation of selection coefficients (S) and estimation of fitness

To estimate the selection coefficient, each evolved isolate was competed with the *metC::Tn10* derivative of the ancestral strain (BW2952) in the same conditions in which isolates had evolved. In each case, both evolved isolates and reference strains were grown independently overnight in glucose-limited chemostats in modified McBride agar (MMA) medium supplemented with 0.02% (w/v) glucose plus 4  $\mu$ g/ml methionine to ensure they were in the same physiological conditions as well as reacclimatize the selection condition. The selection coefficient (S) of evolved isolates was estimated using at least three time points with  $R^2 \geq 0.95$  as previously described in [6]. The selection coefficient for evolved isolates was determined from the slope of the linear regression of  $\ln[p(t)/q(t)]$  against at least three time points [73], where  $p(t)$  and  $q(t)$  are the proportions of two competing strains at time  $t$  hours. Since, competition experiments were performed under two different dilution rates, the fitness of evolved isolates is expressed as the selection per generation,  $S_g$  (calculated as  $(\ln 2)S/D$ , where  $D$  is dilution rate) [73].

#### Cluster analysis

The dendrograms were constructed based on mutations identified by whole genome resequencing using the maximum parsimony method in PAST [69].

#### Estimation of mutation accumulation rates

MARs were estimated using the following equation as described in [52]:

$$MAR = k/(N_e \cdot 2S_g \cdot T)$$

where  $k$  is the average number observed mutations,  $T$  is the number of generations,  $N_e$  is the effective

population size and  $S_g$  is the selection coefficient per generation.

#### Data availability

The sequencing data generated in this study have been deposited at the DNA Data Bank of Japan under accession number DRA000563.

#### Additional material

**Additional file 1: Supplementary Tables S1 to S4 and Figures S1 and S2.** Table 1: phenotypic characteristics of 320 *E. coli* isolates evolving under glucose-limited limited. Table 2: groupings of phenotypic properties found in eight glucose-limited populations of *E. coli*. Table 3: clones chosen for full genome resequencing. Table 4: genomic changes detected in the 44 non-mutator evolved clones. Figure 1: maintenance of phenotypic diversity in an *E. coli* population evolving in a glucose-limited chemostat. Figure 2: relative level of RpoS protein in glucose-limited chemostats.

**Additional file 2: SNPs of 48 strains compared with BW2952.** Sheet 1: SNPs of slow-growth isolates. Sheet 2: SNPs of fast-growth isolates. Sheet 3: the SNPs of isolates with *mutY* mutator.

#### Abbreviations

$\alpha$ -MG: methyl  $\alpha$ -glucoside; bp: base pair; COG: Cluster of Orthologous Groups; IS: insertion sequence; MAR: mutation accumulation rate; nsSNP: non-synonymous SNP; PC: phenotypic class; SNP: single nucleotide polymorphism; SPANC: stress protection and nutritional competence.

#### Acknowledgements

We thank Shona Seeto for the initial characterization of evolved strains. The research was supported by a research grant from the Australian Research Council to TF and PRR. The work was also supported by the Chinese National Science Fund for Distinguished Young Scholars (30788001); the National Natural Science Foundation of China (NSFC) Key Program Grant (31030002); the Fundamental Research Funds for the Central Universities (65020121 and 65020061); the National 973 Program of China Grant (2009CB522603, 2011CB504900, 2012CB721101 and 2012CB721001); and the National Key Programs for Infectious Diseases of China (2009ZX10004-108).

#### Author details

<sup>1</sup>School of Molecular and Microbial Biosciences, University of Sydney, NSW 2006, Australia. <sup>2</sup>TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, Tianjin 300457, PR China. <sup>3</sup>Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, 23 Hongda Street, Tianjin 300457, PR China.

#### Authors' contributions

RM carried out the phenotypic studies, as well as analyses of fitness in chemostats and mutation effects. TF conceived of the study, and participated in its design and coordination and wrote the draft manuscript. PR was involved in data analysis and paper writing. YL and BL carried out sequencing and analysis. LW participated in the design of the study and coordinated the genome sequencing and data analysis. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 1 March 2012 Revised: 25 April 2012 Accepted: 8 June 2012  
Published: 8 June 2012

#### References

1. Feil EJ: Small change: keeping pace with microevolution. *Nat Rev Microbiol* 2004, **2**:483-495.

2. Wirth T, Falush D, Lan RT, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MCJ, Ochman H, Achtman M: **Sex and virulence in *Escherichia coli*: an evolutionary perspective.** *Mol Microbiol* 2006, **60**:1136-1151.
3. Levert M, Zamfir O, Clermont O, Bouvet O, Lespinats S, Hipeaux M, Branger C, Picard B, Saint-Ruf C, Norel F, Balliau T, Zivy M, Le Nagard H, Cruvellier S, Chane-Woon-Ming B, Nilsson S, Gudeli J, Phan K, Ferenci T, Tenaillon O, Denamur E: **Molecular and evolutionary bases of within-patient genotypic and phenotypic diversity in *Escherichia coli* extraintestinal infections.** *PLoS Path* 2010, **6**:e1001125.
4. Hardin G: **Competitive exclusion principle.** *Science* 1960, **131**:1292-1297.
5. Kassen R, Rainey PB: **The ecology and genetics of microbial diversity.** *Annu Rev Microbiol* 2004, **58**:207-231.
6. Maharjan R, Seeto S, Notley-McRobb L, Ferenci T: **Clonal adaptive radiation in a constant environment.** *Science* 2006, **313**:514-517.
7. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ: **The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast.** *PLoS Genet* 2008, **4**:e1000303.
8. Kinnersley MA, Holben WE, Rosenzweig F: **E Unibus Plurim: genomic analysis of an experimentally evolved polymorphism in *Escherichia coli*.** *PLoS Genet* 2009, **5**:e1000713.
9. Levin BR: **Frequency dependent selection in bacterial populations.** *Phil Trans R Soc B* 1988, **319**:459-472.
10. Doebeli M, Ispolatov I: **Complexity and diversity.** *Science* 2010, **328**:494-497.
11. Friesen ML, Saxer G, Travisano M, Doebeli M: **Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*.** *Evolution* 2004, **58**:245-260.
12. Rosenzweig RF, Sharp RR, Treves DS, Adams J: **Microbial evolution in a simple unstructured environment: Genetic differentiation in *Escherichia coli*.** *Genetics* 1994, **137**:903-917.
13. Gudeli J, Weitz JS, Ferenci T, Horner-Devine MC, Marx CJ, Meyer JR, Forde SE: **An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure.** *Ecol Lett* 2010, **13**:1073-1084.
14. Beardmore RE, Gudeli J, Lipson DA, Hurst LD: **Metabolic trade-offs and the maintenance of the fittest and the flattest.** *Nature* 2011, **472**:342-346.
15. Zhang X-S, Hill WG: **Genetic variability under mutation selection balance.** *Trends Ecol Evol* 2005, **20**:468-470.
16. Desai MM, Fisher DS: **Beneficial mutation selection balance and the effect of linkage on positive selection.** *Genetics* 2007, **176**:1759-1798.
17. Wang L, Spira B, Zhou ZM, Feng L, Maharjan RP, Li XM, Li FF, McKenzie C, Reeves PR, Ferenci T: **Divergence involving global regulatory gene mutations in an *Escherichia coli* population evolving under phosphate limitation.** *Genome Biol* 2010, **24**:78-487.
18. Conrad TM, Lewis NE, Palsson BO: **Microbial laboratory evolution in the era of genome-scale science.** *Mol Syst Biol* 2011, **7**:509-509.
19. Elena SF, Lenski RE: **Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation.** *Nat Rev Genet* 2003, **4**:457-469.
20. Barrick JE, Lenski RE: **Genome-wide mutational diversity in an evolving population of *Escherichia coli*.** *Cold Spring Harb Symp Quant Biol* 2009, **74**:119-129.
21. Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, Palsson BO: **Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations.** *Genome Biol* 2009, **10**:R118.
22. Brockhurst MA, Rainey PB, Buckling A: **The effect of spatial heterogeneity and parasites on the evolution of host diversity.** *Proc R Soc Lond Ser B-Biol Sci* 2004, **271**:107-111.
23. Kerr B, Riley MA, Feldman MW, Bohannan BJM: **Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors.** *Nature* 2002, **418**:171-174.
24. Rainey PB, Travisano M: **Adaptive radiation in a heterogeneous environment.** *Nature* 1998, **394**:69-72.
25. Barrett RDH, MacLean RC, Bell G: **Experimental evolution of *Pseudomonas fluorescens* in simple and complex environments.** *Am Nat* 2005, **166**:470-480.
26. Habets M, Rozen DE, Hoekstra RF, de Visser J: **The effect of population structure on the adaptive radiation of microbial populations evolving in spatially structured environments.** *Ecol Lett* 2006, **9**:1041-1048.
27. Tyerman J, Havard N, Saxer G, Travisano M, Doebeli M: **Unparallel diversification in bacterial microcosms.** *Proc R Soc B Biol Sci* 2005, **272**:1393-1398.
28. Rozen DE, Philippe N, Arjan de Visser J, Lenski RE, Schneider D: **Death and cannibalism in a seasonal environment facilitate bacterial coexistence.** *Ecol Lett* 2009, **12**:34-44.
29. Crozat E, Hindre T, Kuhn L, Garin J, Lenski RE, Schneider D: **Altered regulation of the OmpF porin by Fis in *Escherichia coli* during an evolution experiment and between B and K-12 strains.** *J Bacteriol* 2011, **193**:429-440.
30. Philippe N, Crozat E, Lenski RE, Schneider D: **Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*.** *Bioessays* 2007, **29**:846-860.
31. Sniegowski PD, Gerrish PJ, Lenski RE: **Evolution of high mutation rates in experimental populations of *E. coli*.** *Nature* 1997, **387**:703-705.
32. Notley-McRobb L, Seeto S, Ferenci T: **Enrichment and elimination of *mutY* mutators in *Escherichia coli* populations.** *Genetics* 2002, **162**:1055-1062.
33. Maharjan R, Seeto S, Ferenci T: **Divergence and redundancy of transport and metabolic rate-yield strategies in a single *Escherichia coli* population.** *J Bacteriol* 2007, **189**:2350-2358.
34. Ferenci T: **Maintaining a healthy SPANC balance through regulatory and mutational adaptation.** *Mol Microbiol* 2005, **57**:1-8.
35. Notley-McRobb L, Seeto S, Ferenci T: **The influence of cellular physiology on the initiation of mutational pathways in *Escherichia coli* populations.** *Proc R Soc Lond Ser B Biol Sci* 2003, **270**:843-848.
36. Ferenci T, Zhou ZM, Betteridge T, Ren Y, Liu Y, Feng L, Reeves PR, Wang L: **Genomic sequencing reveals regulatory mutations and recombinational events in the widely used MC4100 lineage of *Escherichia coli* K-12.** *J Bacteriol* 2009, **191**:4025-4029.
37. Gaffe J, McKenzie C, Maharjan RP, Coursange E, Ferenci T, Schneider D: **Insertion sequence-driven evolution of *Escherichia coli* in chemostats.** *J Mol Evol* 2011, **72**:398-412.
38. Miller JH: **Spontaneous mutators in bacteria -insights into pathways of mutagenesis and repair.** *Annu Rev Microbiol* 1996, **50**:625-643.
39. Ng PC, Henikoff S: **SIFT: predicting amino acid changes that affect protein function.** *Nucleic Acids Res* 2003, **31**:3812-3814.
40. Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BO: **Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale.** *Nat Genet* 2006, **38**:1406-1412.
41. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF: **Genome evolution and adaptation in a long-term experiment with *Escherichia coli*.** *Nature* 2009, **461**:1243-1247.
42. Ochman H: **Neutral mutations and neutral substitutions in bacterial genomes.** *Mol Biol Evol* 2003, **20**:2091-2096.
43. Chen JQ, Wu Y, Yang HW, Bergelson J, Kreitman M, Tian DC: **Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria.** *Mol Biol Evol* 2009, **26**:1523-1531.
44. Drake JW, Charlesworth B, Charlesworth D, Crow JF: **Rates of spontaneous mutation.** *Genetics* 1998, **148**:1667-1686.
45. Rozen DE, de Visser J, Gerrish PJ: **Fitness effects of fixed beneficial mutations in microbial populations.** *Curr Biol* 2002, **12**:1040-1045.
46. de Visser JAGM, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE: **Diminishing returns from mutation supply rate in asexual populations.** *Science* 1999, **283**:404-406.
47. Blatter EE, Ross W, Tang H, Gourse RL, Ebricht RH: **Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding.** *Cell* 1994, **78**:889-896.
48. Maharjan R, Zhou ZM, Ren Y, Li Y, Gaffe J, Schneider D, McKenzie C, Reeves PR, Ferenci T, Wang L: **Genomic identification of a novel mutation in *hfq* that provides multiple benefits in evolving glucose-limited populations of *Escherichia coli*.** *J Bacteriol* 2010, **192**:4517-4521.
49. Notley-McRobb L, Ferenci T: **The generation of multiple coexisting *mal*-regulatory mutations through polygenic evolution in glucose-limited populations of *Escherichia coli*.** *Environ Microbiol* 1999, **1**:45-52.
50. Maharjan R, McKenzie C, Yeung A, Ferenci T: **The basis of antagonistic pleiotropy in *hfq* mutations that have opposite effects on fitness at slow and fast growth rates.** *Heredity* 2012.

51. Desai MM, Fisher DS, Murray AW: **The speed of evolution and maintenance of variation in asexual populations.** *Curr Biol* 2007, **17**:385-394.
52. Perfeito L, Fernandes L, Mota C, Gordo I: **Adaptive mutations in bacteria: High rate and small effects.** *Science* 2007, **317**:813-815.
53. Kao KC, Sherlock G: **Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*.** *Nat Genet* 2008, **40**:1499-1504.
54. Gualdi L, Tagliabue L, Landini P: **Biofilm formation-gene expression relay system in *Escherichia coli*: Modulation of sigma(s)-dependent gene expression by the CsgD regulatory protein via sigma(s) protein stabilization.** *J Bacteriol* 2007, **189**:8034-8043.
55. White AP, Weljie AM, Apel D, Zhang P, Shaykhtudinov R, Vogel HJ, Surette MG: **A Global metabolic shift is linked to *Salmonella* multicellular development.** *PLoS ONE* 2010, **5**:e11814.
56. Becker G, Klauck E, Hengge-Aronis R: **The response regulator RssB, a recognition factor for sigma(s) proteolysis in *Escherichia coli*, can act like an anti-sigma(s) factor.** *Mol Microbiol* 2000, **35**:657-666.
57. Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS: **The molecular diversity of adaptive convergence.** *Science* 2012, **335**:457-461.
58. Ferenci T: **Bacterial physiology, regulation and mutational adaptation in a chemostat environment.** *Adv Microb Physiol* 2007, **53**:169-229.
59. Doebeli N: **A model for the evolutionary dynamics of cross-feeding polymorphisms in microorganisms.** *Popul Ecol* 2002, **44**:59-70.
60. Dykhuizen DE, Dean AM: **Evolution of specialists in an experimental microcosm.** *Genetics* 2004, **167**:2015-2026.
61. Park SC, Krug J: **Clonal interference in large populations.** *Proc Natl Acad Sci USA* 2007, **104**:18135-18140.
62. Gudelji I, Beardmore RE, Arkin SS, Maclean RC: **Constraints on microbial metabolism drive evolutionary diversification in homogeneous environments.** *J Evol Biol* 2007, **20**:1882-1889.
63. Pfeiffer T, Schuster S, Bonhoeffer S: **Cooperation and competition in the evolution of ATP-producing pathways.** *Science* 2001, **292**:504-507.
64. Novak M, Pfeiffer T, Lenski RE, Sauer U, Bonhoeffer S: **Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*.** *Am Nat* 2006, **168**:242-251.
65. King T, Ishihama A, Kori A, Ferenci T: **A regulatory trade-off as a source of strain variation in the species *Escherichia coli*.** *J Bacteriol* 2004, **186**:5614-5620.
66. Ferenci T: **Regulation by nutrient limitation.** *Curr Opin Microbiol* 1999, **2**:208-213.
67. Notley L, Ferenci T: **Differential expression of *mal* genes under cAMP and endogenous inducer control in nutrient stressed *Escherichia coli*.** *Mol Microbiol* 1995, **16**:121-129.
68. Notley-McRobb L, Ferenci T: **Substrate specificity and signal transduction pathways in the glucose-specific enzyme II (EIIGlc) component of the *Escherichia coli* phosphotransferase system.** *J Bacteriol* 2000, **182**:4437-4442.
69. Hammer O, Harper DAT, Ryan PD: **PAST: Palaeontological Statistics software package for education and data analysis.** *Palaeontologica Electronica* 2001, **4**, Article 4.
70. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**:1754-1760.
71. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**:2078-2079.
72. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Cheetham RK, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu XH, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, et al: **Accurate whole human genome sequencing using reversible terminator chemistry.** *Nature* 2008, **456**:53-59.
73. Dykhuizen DE, Hartl DE: **Selection in chemostats.** *Microbiol Rev* 1983, **47**:150-168.
74. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278**:631-637.

doi:10.1186/gb-2012-13-6-r41

**Cite this article as:** Maharjan *et al.*: The multiplicity of divergence mechanisms in a single evolving population. *Genome Biology* 2012 **13**:R41.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

