PNAS
www.pnas.org

Supplementary Information for

Each of 3,323 metabolic innovations in the evolution of *E. coli* arose through the horizontal transfer of a single DNA segment

Tin Yau Pang & Martin J. Lercher

Martin J. Lercher
Email: Martin.Lercher@hhu.de

This PDF file includes:

Detailed Materials and Methods
Figs. S1 to S9
Captions for Datasets S1 to S6
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S6

Detailed Materials and Methods

Universal GPR rules

We start from the previously published metabolic network reconstructions of 55 extant *E. coli* and *Shigella* strains (1); we excluded the two strains SE11 and 55989, as their position on the phylogeny of vertical descent is uncertain (2).

Mapping genotypes to phenotypes via FBA modeling requires gene-protein-reaction (GPR) rules; these are logical statements that report the proteins or protein complexes (genes linked via AND) required for the reaction to proceed; if alternative proteins or protein complexes exist, these are linked via OR statements. Some of the GPR rules provided in the xml files of the metabolic networks reconstructed by *Monk et al.* (1) are inconsistent between *E. coli* strains. Thus, we first defined a universal set of GPR rules. We merged all alternative GPRs from different strains by a logical OR; *e.g.*, if a reaction is catalyzed by a single protein but by an enzyme complex in another strain, we assumed that the reaction can proceed as long as either an ortholog of the gene encoding the single protein or orthologs of all genes encoding the enzyme complex members are present in the genome. Some reactions occur spontaneously and are thus not associated with any genes, *i.e.*, they have empty GPR rules. We assumed that a reaction occurs spontaneously (*i.e.*, we allowed the reaction to proceed without genomic evidence) only if the GPR was empty in all 53 metabolic models. All other reactions had non-empty GPRs; they were only included in a strain-specific model if the strain's genes fulfilled the corresponding GPR rules. Dataset S3 lists the compiled universal GPR rules.

We replaced the existing GPR rules with the universal GPR rules in each of the 53 metabolic models. If one of the alternative enzymes or enzyme complexes linked by logical ORs is not fully supported by a strain's genome, this part of the GPR was removed; if none of the enzymes or enzyme complexes linked by OR is supported by the genome, the reaction was removed from the metabolic model.

Ancestral genomes and HGT

We obtained orthologous gene families and their presence in the genomes of ancestral strains from Pang & Lercher (2). In this previous publication, we had used the program ProteinORTHO (3) to cluster protein sequences from the 53 strains and 17 outgroup genomes into orthologous gene families. 1,334 protein families had exactly one member in each of the genomes, and

these were used to reconstruct a maximum-likelihood phylogeny representing vertical descent; each internal branch of this phylogeny had a bootstrap support of at least 60%. Gains and losses of orthologous gene families as well as the gene content of ancestral strains were estimated using the maximum-likelihood algorithm in GLOOME (4). If the probability of an orthologous gene family was predicted by GLOOME to be ≥ 0.5 at an internal node of the phylogeny, then the gene family was assumed to be present, else it was considered absent (2).

We also investigated alternative methods to infer HGT events and the gene content of ancestral strains. Horizontally acquired genes are generally identified through one of three classes of methods: (i) explanation of gene presence/absence patterns on a tree representing vertical descent through gene gains and losses (as perfomed by GLOOME (4), employed in Pang & Lercher (2)); (ii) detection of unusual DNA sequence composition or codon usage; and (iii) reconciliation of gene trees with a reference tree representing vertical descent. The second of these algorithm classes relies on significant sequence composition differences between the donors and the recipients of horizontally transferred sequences. This type of method is thus not suited for the detection of HGT or recombination between strains from the same species. To explore alternatives to the gene/presence method employed in Ref. (2), we thus applied two methods that infer HGT events based on the reconciliation of individual gene trees with the phylogeny of vertical descent, RANGER-DTL (5) and ALE (6).

When applied to the 53 genomes in our dataset, the GLOOME algorithm employed in Ref. (2) predicted ancestral genome sizes of 4,242±269 genes, compared to 4,509±351 genes in the extant strains. In contrast, both tree reconciliation methods predicted massively reduced ancestral genomes of 2,974±1,076 (ALE) and 2,084±641 (RANGER-DTL with default parameter settings), with older ancestors possessing increasingly reduced genomes (SI Appendix, Fig. S9a). For RANGER-DTL, it was possible to increase the average size of ancestral genomes by reducing the gene loss penalty (we tried gene loss penalties of 1 (default), 0.5, 0.1, and 0.01); note that for each parameter setting, the spread of genome sizes was much higher than in either the extant genomes or in the GLOOM reconstructions (SI Appendix, Fig. S9a).

Importantly, regardless of parameter setting, the total number of metabolic genes per genome was much smaller in the reconstructions obtained through the tree reconciliation methods compared to either the extant genomes or the GLOOME reconstructions (SI Appenix, Fig. S9b). Accordingly, none of the metabolic networks reconstructed from the ancestral genomes predicted by ALE and RANGER-DTL was capable of producing biomass even in the most nutrient-rich environments. Thus, it appears that gene tree reconciliation is unsuited for the

dataset of closely related genomes analysed here, supporting the choice of the gene presence/absence-based method employed in Ref. (2).

Reconstruction of horizontally transferred DNA segments and their metabolic genes

Based on the gene acquisitions inferred by Pang & Lercher (2), we identified sets of genes that were likely co-transferred on a single DNA segment. This was done based on the observation that DNA segments acquired via HGT and retained by *E. coli* strains are at most 30kb in length (2).

Let **S** be the set of all orthologous gene families ("genes") acquired along a given branch. Let **E** be the set of extant descendants of this branch. We carried out the following iterative algorithm to group the horizontally transferred genes into subsets of genes that were likely co-transferred on a single DNA segment:

1. identify the start positions of every orthologous gene family in **S** in the genomes in **E**;
2. pick a random orthologous gene family g_A in **S** and put it into a new set **P**;
3. for each gene *g* in **S** not included in **P** ($\forall g \in \mathbf{S} \backslash \mathbf{P}$), count the number of extant genomes in **E** where *g* together with all genes in **P** is contained within a 30kb segment of the genome;
4. pick the one gene *g* in **S** not included in **P** ($g \in \mathbf{S} \backslash \mathbf{P}$) supported by the highest segment count, and add it to **P** (if multiple genes are supported equally, pick one randomly);
5. repeat steps 3-4 until no more genes can be added to **P;**
6. the genes in **P** are then grouped into a single HGT event; these genes are removed from **S**;
7. Repeat steps 2-6 to reconstruct another HGT event until **S** is empty (*i.e.*, all horizontally acquired genes have been assigned to a DNA segment).

Dataset S4 lists our reconstruction of HGT segments and their associated genes. To test the reliability of our reconstruction, we repeated our algorithm 1,000 times. We observed that a set of genes grouped into a DNA segment in one realization had a 99.3% chance to also occur as a co-transferred set in another realization, validating the accuracy of the greedy assignments.

Note that the probability for two randomly chosen metabolic genes to reside within 30kb of each other on at least one extant genome is 0.0081; this demonstrates that two genes acquired along the same phylogenetic branch are very unlikely to be found within 30kb on the genome of one of the descendants unless they were indeed co-transferred.

Reconstruction of metabolic networks from ancestral genomes

We applied the universal GPR rules to the gene content of all ancestral strains (2) to assess which metabolic reactions are supported by their genomes, and assembled corresponding metabolic networks; we always included all reactions with an empty universal GPR rule. Reaction reversibilities and biomass reactions are universal across the metabolic reconstructions published by Monk *et al*. (1); these were transferred to the ancestral metabolic models. We used the standard biomass reaction (Ec_biomass_iJO1366_core_53p95M) in all simulations. All models also contained a non-growth related maintenance energy consumption term, a flux of at least 3.15 through the ATPM reaction, which converts ATP to ADP to simulate non-growth related energy requirements. The numbers of total genes, metabolic genes, and metabolic reactions in each of the extant and ancestral strains are shown in Dataset S1.

Identifying the metabolic phenotypes of extant and ancestral E. coli strains in different nutritional environments

We applied flux balance analysis (FBA) (7) to calculate the biomass production rate of each extant and ancestral metabolic model across 202,418 different nutritional environments. 2,418 of these nutritional environments are based on those used by *Szappanos et al.* (8), who tested 1,776 aerobic nutritional environments. Of these, 1,209 nutritional environments remained after removing those that contained exchange reactions not found in the metabolic models of *Monk et al.* (1)**.** We doubled this set by adding corresponding anaerobic environments.

We further generated 100,000 different random nutritional environments, constructed by choosing a random source each of carbon, nitrogen, phosphorus, and sulphur (see Dataset S5 for the lists of compounds used as possible carbon, nitrogen, phosphorus, and sulphur sources). To each environment, we added 12 other essential nutrients: calcium (ca2), chloride (cl), copper (cu2), cobalt (cobalt2), iron (fe3), water (h2o), potassium (k), magnesium (mg2), manganese (mn2), molybdate (mobd), zinc (zn2), and nickel (ni2). These 100,000 nutritional environments are anaerobic; adding the corresponding aerobic environments resulted in a total of 200,000 random environments. Dataset S2 lists the nutrients in each of the 202,418 environments tested. When performing FBA simulations, we set the lower bound of all exchange reactions for nutrients present in the tested environment to -10; the lower bound of all other exchange reactions was set to 0.

We used the *sybil* package (9) in R (10) to perform FBA simulations on the 53 extant and 52 ancestral strains across the 202,418 nutritional environments. Maximal biomass reaction

rates $\leq 10^{-6}$ were interpreted as an inability of the strain to grow in the tested environment. The bimodal distribution of the number of viable environments per strain is shown in the SI Appendix (Fig. S3). 76 of the metabolic models (46 ancestral and 30 extant strains) can grow in more than 20% of the tested environments, while the remaining 29 models can produce biomass in at most 0.5% of the tested environments; 26 models cannot produce biomass in any of the tested environments. The 29 phenotypically severely restricted strains were either highly metabolically specialized or had incomplete metabolic pathways (*e.g.*, because reactions added by Monk *et al.* (1) without genomic evidence were removed in our application of the universal GPR rules). Thus, we excluded these strains from further analysis and present results only for the remaining 76 strains (30 extant and 46 ancestral).

Phenotypic and genomic distances between strains

We defined the phenotypic distance of a pair of strains as 1-*J*, where *J* is the Jaccard index of the subsets of nutritional environments (out of the 202,418 tested environments) in which each of the two strains can grow according to the FBA simulations. We defined the genomic distance between two strains as the amino acid sequence distance calculated per amino acid site from the concatenated global alignments of 1,334 core genes (2).

Phenotypic innovations and corresponding DNA segment transfers in the real data and in a horizontal gene transfer (HGT) model

To evaluate the phenotypic innovations conferred by DNA segment transfer events in the real data, we applied FBA to the metabolic networks at the beginning and end of the phylogenetic branch on which one or more DNA transfers occurred. As before, we used a threshold of 10^{-6} to distinguish growth ("non-zero" biomass production) from non-growth ("zero" biomass production). We compared the biomass production of the ancestral and derived strains in each of the 202,418 environments, and selected those environments that have phenotypic innovations that fall into one of two phenotypic innovation types:

I. *New phenotypes*, where the derived strain can grow in the tested environment while the ancestral strain could not.
II. *Yield-improved phenotypes*, where both ancestral and derived strains can grow in the tested nutritional environment, but the maximal biomass flux at least doubled in the derived compared to the ancestral strain.

To study more generally how HGT can lead to phenotypic innovations, we developed a model based on the 30 non-auxotrophic extant strains. In each strain, we identified all non-universal (accessory) metabolic genes, with their genomic position represented by their transcription start sites. We then searched for all possible gene combinations that can be transferred together (*i.e.*, that are found within a 30kb segment of the genome (2)) and that are not a subset of another set. The sets of genes defined in this way thus correspond to all metabolically relevant potential horizontal gene transfers of DNA segments between extant strains (listed in Dataset S6). For each segment, we modeled the effects of its transfer to those of the other 29 non-auxotrophic extant strains that miss at least one of its genes: we calculated the maximal biomass flux in each of the 202,418 nutritional environments and counted the two types of phenotypic innovations.

Phenotype-transferability between a pair of donor and recipient strains

To test the extent to which pairs of strains can benefit from HGTs, we examined every pair of extant non-auxotrophic strains. Given a donor and a recipient, we considered the phenotypes present in the donor but not in the recipient. Such a "unique" phenotype is single-segment-transferrable if there exists a DNA segment of at most 30kb in the donor such that the transfer of the segment to the recipient leads to the transfer of the phenotype. The phenotype transferability of a donor-recipient pair is defined as the proportion of single-segment-transferrable phenotypes out of all unique phenotypes of the donor relative to the recipient.

Segment transfers resulting in phenotype transfers

We also calculated the probability that a segment bestows any new phenotype on its recipient, regardless of whether that phenotype was present in the donor. For each donor/recipient pair of extant non-auxotrophic strains, we calculated the fraction of donor DNA segments containing metabolic genes not present in the recipient that give rise to at least one new phenotype in the recipient strain.

We further tested if the transfer of the same DNA segment into two recipients leads to the same phenotypic innovation. For a pair of extant non-auxotrophic strains A and B, we considered all environments in which both A and B cannot grow. We then examined all DNA segments of other strains that can bring at least one new metabolic gene to each of their genomes and simulated the consequences of their acquisition by both A and B using FBA.

Based on these results, we calculated the probability for a segment to give rise to the same phenotypic innovation in both A and B.

Minimal number of DNA segments required for a phenotypic innovation

To estimate the number of DNA segments required for a phenotypic innovation observed in the real data, we first performed parsimonious FBA (pFBA) (11) on the metabolic network of the derived strain of the branch in the environment for which a phenotypic innovation occurred. pFBA approximates efficient enzyme usage and reports the reactions most likely to be active in the nutritional phenotype (11). From these, we identified the subset of reactions acquired along the branch. Based on the previously identified transferred DNA segments (Dataset S4, see above), we then identified the number of segments involved in the phenotypic innovation. As pFBA only approximately describes the physiology of biological systems (11), this procedure might inflate the number of segments contributing to the phenotypic innovation. Hence, when pFBA suggested that more than one segment contributed to a phenotypic innovation, we simulated transferring all possible combinations of those predicted segments to the ancestral network to see if a smaller number of segments was already able to bestow the observed phenotypic innovation.

As a reference, we also investigated the number of segments required for each possible phenotypic innovation achievable through HGT within the *E. coli* pan-genome. We first constructed the metabolic supermodel that contains all reactions from the 30 extant non-auxotrophic strains, and performed pFBA on the 202,418 nutritional environments to obtain the maximal biomass flux and the corresponding reaction pathways in the supermodel. For each environment, we compared the biomass flux in the supermodel with that in each non-auxotrophic extant strain; we thereby identified all possible phenotypic innovations for the extant strains within the *E. coli* pan-genome. Those reactions that are, according to pFBA, involved in the supermodel phenotype but are not present in the considered extant recipient strain are those that bestow the considered phenotypic innovation; these reactions must be acquired by the recipient through HGT. We searched for the minimum number of segments in the extant genomes that can provide those extra reactions (see Dataset S6 for potential segments). If this method predicted multiple segments, we simulated transferring each segment found on the extant strains (Dataset S6) to the recipient strain to test if an alternative single segment transfer can already result in the phenotypic innovation (note that the pFBA-based method is computationally much faster, which is why we first used it to identify all clear cases of single-segment transfers).

In the analysis of the real data, we identified 1,998 new phenotypes (type I) and 1,325 yield improvements (type II), and found that none of them required more than a single DNA segment transfer to arise. In the HGT model, the fraction of potential innovations that require multiple DNA segment transfers was 0.0239 for new phenotypes (type I) and 0.0739 for yield-improved phenotypes (type II). We applied a single-sided binomial test to examine the hypothesis that the chance of multi-segment phenotypic innovations in the real data is lower than expected from the HGT model of all possible innovations, which is the case ($P<10^{-15}$ for each type individually and for both types combined).

Exaptation of earlier HGT events for later phenotypic innovations

In order to assess the role of exaptation in *E. coli* metabolic evolution, we identified phenotypic innovations that utilized enzymes or transporters transferred into the lineage on previous branches of the phylogeny. For each ancestral and extant metabolic network, we used the pFBA results calculated above to identify the active reactions involved in any newly evolved phenotypic innovation (types I and II). Each of these reactions was either (i) inherited from the most recent common ancestor (MRCA) of all examined strains—the root node of the phylogeny—, (ii) acquired via HGT on the phylogenetic branch immediately leading to the examined strain—reactions directly responsible for the innovation—, or (iii) acquired via HGT on an earlier branch. The reactions in category (iii) were thus acquired much earlier than those in category (ii) and were exapted for the considered phenotype. We identified the corresponding minimal number of DNA segments gained via HGT on earlier branches (see Dataset S4 for the segments transferred into different branches).

Figures

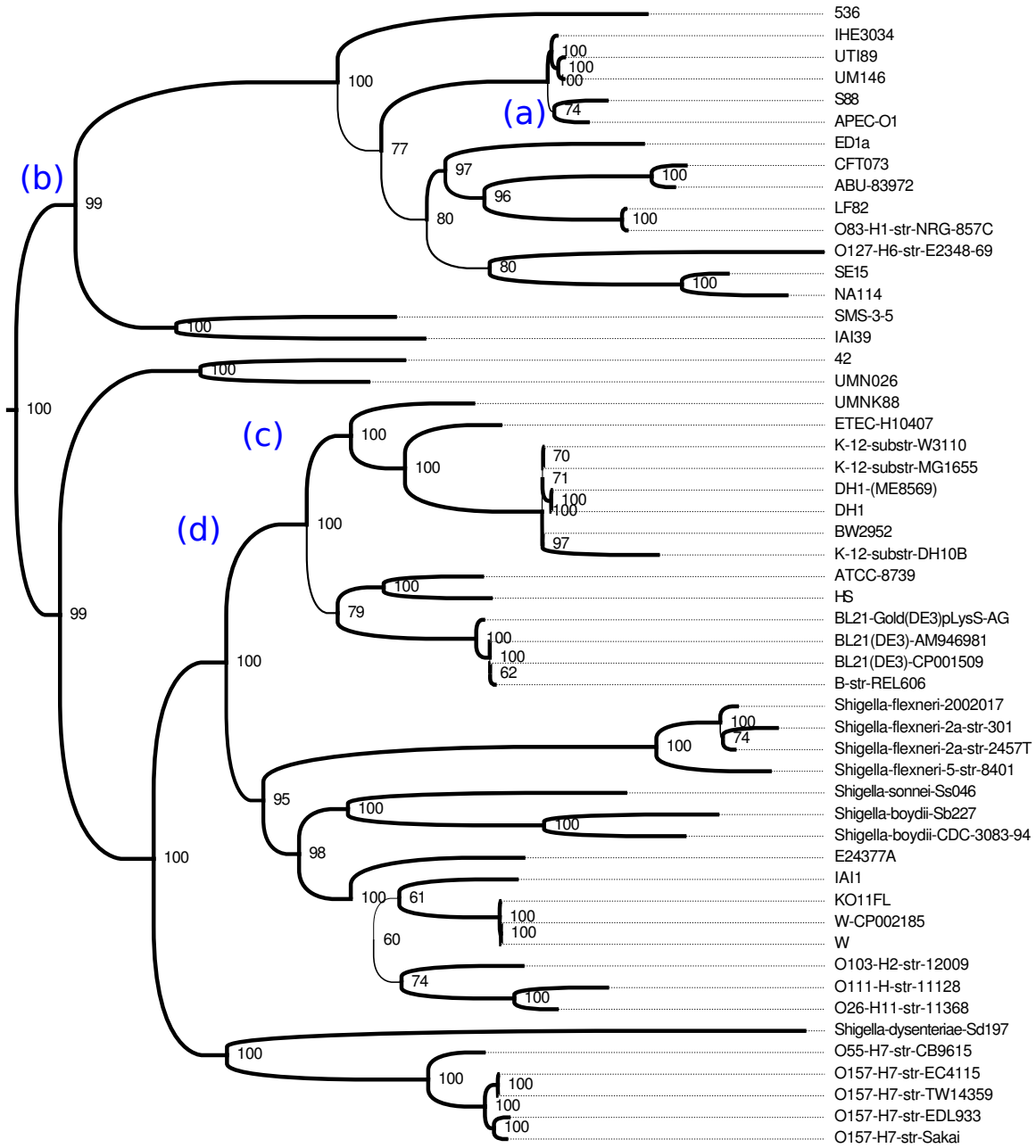

Figure S1. Phylogenetic tree describing vertical inheritance among the 53 *E. coli* strains (including *Shigella*) analysed in this work (redrawn from Figure S2 of Pang & Lercher (2)). Numbers to the right of branches (as well as branch line widths) reflect the maximum-likelihood bootstrap support values in percent. Reactions contributing to the complex phenotypic innovation described in Figs. S6 and S7 were acquired via HGT on the branches marked with (a) and (b) to the left of the respective branches; reactions contributing to the complex phenotypic innovation described in Fig. S8 were acquired via HGT on the branches marked with (c) and (d) to the left of the respective branches. For a dendrogram putting these strains into the wider context of the *E. coli* species, see Fig. S2.

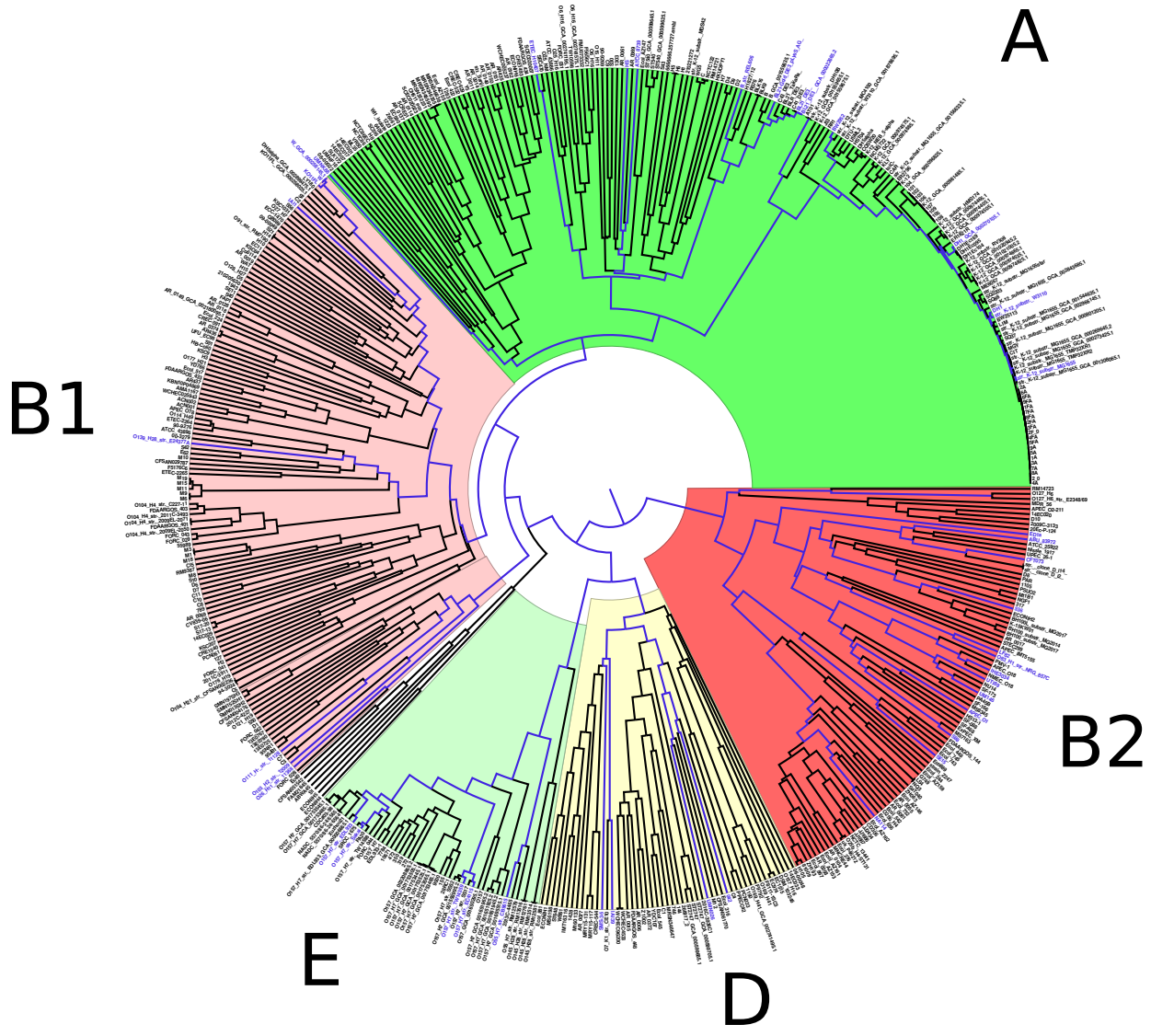

Figure S2. Dendrogram placing the *E. coli* strains analyzed in this work (blue labels and blue branches) into the larger context of the *E. coli* species. The colored background distinguishes the major *E. coli* clades (A, B1, B2, D, E).

The dendrogram was obtained by reducing the *E. coli* "Genome tree report" provided by NCBI (https://www.ncbi.nlm.nih.gov/genome/tree/167), which is based on whole-genome sequence similarities estimated with BLAST, to fully sequenced genomes (*N*=591) and complete chromosomes (*N*=59); we also added the contig representing *E. coli* strain NA114, which had recently been demoted from whole-genome status by NCBI.

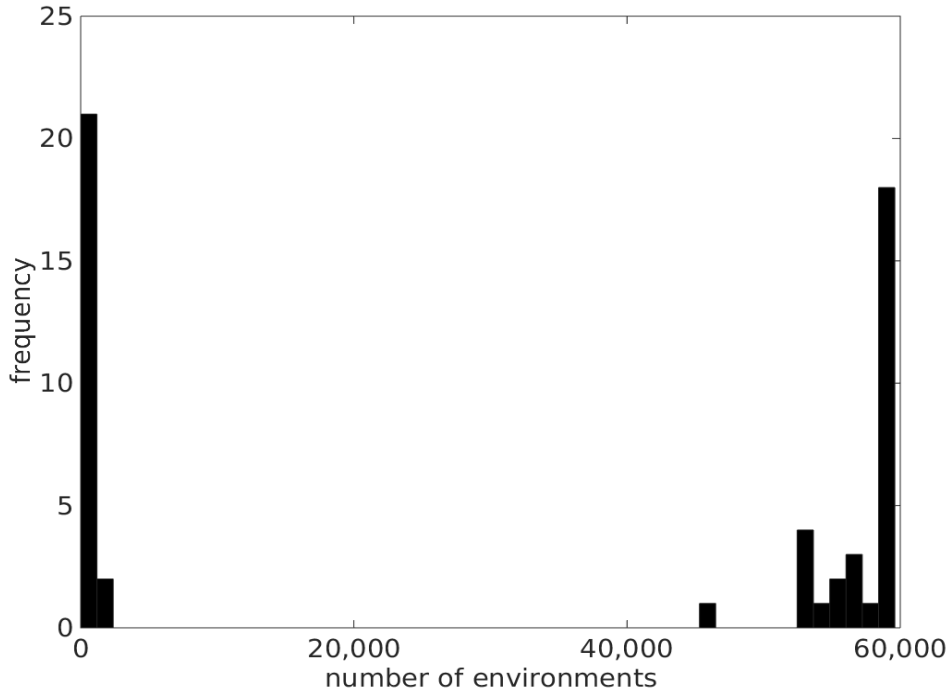

Figure S3. Distribution of the number of nutritional environments in which individual extant strains can grow (including auxotrophic strains). The strains in the left half of the histogram are either highly metabolically specialized or have incomplete metabolic models and where excluded from the evolutionary analyses.

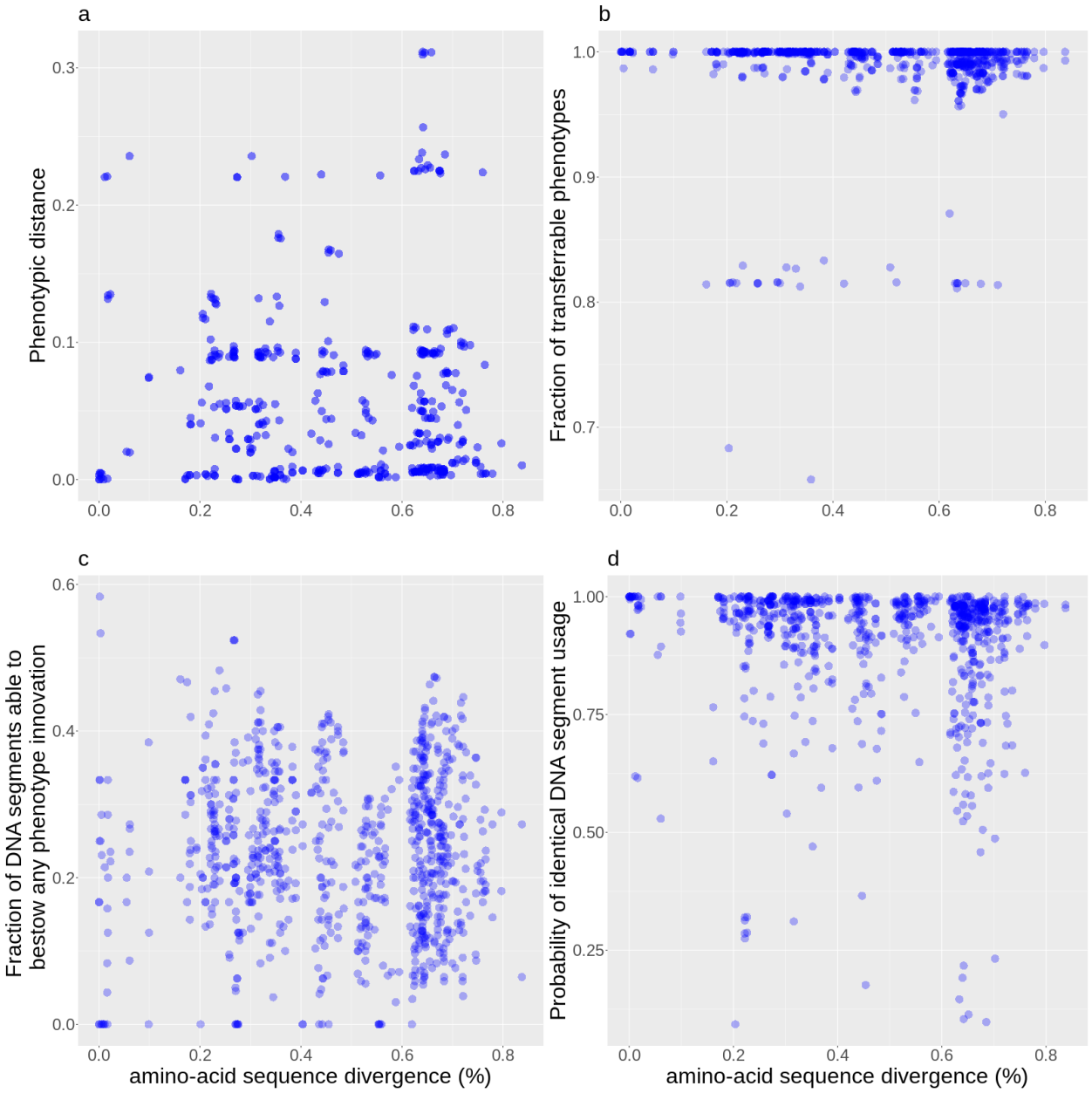

Figure S4. Amino acid sequence divergence between two strains does not correlate with phenotypic distance or the probability that HGT between them transfers a phenotype, but correlates weakly with the transferability of specific phenotypes and with the equal utilization of DNA segments. **a,** Phenotypic distance, defined based on the Jaccard index of the sets of environments in which two strains can grow (Spearman's ρ=0.0052, P=0.88). **b,** Fraction of complementary phenotypes transferrable from a donor to a recipient (Spearman's ρ=-0.24, $P<10^{-12}$). **c,** Fraction of 30kb DNA segments of the donor that bestow any phenotypic innovation on the recipient (Spearman's ρ=-0.0064, P=0.85). **d,** Probability that two recipients acquire the same phenotypic innovations from a given 30kb DNA segment (Spearman's ρ=-0.20, $P=10^{-9}$). Each circle in these figures represents the binning of 50 neighboring data points.

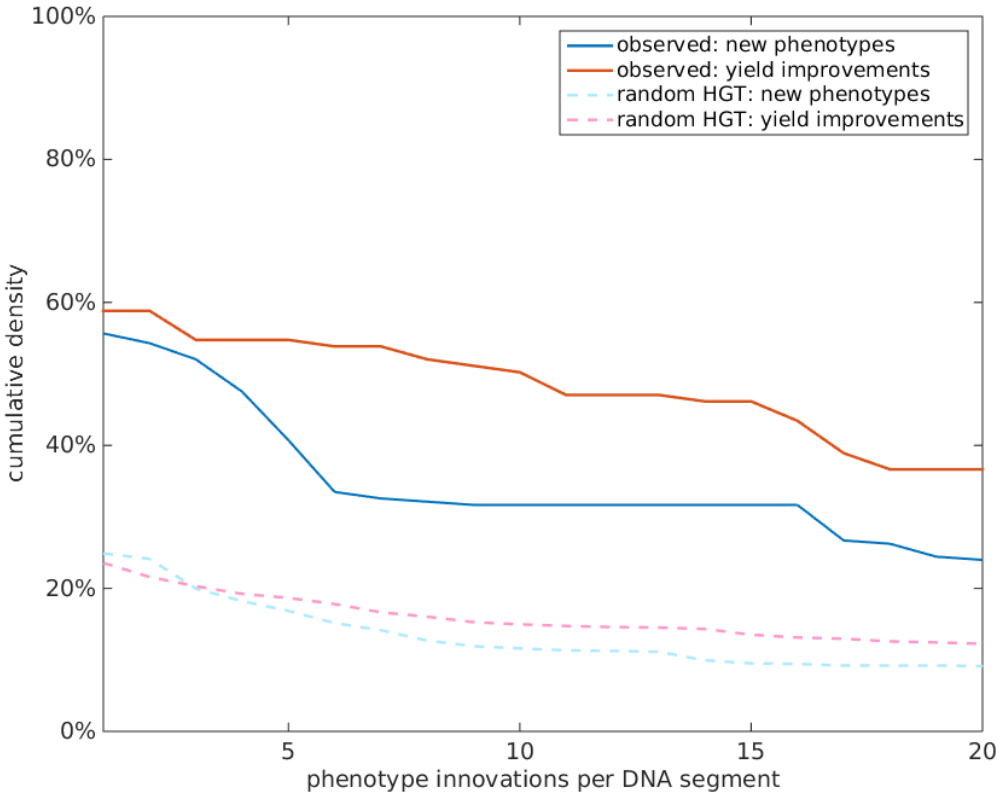

Figure S5. The successful horizontal transfer of DNA segments leads much more often to phenotypic innovations than expected for randomly chosen DNA segments found in extant *E. coli* strains. Cumulative distribution of the number of phenotypic innovations per 30kb DNA segment.

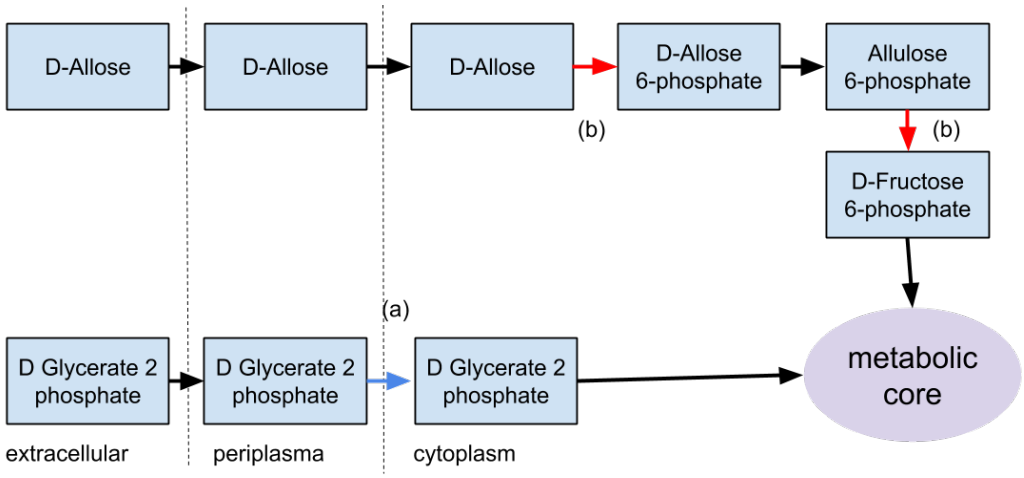

Figure S6. Illustration of a complex phenotypic innovation – the ability to grow in the environment corresponding to line 73,556 in Dataset S2, distinguished by the unusual carbon source D-allose and the unusual phosphorous source D-glycerate 2-phosphate. The innovation arose in the most recent common ancestor of the APEC-O1 and S88 strains. The three relevant enzymes catalyze the reactions marked by blue and red arrows. They were acquired in two distinct horizontal transfer events (marked by (a) and (b)) on two separate branches of the phylogeny, highlighted on the phylogenetic tree in Figure S1.

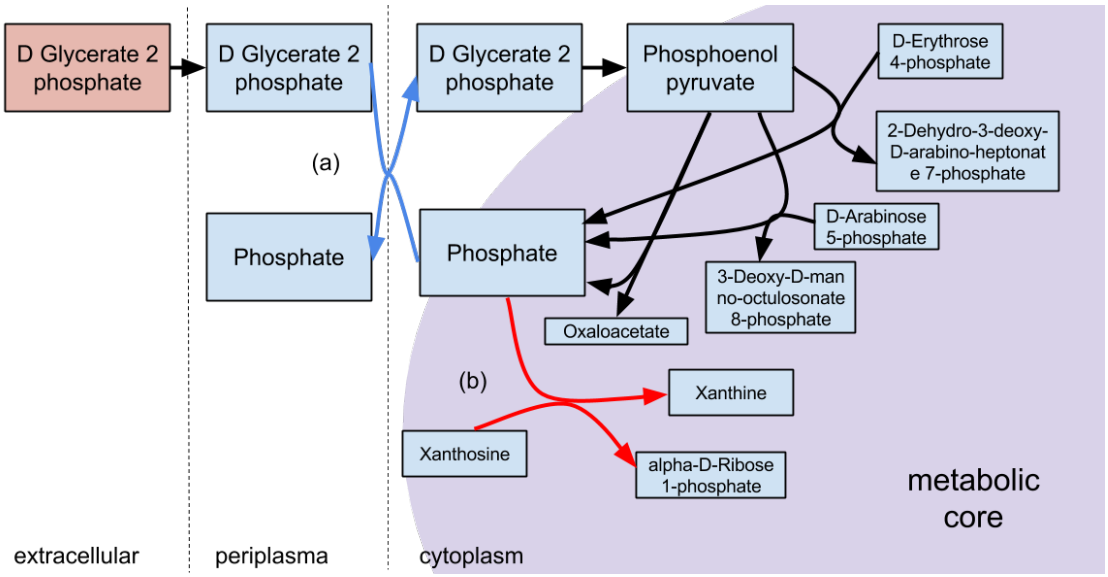

Figure S7. Illustration of a complex phenotypic innovation – the ability to grow in the environment corresponding to line 84,924 in Dataset S2, distinguished by the unusual phosphorous source D-glycerate 2-phosphate. The innovation arose in the most recent common ancestor of the APEC-O1 and S88 strains. The two relevant enzymes and transporter catalyze the reactions marked by blue and red arrows. They were acquired in two distinct horizontal transfer events (marked by (a) and (b)) on two separate branches of the phylogeny, highlighted on the phylogenetic tree in Figure S1.

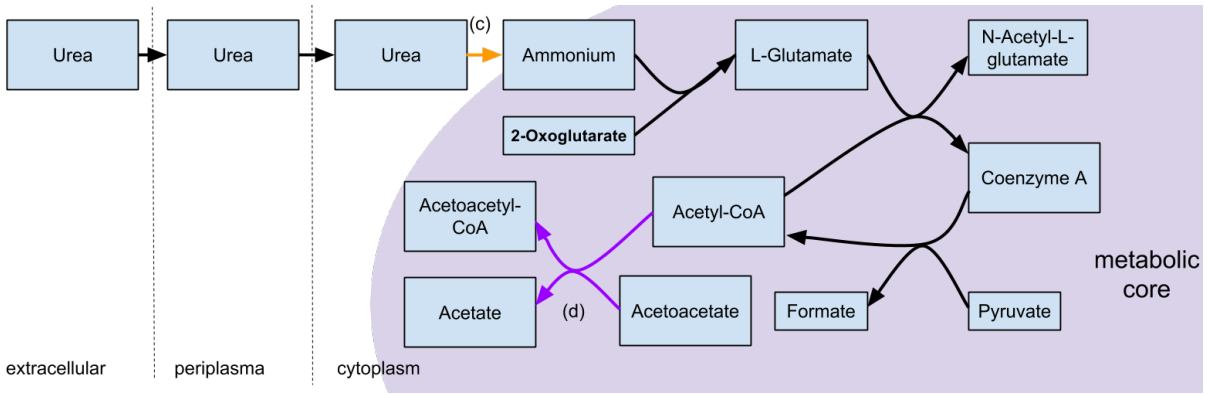

Figure S8. Illustration of a complex phenotypic innovation – the ability to grow in the environment corresponding to line 15,194 in Dataset S2, distinguished by the unusual nitrogen source urea. This innovation arose in the UMNK88 strain. The five relevant enzymes catalyze the reactions marked by orange and magenta arrows. They were acquired in two distinct horizontal transfer events (marked by (c) and (d)) on two separate branches of the phylogeny, highlighted on the phylogenetic tree in Figure S1.

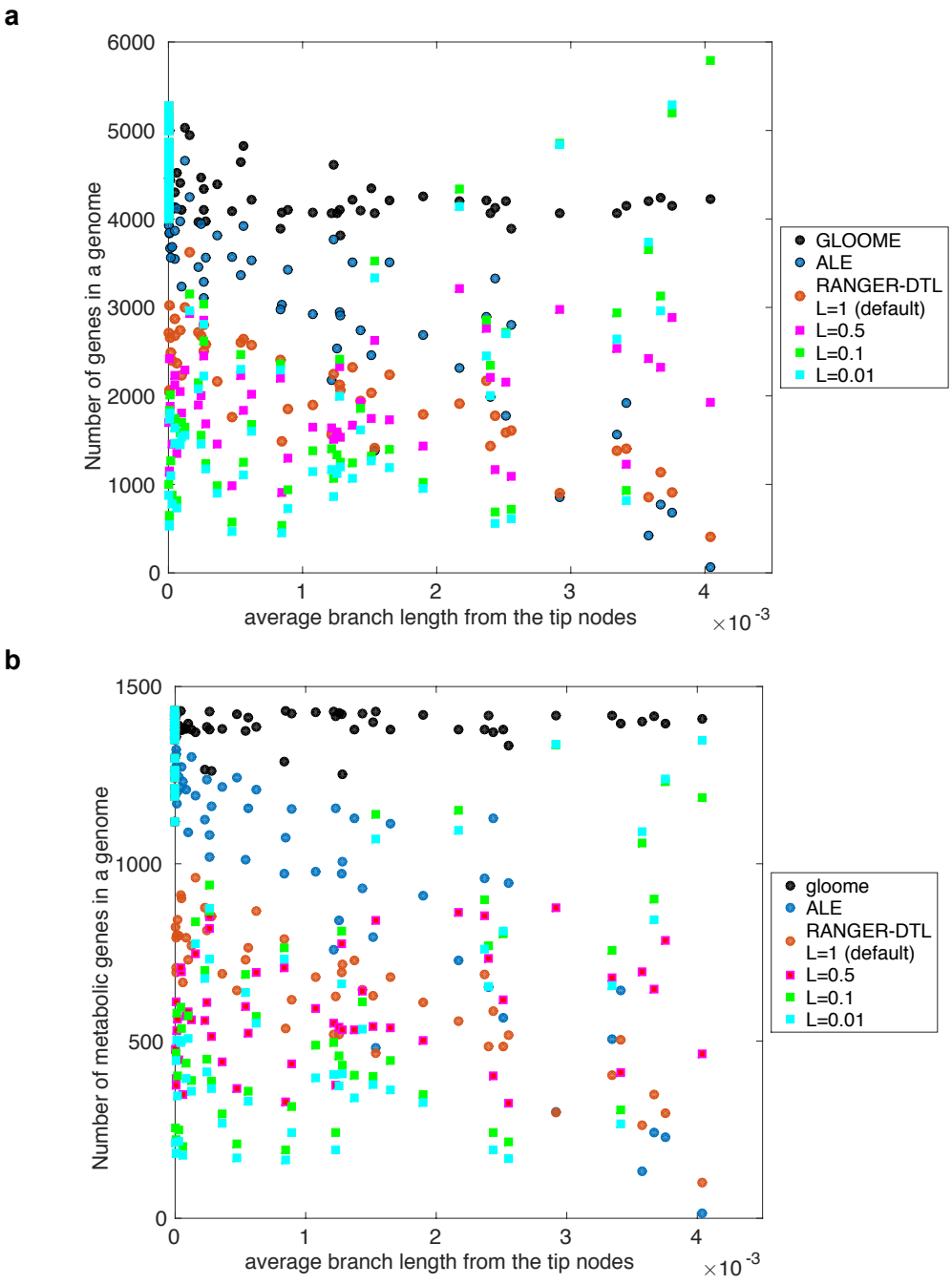

Figure S9. Inferred genome sizes **(a)** and number of metabolic genes **(b)** of ancestral strains are similar to those of extant strains when reconstructed based on gene presence/absence data (GLOOME (4), as employed in Ref. (2) and utilized here), but are increasingly reduced for older ancestors when inferred based on the reconciliation of gene trees with a tree representing vertical descent (ALE (6), RANGER-DTL (5) with default settings). While gene loss penalties lower than the default (L=1) result in larger ancestral genome predictions, these show high variance; moreover, the number of metabolic genes is still massively reduced even for the lowest tested parameter value (L=0.01). Accordingly, none of the metabolic networks predicted using ALE or RANGER-DTL is capable of producing biomass in any environment.

Datasets

Dataset S1. Details of the 53 strains, including their genbank ID, name, and numbers of genes, metabolic genes, metabolic reactions, and tested environments that support growth.

Dataset S2. Exchange reactions in each of the 202,418 nutritional environments tested in our study. The first 2,418 environments in the list come from *Szappanos et al.* (8), while the remaining 200,000 environments represent random combinations of carbon, nitrogen, phosphorous, and sulphur sources (see Dataset S5).

Dataset S3. Compiled universal gene-reaction (GPR) rules for inferring the metabolic network from the genomic presence and absence of orthologous gene families. Gene IDs in this Dataset are mapped to locus tags in the genbank files through Table S3 in Pang & Lercher (2); reaction names follows that of the metabolic models in Monk *et al.* (1).

Dataset S4. Reconstructed DNA segments containing enzymes and/or transporters that have been transferred into different branches of the *E. coli* phylogeny. Gene IDs in this Dataset are mapped to locus tags in the genbank files through Table S3 in Pang & Lercher (2).

Dataset S5. Exchange reactions of the carbon, nitrogen, phosphorous, and sulphur sources used to generate the random nutritional environments.

Dataset S6. Non-universal DNA segments present in the extant genomes containing enzymes and/or transporters, which could be the source of HGT events leading to phenotypic innovations. Gene IDs in this Dataset are mapped to locus tags in the Genbank files through Table S3 in Pang & Lercher (2).

References

1. Monk JM*, et al.* (2013) Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. *P Natl Acad Sci USA* 110(50):20338-20343.
2. Pang TY & Lercher MJ (2017) Supra-operonic clusters of functionally related genes (SOCs) are a source of horizontal gene co-transfers. *Sci Rep-UK* 7:40294.
3. Lechner M*, et al.* (2011) Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *Bmc Bioinformatics* 12:124.
4. Cohen O, Ashkenazy H, Belinky F, Huchon D, & Pupko T (2010) GLOOME: gain loss mapping engine. *Bioinformatics* 26(22):2914-2915.
5. Bansal MS, Kellis M, Kordi M, & Kundu S (2018) RANGER-DTL 2.0: Rigorous Reconstruction of Gene-Family Evolution by Duplication, Transfer, and Loss. *Bioinformatics* 34(18):3214-3216.
6. Szollosi GJ, Rosikiewicz W, Boussau B, Tannier E, & Daubin V (2013) Efficient exploration of the space of reconciled gene trees. *Syst Biol* 62(6):901-912.
7. Orth JD, Thiele I, & Palsson BO (2010) What is flux balance analysis? *Nat Biotechnol* 28(3):245-248.
8. Szappanos B*, et al.* (2016) Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nat Commun* 7:11607.
9. Gelius-Dietrich G, Desouki AA, Fritzemeier CJ, & Lercher MJ (2013) sybil - Efficient constraint-based modelling in R. *Bmc Syst Biol* 7:125.
10. R Core Team (2016) R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria).
11. Holzhütter HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur J Biochem* 271(14):2905-2922.