Author for correspondence:
Richard E. Lenski
e-mail: lenski@msu.edu

†These authors contributed equally to this study.

THE ROYAL SOCIETY
PUBLISHING

Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*

Richard E. Lenski[1,2,3,4,†], Michael J. Wiser[1,3,†], Noah Ribeck[1,2], Zachary D. Blount[1,2], Joshua R. Nahum[1], J. Jeffrey Morris[1,5], Luis Zaman[1,6], Caroline B. Turner[1,3,7], Brian D. Wade[4], Rohan Maddamsetti[1,3], Alita R. Burmeister[1,2], Elizabeth J. Baird[1,2], Jay Bundy[1,3], Nkrumah A. Grant[1,2], Kyle J. Card[1,2], Maia Rowles[1], Kiyana Weatherspoon[1], Spiridon E. Papoulis[1,8], Rachel Sullivan[2], Colleen Clark[2], Joseph S. Mulka[2] and Neerja Hajela[2]

[1]BEACON Center for the Study of Evolution in Action, [2]Department of Microbiology and Molecular Genetics, [3]Department of Integrative Biology, and [4]Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI, USA
[5]Department of Biology, University of Alabama at Birmingham, Birmingham, AL, USA
[6]Department of Biology, University of Washington, Seattle, WA, USA
[7]Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[8]Department of Microbiology, University of Tennessee, Knoxville, TN, USA

Many populations live in environments subject to frequent biotic and abiotic changes. Nonetheless, it is interesting to ask whether an evolving population's mean fitness can increase indefinitely, and potentially without any limit, even in a constant environment. A recent study showed that fitness trajectories of *Escherichia coli* populations over 50 000 generations were better described by a power-law model than by a hyperbolic model. According to the power-law model, the rate of fitness gain declines over time but fitness has no upper limit, whereas the hyperbolic model implies a hard limit. Here, we examine whether the previously estimated power-law model predicts the fitness trajectory for an additional 10 000 generations. To that end, we conducted more than 1100 new competitive fitness assays. Consistent with the previous study, the power-law model fits the new data better than the hyperbolic model. We also analysed the variability in fitness among populations, finding subtle, but significant, heterogeneity in mean fitness. Some, but not all, of this variation reflects differences in mutation rate that evolved over time. Taken together, our results imply that both adaptation and divergence can continue indefinitely—or at least for a long time—even in a constant environment.

1. Introduction

In nature, the process of adaptation by natural selection appears inexhaustible and open-ended in its creativity [1]. Such sustained adaptation is usually thought to occur in response to changing environmental conditions, including those produced by the evolution of other organisms with which the focal population interacts. As a corollary, it is also often presumed that evolving organisms must eventually 'run out' of ways to improve, absent changes in their environment. In the parlance of adaptive landscapes, a population will, sooner or later, reach a local fitness peak [2,3].

Wiser *et al.* [4] challenged the presumption that there must be an upper bound to organismal fitness. They measured the fitness trajectories over 50 000

generations for *Escherichia coli* populations in the long-term evolution experiment (LTEE). They compared the fit of two simple models—a hyperbolic model and a power-law model—that both predict a decelerating fitness trajectory (i.e. a declining rate of fitness improvement), but only the former has an upper limit, or asymptote. The power-law model, by contrast, predicts that the logarithm of fitness will increase with the logarithm of time, a relationship that has no asymptote. Both models fit the observed fitness trajectories well, but the power-law model fit much better. Moreover, if truncated datasets (e.g. from only the first 20 000 generations) were used to predict the subsequent trajectories, the hyperbolic model consistently underestimated the extent of future improvement, whereas the power-law model accurately predicted the changes seen in later generations. Despite having no upper limit, but owing to its logarithmic dependence on time, the power law did not lead to absurd predictions that would seem to violate physical constraints. Indeed, when the power-law model was extrapolated millions of generations into the future, the predicted fitness levels correspond to growth rates that are within the range that some bacterial species can achieve under optimal conditions.

In addition, Wiser *et al.* [4] presented a dynamical model of fitness evolution for large asexual populations that also generated a power-law relationship. The dynamical model included two particularly important phenomena that have been documented in the LTEE—clonal interference, in which contemporaneous lineages with different beneficial mutations compete for fixation [5,6]; and diminishing-returns epistasis, in which the fitness advantage of beneficial mutations tends to be smaller on more-fit genetic backgrounds [7]. That model also predicted that several populations that evolved hypermutable phenotypes [8] early in the LTEE would show faster rates of fitness improvement, and this prediction was confirmed.

Here, we perform competition assays using population samples from 40 000, 50 000 and 60 000 generations to test whether the power-law model's predictions continue to hold. Because the magnitude of fitness changes are predicted to become increasingly small—and the curvature of the fitness trajectory increasingly subtle—as time goes by, we performed more than 1100 competitions to assess the fit of the model. Our results add support to the hypothesis that adaptive evolution is unbounded even in a constant environment. This extensive replication also allowed us to estimate with some precision the among-population variance component for mean fitness (i.e. the variation above the level attributable to the measurement error in replicate assays). From previous studies, we know that one of the 12 LTEE populations followed a different path from the others, when it evolved the ability to grow on citrate in the medium (included as a chelating agent), which neither the ancestor nor any of the other populations can exploit [9,10]. We also know that the populations that became hypermutable achieved a boost in their fitness trajectories [4]. Early in the LTEE, before these exceptional cases had evolved, previous estimates of the among-population variance indicated subtle but significant heterogeneity in mean fitness [11]. However, it was not known whether that variation was transient or would be sustained over the long term. Our results show that the populations—even those that appear unexceptional in other respects—continue to vary in their mean fitness levels after many tens of thousands of generations.

2. Material and methods

(a) Strains

The *E. coli* LTEE is described in detail elsewhere [11,12]. Our analyses used whole-population samples from nine of the 12 LTEE populations taken at three time points: 40 000, 50 000 and 60 000 generations (electronic supplementary material, table S1), for a total of 27 samples. We excluded three populations (Ara−2, Ara−3 and Ara+6) that no longer make colonies that can be reliably counted in the standard fitness assays or that evolved the ability to consume the citrate in the culture medium, which also precludes using the standard assays [4]. Of the nine populations used in our study, three evolved hypermutable phenotypes: Ara−1, Ara−4 and Ara+3 [8,13]. In addition to the whole-population samples, we used two clones as common competitors, REL10948 and REL11638 [4]. The former is an Ara− clone isolated from population Ara−5 at 40 000 generations; the latter is a spontaneous Ara+ mutant of that clone (electronic supplementary material, table S1). The Ara marker serves to distinguish competitors during fitness assays, as Ara− and Ara+ cells make red and white colonies, respectively, on tetrazolium–arabinose (TA) indicator plates. This marker has been shown to be selectively neutral under the glucose-limited conditions of the LTEE [14,15]. The use of an evolved strain as a common competitor should improve the precision of estimates when the fitness differential between the evolved and ancestral types is large [16]. We also used the LTEE's ancestral strain, REL606 [11,17] and three clones from populations Ara+4 and Ara+5 at 60 000 generations (electronic supplementary material, table S1) in assays to estimate mutation rates.

(b) Culture conditions

The culture conditions for the LTEE are described elsewhere [11,12]. In brief, each population is maintained by transferring 0.1 ml of culture into 9.9 ml of fresh medium every 24 h. The medium, called DM25, contains $25\,\mu g\,ml^{-1}$ glucose, which is the limiting resource. Cultures are kept at $37°C$ in a shaking incubator for aeration. The 100-fold dilution and regrowth allow approximately 6.6 generations per day. Every 500 generations (75 days), samples of each population are stored in 10% glycerol at $-80°C$; the bacteria in those samples remain viable and are available for later study.

(c) Fitness assays

The fitness assays followed the same procedures as used in the first experiment reported by Wiser *et al.* [4]. In brief, fitness was measured in the same environment as used in the LTEE by competing a population sample against a reference strain, either REL10948 or REL11638, with the opposite Ara marker state. Prior to the start of the assay, competitors were removed from the freezer and separately acclimated to the culture medium and other conditions described above. The competitors were then mixed at an equal volumetric ratio, and a sample was spread on a TA plate to estimate their initial abundances based on colony counts. The mixed competition cultures were propagated for 3 days by 1:100 daily dilutions into fresh medium and, at the end of the third day, another sample was spread on a TA plate to estimate the final abundances of the two competitors. From the initial and final counts of each type, and taking into account the dilution factors, we calculated for each competitor its realized growth rate during the assay. We then calculated fitness as the ratio of the evolved population's growth rate to that of the reference strain. The assays were performed in 42 complete blocks of 27 assays each. Twenty authors performed one block each, and one author (M.J.W.) performed 22 blocks.

(d) Missing values and outliers

Eight of the 1134 assays failed to yield fitness estimates owing to procedural errors. In addition to the missing values, we screened the remaining estimates for outliers as follows. We first log-transformed the estimates to make random deviations symmetric around the mean. We then computed the mean and standard deviation of the transformed estimates for each of the 27 population samples, and we converted each transformed estimate to a z-score by taking the absolute value of its deviation from the mean and dividing by the standard deviation. Assuming normality, one expects approximately 0.3% of the z-scores to be greater than 3 and only approximately 0.01% to be greater than 4. Thus, one would expect to see among the 1126 estimates only a few z-scores greater than 3 and none or at most one above 4. However, 13 z-scores were over 3, including five that exceeded 4. There were also 44 scores that fell between 2 and 3; assuming normality, we expect 48 such values, and so there are not excessive deviations in this range. Based on this analysis, we eliminated the estimates that produced the 13 most extreme z-scores from our analyses, leaving a total of 1113 values. The number of estimates ranged from 39 to 42 for the 27 population samples (electronic supplementary material, table S2). We also computed quantities that describe changes in fitness over time, which require estimates from multiple time points. Owing to the missing values and outliers, and because we calculated these changes using fitness estimates from the same population and block, we often had fewer estimates for these quantities. In the Results, we describe the criteria used for including data in each analysis.

(e) Fluctuation tests

We performed fluctuation tests [18] to compare the mutation rates of the ancestor and three 60 000-generation clones from each of two populations, Ara+4 and Ara+5. The bacteria were revived by inoculating 15 μl of frozen stock into 10 ml of LB broth, then incubating the cultures overnight at 37°C in a shaking incubator for aeration. Each resulting stationary-phase culture was then diluted 10 000-fold into 9.9 ml of DM25 and incubated for 24 h at 37°C with aeration. From each DM25 culture, we transferred 150–600 cells to 24 replicate 600-μl cultures of DM250 (containing 250 μg ml^{-1} glucose) in 96-well plates, which were incubated for 24 h at 37°C. Samples from six replicate cultures were then diluted and spread on TA plates to estimate the average number of cells, and 200-μl aliquots (one-third of the total volume) from all 24 cultures were spread onto LB plates supplemented with either 100 μg ml^{-1} rifampicin or 30 μg ml^{-1} nalidixic acid. The antibiotic-containing plates were incubated at 37°C for 48 h, and the number of plates with one or more resistant colonies was counted. The mutation rate, μ, was estimated using the p_0 method [18] as $-\ln(p_0)/N$, where p_0 is the proportion of replicate cultures without any mutants and N is the number of cells tested per replicate culture. If none of the replicates in a test produced any resistant mutant, then the mutation rate was estimated, by convention, using 0.5 as the number of positive replicates.
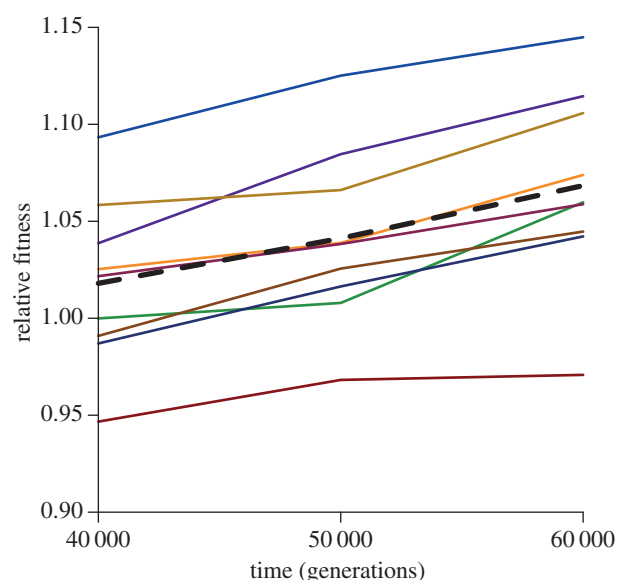
(f) Statistical analyses

Population and block are random factors in ANOVAs. Statistical analyses were performed using R v. 3.0.2 [19].

3. Results

(a) Fitness continues to increase

In all nine populations tested, the estimated mean fitness increased over both the 40 000–50 000- and the 50 000–60 000-generation intervals (figure 1; electronic supplementary

Figure 1. Trajectories of mean fitness for nine *E. coli* populations from the LTEE (thin lines) and grand-mean fitness (thick dashed line). Data from the electronic supplementary material, table S2.

material, table S2). The probability that all nine populations would, by chance, yield monotonically increasing point estimates across three time points is $(1/3 \times 1/2)^9 < 10^{-7}$.

When each population is analysed individually (table 1), the cumulative increase between 40 000 and 60 000 generations is highly significant (all $p < 0.001$). Six populations showed significant fitness gains over both component intervals, while three had significant gains in only one interval (all $p < 0.05$).

One can apply a table-wide sequential Bonferroni correction [20,21] to the data in table 1 to account for the facts that (i) we performed 27 significance tests in total and (ii) the sets of three tests for each population are not independent. However, the results hardly change—only three of the 24 tests that were individually significant at $p < 0.05$ become non-significant using this more conservative approach.

(b) Gains more consistent with the power-law model than with the hyperbolic model

Wiser *et al.* [4] presented several lines of evidence that fitness trajectories of the LTEE populations are better described by a power-law model than by a hyperbolic model. Both models were fit to trajectories based on competitions between evolved population samples and their genetically marked ancestors using data through 50 000 generations. The best fit of the hyperbolic model gave $w = 1 + 0.7007 \, t/(t + 4431)$, where w is the grand-mean fitness relative to the ancestor and t is time in generations. This equation predicts grand-mean fitness values relative to the ancestor of 1.6308, 1.6437 and 1.6525 at 40 000, 50 000 and 60 000 generations, respectively. For the power-law model, the best fit yielded $w = (0.00515 \, t + 1)^{0.0950}$, which predicts values of 1.6597, 1.6951 and 1.7246 at the same time points.

In our study, the evolved population samples competed against a high-fitness strain (a 40 000-generation clone from one of the LTEE populations), not against their much less fit ancestor. This high-fitness strain was used so that the fitness differences between the competitors were much

Table 1. Changes in mean fitness between 40 000 and 60 000 generations in the LTEE populations. Based on *N* blocks for which fitness values were obtained at all three time points. All *p*-values were calculated using one-tailed *t*-tests, given the expectation that fitness should increase. All tests are significant ($p <$ 0.05) except three that were non-significant based on individual tests (bold) and three others that were non-significant after a table-wide sequential Bonferroni correction (italics).

generations		40 000 – 60 000			40 000 – 50 000			50 000 – 60 000		
population	N	mean	s.d.	*p*-value	mean	s.d.	*p*-value	mean	s.d.	*p*-value
Ara+1	39	0.0250	0.0383	0.0001	0.0227	0.0410	0.0007	0.0023	0.0359	**0.3436**
Ara+2	38	0.0527	0.0416	<0.0001	0.0192	0.0533	*0.0164*	0.0335	0.0686	0.0023
Ara+3	38	0.0527	0.0489	<0.0001	0.0335	0.0391	<0.0001	0.0192	0.0502	*0.0120*
Ara+4	41	0.0603	0.0437	<0.0001	0.0085	0.0342	**0.0598**	0.0519	0.0468	<0.0001
Ara+5	40	0.0551	0.0410	<0.0001	0.0297	0.0329	<0.0001	0.0254	0.0420	0.0002
Ara−1	41	0.0756	0.0641	<0.0001	0.0464	0.0528	<0.0001	0.0292	0.0771	*0.0099*
Ara−4	39	0.0472	0.0425	<0.0001	0.0059	0.0490	**0.2289**	0.0414	0.0483	<0.0001
Ara−5	40	0.0541	0.0329	<0.0001	0.0348	0.0272	<0.0001	0.0193	0.0332	0.0003
Ara−6	42	0.0371	0.0267	<0.0001	0.0165	0.0311	0.0007	0.0205	0.0329	0.0001

Table 2. Comparison between the observed changes in the grand-mean fitness and those predicted by the hyperbolic and power-law models. Two-tailed *p*-values were calculated by comparing the empirical grand mean to each model's prediction using a *t*-test with 8 d.f.

generations	40 000 – 60 000	40 000 – 50 000	50 000 – 60 000
grand mean	0.0511	0.0241	0.0270
standard deviation	0.0142	0.0132	0.0143
hyperbolic model	0.0133	0.0078	0.0054
two-tailed *p*	<0.0001	0.0060	0.0019
power-law model	0.0391	0.0213	0.0174
two-tailed *p*	0.0350	0.5428	0.0804

smaller, which allowed the competitions to run for 3 days and, in turn, provided more precise estimates of small differences in relative fitness. Therefore, to compare the fitness gains estimated in our study with the predictions of the models from Wiser *et al.* [4], we needed to analyse the changes in fitness, rather than fitness itself. The hyperbolic model predicts a proportional fitness increase of 0.79% (i.e. 1.6437/1.6308−1) from 40 000 to 50 000 generations, and it predicts an increase of 0.54% from 50 000 to 60 000 generations; the predicted increase over the two intervals combined is 1.33%. The power-law model predicts increases of 2.13 and 1.74% over the same two intervals, with a cumulative gain of 3.91%.

We can test the actual fitness gains against these two sets of predictions using the means for the nine populations (table 2), where the independently evolving populations serve as the appropriate unit of replication in the comparisons. The observed fitness gains significantly exceed the gains predicted by the hyperbolic model over all three intervals (all $p < 0.01$). This outcome agrees with the pattern reported by Wiser *et al.* [4] when they truncated their 50 000-generation dataset and

used shorter-duration subsets to predict the future trajectory of mean fitness—namely, the hyperbolic model consistently underestimated the potential for future gains. The observed gains are much closer to those predicted by the power-law model, although the difference between the model and observations was marginally significant ($p = 0.035$) over the entire 40 000–60 000-generation interval (table 2). The observed improvement during that time was 5.1%, whereas the power-law model predicted a gain of 3.9%.

The power-law model has no asymptote (unlike the hyperbolic model), but it does predict that the rate of fitness increase will decline over time. Indeed, a decelerating fitness trajectory was clearly evident over the full 50 000 generations analysed by Wiser *et al.* [4]. However, the deceleration becomes less apparent as time goes on because the fitness gains become smaller, making it difficult to tell whether the changes between two successive intervals differ from one another. Using the same fitness gains predicted by the power-law model as before, we can express the predicted deceleration as the difference in gains between 40 000 and 50 000 generations and between 50 000 and 60 000 generations, which equals 0.39%. Averaged across the nine populations, the observed deceleration was −0.29% (i.e. there was a slight acceleration), but this deviation from the power-law model prediction is not significant given the variability among populations (two-tailed $p = 0.4175$). In addition to the small size of the fitness gains over these two periods, the absence of any observable deceleration may reflect limited statistical power given the small number of populations and the resulting noise associated with the stochastic occurrence of beneficial mutations.

(c) Data consistent with assumption of transitive fitness

Wiser *et al.* [4] competed the evolved populations against their common ancestor, whereas we competed them against a clone isolated from one of the evolving populations in order to reduce the fitness differences and thereby obtain more precise estimates. Thus, an implicit assumption of our analyses is that fitness is transitive (i.e. one can predict the

Table 3. Among-population variance component for fitness (V_{pop}) and corresponding standard deviation (σ_{pop}).

generation	40 000		50 000		60 000	
	V_{pop}	σ_{pop}	V_{pop}	σ_{pop}	V_{pop}	σ_{pop}
all nine populations	0.001956	0.0442	0.002005	0.0448	0.002614	0.0511
excluding hypermutators	0.000868	0.0295	0.000694	0.0263	0.001457	0.0382
also excluding Ara+1	0.000342	0.0185	0.000210	0.0145	0.000182	0.0135

fitness of C relative to A based on the fitness of C relative to B and of B relative to A). It is certainly possible that this assumption is not true in the LTEE. However, de Visser & Lenski [22] performed hundreds of competitions using 21 samples from one LTEE population, and they found no significant deviations from transitivity.

Using the data from Wiser *et al.* [4] and our study, we examined whether transitivity spans the two studies and the different competitors they used. Wiser *et al.* measured fitness values at 41 time points for each population relative to the ancestor; from those data they fit the power-law model separately to each population (table S4 in their supplement). From their fitted models, we can calculate the predicted fitness relative to the ancestor for each population at generation 60 000. We then assume transitivity and divide each population's predicted fitness at generation 60 000 by that predicted for population Ara−5 at generation 40 000 (i.e. the source of the common competitor used in our study). This approach allows us to compare the predicted, rescaled values for the nine populations with their mean fitness values measured here at 60 000 generations based on competitions against the 40 000-generation clone. The observed and predicted fitness values were both measured with error (i.e. neither is a precisely known variable); therefore, a correlation analysis is appropriate but a standard regression is not [22]. The correlation analysis indicates a highly significant positive relationship ($r = 0.822$, 7 d.f., one-tailed $p = 0.0033$) between the observed and predicted fitness values at 60 000 generations. In principle, it is possible to obtain a significant correlation, but with the observed values systematically greater or less than the predicted values. That was not the case, however, as the observed fitness values exceeded the population-specific predictions in four cases, with five cases in the other direction. On balance, then, the two datasets are consistent with the assumption of transitivity.

(d) Heterogeneity among populations in fitness

We can use the replicate assays for each population in analyses of variance to evaluate whether the fitness differences among the populations are significant, and to estimate the among-population variance component [23]. For each generation tested, we ran a two-way ANOVA with population and block as random factors (electronic supplementary material, tables S3–S5). We excluded blocks with missing values in order to fulfil a complete-block design. At all three generations, there was highly significant variation in mean fitness among the populations (all $p < 0.0001$). We found equally strong support for that variation using the rank-based, non-parametric Friedman's method [23].

The among-population variance component, V_{pop}, reflects the heterogeneity in mean fitness among populations that is above and beyond the variability caused by measurement noise (including block effects). We took the square root of the variance component to generate a corresponding standard deviation, σ_{pop}, that is commensurate in scale to fitness (table 3). At all three time points, σ_{pop} was between about 4.5 and 5%, so that a typical pair of populations differs in mean fitness by several per cent.

One source of among-population variability is that three of the nine populations in our study had evolved hypermutability, such that their point-mutation rates increased by roughly 100-fold [8,13]. Both theory and prior empirical evidence indicate that these hypermutable populations should increase in fitness somewhat faster than populations with the ancestral mutation rate, at least so long as there remain some point mutations that confer fitness benefits substantially above the increased load of deleterious mutations suffered by the hypermutators [4,12,13,24]. Indeed, we see that all three hypermutator populations—Ara+3, Ara−1 and Ara−4— have higher mean fitness than any of the other six populations at all three time points tested here (electronic supplementary material, table S2). The probability of this outcome occurring by chance alone, at any given time point, is ($3/9 \times 2/8 \times 1/7$) = 0.0119. (Owing to temporal autocorrelation, such that high and low fitness populations tend to retain their relative ranks, it is inappropriate to combine probabilities across time points.) Even if we drop the hypermutator populations, the among-population variance component for fitness remains highly significant (all $p < 0.0001$) at all three generations (electronic supplementary material, tables S3–S5). However, the corresponding standard deviation, σ_{pop}, declines to between 2.5 and 4% (table 3).

Another population, Ara+1, also stands out as unusual in two respects. First, it experienced many insertions and other non-point mutations caused by increased activity of the IS150 mobile element [25,26]. Second, the fitness trajectory for Ara+1 was notably lower than that of any other population [4]. Indeed, in our data, the mean fitness of Ara+1 was at least 4% lower than any other population at all three time points tested (electronic supplementary material, table S2). If the unusual activity of IS150 in this population is causally related to its lower fitness gains—and that connection has not been proven—then it would imply that some of the IS150-mediated mutations moved this population into a relatively unproductive region of genotypic space, one with fewer or smaller opportunities for continued adaptation. In any case, we can remove Ara+1 from our analysis, along with the three hypermutators, and test whether the five remaining— and seemingly normal—populations vary in their fitness

levels. In fact, we still see highly significant variation (all $p <$ 0.0001) at all three time points (electronic supplementary material, tables S3–S5), although the standard deviation for fitness declines further to between 1 and 2% (table 3).

(e) Heterogeneity among populations in rates of fitness gain and deceleration

We also performed analyses of variance to assess whether the rates of fitness increase from generations 40 000 to 60 000 were homogeneous or heterogeneous across the populations, and whether the extent of deceleration over the second half of that interval relative to the first varied among the populations. These analyses required blocks with no missing values or outliers for any of the 27 samples (i.e. nine populations at three time points), because they involve calculations across the generations as well as comparisons among the populations. Twenty-nine of the 42 total blocks qualified.

There is highly significant variability among the populations in the extent of their improvement between 40 000 and 60 000 generations (electronic supplementary material, table S6). The corresponding σ_{pop} is about 1.5%. In other words, a typical pair of populations differs in their fitness gains over this period by 1 or 2%. These results are essentially unchanged whether we include or exclude the populations that became hypermutable (electronic supplementary material, table S6), and the variation remains significant using Friedman's non-parametric method. One might expect a negative correlation between the extent of the fitness gain and the initial fitness at the start of this interval, which would be consistent with the tendency for diminishing-returns epistasis in the LTEE [7] and many other evolution experiments [27–32]. However, using the data from table 1 and electronic supplementary material, table S2, we find that the correlation runs in the opposite direction ($r = 0.3921$), owing largely to the fact that population Ara+1 had both the lowest fitness at 40 000 generations and the smallest gain between 40 000 and 60 000 generations. This association provides further evidence that Ara+1 has moved into a region of genotypic space with less potential for sustained adaptation. If we exclude this population, the correlation becomes negative, as expected, although it is not significant ($r = -0.0623$, 6 d.f., one-tailed $p = 0.4417$).

We also find significant variation among populations in the extent to which their fitness trajectories decelerated or accelerated (i.e. $w_{50k}/w_{40k} - w_{60k}/w_{50k}$) in the two successive 10 000-generation intervals (electronic supplementary material, table S7). This variability in the curvature of the trajectories is again significant whether or not the hypermutable populations are included, and it remains significant using Friedman's rank-based method.

We considered the possibility that a population that was not a hypermutator at 50 000 generations might have evolved hypermutability between 50 000 and 60 000 generations. In particular, population Ara+4 had the largest fitness gain of any population between 50 000 and 60 000 generations, and it also showed the most acceleration over that period compared to the prior 10 000 generations (table 1). We performed fluctuation tests [18] using clones from Ara+4 at 60 000 generations to assess whether their mutation rate was elevated, using the LTEE ancestral strain and clones from a more typical population, Ara+5, as comparisons. These tests show no indication of an elevated mutation rate in Ara+4 (electronic supplementary material, table S8), whereas the populations that previously evolved hypermutability showed approximately 100-fold increases in their mutation rates [8,13].

4. Discussion

Many natural populations live in environments that change often as the result of coevolving species, abiotic perturbations or both. This environmental variability may prevent natural populations from reaching their adaptive limits. However, it is also worthwhile examining whether a population's mean fitness can increase indefinitely, and potentially without limit, in a constant environment. Doing so may provide new insights into the limits of adaptation by natural selection. Also, while an organism's overall fitness may be subject to changing environments, it is possible that some aspects of its performance—such as core metabolic processes—might experience constant pressures that would favour little improvements even after eons of selection.

The bacteria in the LTEE live and evolve in a deliberately simple and uniform environment. Except for the daily fluctuations produced by the transfers, the exogenous environment is kept as constant as feasible by using a chemically defined medium and simple protocols. The limiting resource, glucose, is provided at a low concentration by laboratory standards, resulting in cell densities and levels of secreted metabolites that are likewise low, thereby reducing, but not eliminating, opportunities for complex frequency-dependent interactions. Although some frequency-dependent interactions have evolved in the LTEE, including a transiently stable polymorphism recently discovered in one of the populations used in our study [6], there is no reason to expect that such interactions should lead to systematic increases in fitness relative to a distant ancestor or other reference competitor. That is, any context-specific interactions should favour mutations that are beneficial relative to an organism's immediate competitors and predecessors. By contrast, fitness gains relative to a distant ancestor or other reference competitor indicate improvements in the shared, constant aspects of the environment.

Wiser et al. [4] found that a simple, two-parameter power-law model—where the rate of improvement declines over time, but fitness has no upper bound—provided an excellent description of the grand-mean fitness trajectory in the LTEE over 50 000 generations. They also showed that if the dataset was truncated, the power-law model still predicted the future trajectory with impressive accuracy. In this study, we extended the duration of that prior study by 'only' 10 000 generations (i.e. 20%). However, the changes in fitness are much smaller, and the curvature in the fitness trajectory is much more subtle, over this period than during the experiment as a whole. Therefore, to ensure that our study had the power to detect the predicted changes, we obtained about 40 fitness estimates for each population at each of 40 000, 50 000 and 60 000 generations.

Our data and analyses provide strong, albeit imperfect, support for the power-law model. The mean fitness of the populations, individually and collectively, continued to improve significantly even over these later generations (figure 1 and table 1). The average increase in relative fitness was reasonably consistent with the predictions of the power-law model based entirely on the previous data, although the difference between the 5.1% increase we observed and the predicted gain of 3.9% was marginally significant (table 2). In any case, the fit of the power-law model was

much better than that of the hyperbolic model, which predicted only a 1.3% increase in fitness (table 2). The experimental data also showed no clear decline in the rate of fitness increase from 50 000 to 60 000 generations relative to that from 40 000 to 50 000 generations; however, the predicted difference between these periods is small, and the observed and predicted values do not differ significantly. In any case, if an alternative to the power-law model were to be sought based on our data, it would have even less tendency towards deceleration, not more—in other words, our new data, like the data analysed by Wiser *et al*. [4], do not support a model in which fitness has an upper bound.

The extensive replication in our study also enabled us to quantify the divergence of the evolving populations' fitness trajectories. The among-population variation in mean fitness provides information on the structure of the fitness landscape that the grand-mean fitness trajectory cannot [11,33,34]. In particular, this variability sheds light on the form of epistasis, which is a key feature of the dynamic model developed by Wiser *et al*. [4] which gives rise to the power-law relationship. In some respects, it was already known that the fitness trajectories are not strictly parallel. First, one population evolved the ability to grow on citrate [9,10], a resource in the medium of the LTEE that remains unavailable to the other populations. This population was excluded from our analysis (and, after the function evolved, from Wiser *et al*. [4]) because the strong density- and frequency-dependent effects of that phenotype make the assays used to measure fitness inappropriate. (Two other populations were excluded from both studies because, in later generations, they do not make colonies on the plates used to enumerate competitors in the fitness assays.) Second, Wiser *et al*. [4] showed that hypermutable populations had faster-rising fitness trajectories than the other populations. However, they did not examine whether there was variation in fitness among the populations that retained the low ancestral mutation rate throughout the LTEE. Third, there was significant among-population fitness variation in the early generations of the LTEE [11,35]; however, it was not known whether it would persist or, alternatively, diminish if the populations converged on the same fitness level over time.

In this study, we found significant among-population variation in fitness at all three late-generation time points tested (electronic supplementary material, tables S3–S5). Moreover, this variation was significant even if those populations that became hypermutable were excluded from the analysis. The square root of the among-population variance component (i.e. comparable to a standard deviation) over the period from 40 000 to 60 000 generations was about 5% for all of the populations in our analysis and about 3% without the hypermutable populations (table 3). Population Ara+1 was a notable outlier among the non-hypermutable populations, having the lowest fitness of all populations at each generation we tested; this population also had the lowest trajectory in the study by Wiser *et al*. [4].

Given the empirical and theoretical support for pervasive diminishing-returns epistasis in the LTEE [4,7], one would expect the Ara+1 population to show a propensity to improve faster than the other non-hypermutable populations. That is, its low fitness implies greater scope for improvement under the assumption that all populations are subject to the same form and strength of diminishing-returns epistasis. In fact, however, Ara+1 had the least improvement from 40 000 to 60 000 generations of any population—in other words, the opposite of that expectation. This result implies that this population has, in some sense, gotten stuck in a genotypic region of the fitness landscape that constrains its evolvability. It follows that the coefficient that describes the strength of diminishing-returns epistasis is not a constant—even in the constant environment of the LTEE—but instead it must vary between local neighbourhoods in genotypic space.

Besides its unique fitness trajectory, population Ara+1 also had an unusually high number of mutations caused by the IS*150* mobile element [25,26]. In some respects, that would seem to make Ara+1 similar to the populations that became hypermutable and that had the fastest rates of fitness improvement. Indeed, some genes that acquired mutations caused by IS elements in one population accumulated point mutations in other populations, implying that knockout or knockdown mutations in those genes were beneficial [36,37], which has been confirmed in some cases [38]. However, insertions are typically more disruptive of gene functions than are point mutations. Therefore, a population that evolves insertion-mediated hypermutability may have fewer opportunities for subtle refinements that could compensate for earlier mutations that, while beneficial, overshot some optimum effect. We emphasize this connection between IS-mediated hypermutability and reduced evolvability is, at present, merely conjecture. However, we think it is a hypothesis worthy of further exploration and testing.

How rugged is the fitness landscape on which the LTEE populations are evolving? De Visser & Krug [39] present several metrics that describe the ruggedness of fitness landscapes based on the individual and combined effects of mutations found in evolved genotypes. Using data from the first five mutations to fix in one population [7], the LTEE landscape appears smooth when compared to other studies in which mutations were combined and their fitness effects quantified. Smoother landscapes presumably tend to promote more repeatable adaptive trajectories and, indeed, many studies have identified parallel phenotypic and genetic changes in the LTEE [35,37,38,40–42]. Nonetheless, there are many examples of divergence as well. Some of these seem minor, such as the fact that very few mutations are identical at the sequence level, even when the same gene has beneficial mutations in most or all of the LTEE populations [37]. Other cases of divergence have more obvious and important consequences, such as the evolution of hypermutability in some populations [4,8]. The most striking case of divergence in the LTEE is the new ability to grow on the citrate that is present in the medium, which evolved in only one population [9,10]. In this study, we demonstrated additional divergences impacting the fitness trajectories of the seemingly ordinary populations that did not become hypermutable or evolve the ability to consume citrate. We showed that population Ara+1 is peculiar in having both the lowest and shallowest fitness trajectory. Even among the other non-hypermutable populations there is significant variation in their fitness levels. Hence, subtle divergences are ubiquitous even in the most ordinary populations. In closing, both adaptation and divergence are continuing unabated in the LTEE, even after many tens of thousands of generations in a constant environment.

References

1. Thompson JN. 2013 *Relentless evolution*. Chicago, IL: University of Chicago Press.
2. Wright S. 1932 The role of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proc. Sixth Intl. Cong. of Genetics* (ed. DF Jones), pp. 356–366. Brooklyn, NY: Brooklyn Botanic Garden.
3. Wright S. 1984 *Evolution and the genetics of populations,* vol. 3. Chicago, IL: University of Chicago Press.
4. Wiser MJ, Ribeck N, Lenski RE. 2013 Long-term dynamics of adaptation in asexual populations. *Science* **342**, 1364–1367. (doi:10.1126/science.1243357)
5. Gerrish PJ, Lenski RE. 1998 The fate of competing beneficial mutations in an asexual population. *Genetica* **102/103**, 127–144. (doi:10.1023/A:1017067816551)
6. Maddamsetti R, Lenski RE, Barrick JE. 2015 Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* **200**, 619–631. (doi:10.1534/genetics.115.176677)
7. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011 Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196. (doi:10.1126/science.1203801)
8. Sniegowski PD, Gerrish PJ, Lenski RE. 1997 Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature* **387**, 703–705. (doi:10.1038/42701)
9. Blount ZD, Borland CZ, Lenski RE. 2008 Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105**, 7899–7906. (doi:10.1073/pnas.0803151105)
10. Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012 Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**, 513–518. (doi:10.1038/nature11514)
11. Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991 Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. *Am. Nat.* **138**, 1315–1341. (doi:10.1086/285289)
12. Lenski RE. 2004 Phenotypic and genomic evolution during a 20 000-generation experiment with the bacterium *Escherichia coli*. *Plant Breed. Rev.* **24**, 225–265.
13. Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, Chane-Woon-Ming B, Médigue C, Lenski RE, Schneider D. 2013 Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc. Natl Acad. Sci. USA* **110**, 222–227. (doi:10.1073/pnas.1219574110)

14. Lenski RE. 1988 Experimental studies of pleiotropy and epistasis in *Escherichia coli*. I. Variation in competitive fitness among mutants resistant to virus T4. *Evolution* **42**, 425–432. (doi:10.2307/2409028)
15. Woods RJ, Barrick JE, Cooper TF, Shrestha U, Kauth MR, Lenski RE. 2011 Second-order selection for evolvability in a large *Escherichia coli* population. *Science* **331**, 1433–1436. (doi:10.1126/science.1198914)
16. Wiser MJ, Lenski RE. 2015 A comparison of methods to measure fitness in *Escherichia coli*. *PLoS ONE* **10**, e0126210. (doi:10.1371/journal.pone.0126210)
17. Jeong H *et al.* 2009 Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol.* **394**, 644–652. (doi:10.1016/j.jmb.2009.09.052)
18. Luria SE, Delbrück M. 1943 Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511.
19. R Core Team. 2013 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. (http://www.R-project.org/)
20. Holm S. 1979 A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70.
21. Rice WR. 1989 Analyzing tables of statistical tests. *Evolution* **43**, 223–225. (doi:10.2307/2409177)
22. de Visser JAGM, Lenski RE. 2002 Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. *BMC Evol. Biol.* **2**, 19. (doi:10.1186/1471-2148-2-19).
23. Sokal RR, Rohlf FJ. 1981 *Biometry,* 2nd edn. New York, NY: WH Freeman.
24. 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. (doi:10.1126/science.283.5400.404)
25. Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE, Blot M. 1999 Genomic evolution during a 10 000-generation experiment with bacteria. *Proc. Natl Acad. Sci. USA* **96**, 3807–3812. (doi:10.1073/pnas.96.7.3807)
26. Raeside C *et al.* 2014 Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. *mBio* **5**, e01377-14. (doi:10.1128/mBio.01377-14)

27. Lenski RE. 1988 Experimental studies of pleiotropy and epistasis in *Escherichia coli*. II. Compensation for maladaptive pleiotropic effects associated with resistance to virus T4. *Evolution* **42**, 433–440. (doi:10.2307/2409029)
28. Travisano M, Mongold JA, Bennett AF, Lenski RE. 1995 Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**, 87–90. (doi:10.1126/science.7809610)
29. Moore FB-G, Rozen DE, Lenski RE. 2000 Pervasive compensatory adaptation in *Escherichia coli*. *Proc. R. Soc. Lond. B* **267**, 515–522. (doi:10.1098/rspb.2000.1030)
30. Barrick JE, Kauth MR, Strelioff CC, Lenski RE. 2010 *Escherichia coli rpoB* mutants have increased evolvability in proportion to their fitness defects. *Mol. Biol. Evol.* **27**, 1338–1347. (doi:10.1093/molbev/msq024)
31. Chou H-H, Chiu H-C, Delaney NF, Segrè D, Marx CJ. 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192. (doi:10.1126/science.1203799)
32. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. 2014 Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522. (doi:10.1126/science.1250939)
33. Nahum JR, Godfrey-Smith P, Harding BN, Marcus JH, Carlson-Stevermer J, Kerr B. 2015 A tortoise–hare pattern seen in adapting structured and unstructured populations suggests a rugged fitness landscape in bacteria. *Proc. Natl Acad. Sci. USA* **112**, 7530–7535. (doi:10.1073/pnas.141063111)
34. McCandlish DM, Otwinowski J, Plotkin JB. 2015 Detecting epistasis from an ensemble of adapting populations. *Evolution* **69**, 2359–2370. (doi:10.1111/evo.12735)
35. Lenski RE, Travisano M. 1994 Dynamics of adaptation and diversification: a 10 000-generation experiment with bacterial populations. *Proc. Natl Acad. Sci. USA* **91**, 6808–6814. (doi:10.1073/pnas.91.15.6808)
36. Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M. 2000 Long-term experimental evolution in *Escherichia coli*. IX. Characterization of IS-mediated mutations and rearrangements. *Genetics* **156**, 477–488.
37. Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. 2006 Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*.

8

rspb.royalsocietypublishing.org *Proc. R. Soc. B* **282**: 20152292

Proc. Natl Acad. Sci. USA **103**, 9107–9112. (doi:10.1073/pnas.0602917103)

38. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009 Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243–1247. (doi:10.1038/nature08480)

39. de Visser JAGM, Krug J. 2014 Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Gen*. **15**, 480–490. (doi:10.1038/nrg3744)

40. Vasi F, Travisano M, Lenski RE. 1994 Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. *Am. Nat*. **144**, 432–456. (doi:10.1086/285685)

41. Cooper TF, Rozen DE, Lenski RE. 2003 Parallel changes in gene expression after 20 000 generations of evolution in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **100**, 1072–1077. (doi:10.1073/pnas.0334340100)

42. Crozat E, Winkworth C, Gaffé J, Hallin PF, Riley MA, Lenski RE, Schneider D. 2010 Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of *Escherichia coli*. *Mol. Biol. Evol*. **27**, 2113–2128. (doi:10.1093/molbev/msq099)