PNAS
www.pnas.org

Supplementary Information for

Adaptive evolution reveals a tradeoff between growth rate and oxidative stress during naphthoquinone based aerobic respiration

Anand et al.

Bernhard O. Palsson.
E-mail: palsson@ucsd.edu

This PDF file includes:

Figs. S1 to S5
Tables S1 to S4
Caption for Database S1
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Database S1

Materials and methods

Materials. *E. coli* K-12 MG1655 (ATCC 700926) was used as the wild type strain. P1 phage transduction method was used to generate the knockout strains(1) and strains from Keio collection were used as donor for the gene knockout cassettes(2). Knockouts confirmation was done by gene specific PCR and genome re-sequencing (PCR confirmation primers are given in Table S4). Bioscreen C Reader system was used for the growth profiling with 200 µL culture volume per well using a minimum of three biological replicates. Media components and Paraquat dichloride were purchased from Sigma-Aldrich (St. Louis, MO). Hydrogen peroxide was purchased from Fisher scientific (H325).

Methods.

Adaptive laboratory evolution (ALE) and DNA resequencing. ALE was performed using 3 independent replicates of $\Delta ubiC$ strain. Cultures were serially propagated on M9 minimal medium with 4 g/L glucose at 37°C and well-mixed for proper aeration using an automated system that passed the cultures to fresh flasks once they had reached an A_{600} of 0.3 (Tecan Sunrise plate reader, equivalent to an A_{600} of 1 on a traditional spectrophotometer with a 1 cm path length). Cultures were always maintained in excess nutrient condition assessed by non-tapering exponential growth. The laboratory evolution was performed for a sufficient time interval to allow the cells to reach its fitness plateau. The fitness jump was observed in about 200 generations; however, the experiment was continued for approximately 900 generations to explore the possibility of any secondary fitness jump. Further passaging was stopped due to the absence of any appreciable growth rate increase in about 700 generations. The slope of $\ln(A_{600})$ vs. time of four A_{600} measurements from each flask was used to determine the growth rate. A cubic interpolating spline constrained to be monotonically increasing was fit to these growth rates to obtain the fitness trajectory curves. DNA resequencing was performed on a clone from the end points of evolved strains as described earlier(3).

Transcriptomics. Total RNA was sampled from two biological replicates. The strains were grown in a condition same as that used during ALE. Total RNA isolation, rRNA removal and sequencing library preparation was performed as previously described(3). Libraries were ran on a HiSeq and/or NextSeq (illumina). Expression profiling was performed as previously described(4). Raw sequencing reads were mapped to the reference genome (NC_000913.3) using bowtie v1.1.2(5) with a maximum insert size of 1000 and two maximum mismatches after trimming 3 bp at 3' ends. Transcript abundance was quantified using summarizeOverlaps from the R GenomicAlignments package, with strand inversion for the dUTP protocol and strict intersection mode(6). We then estimated the dispersion and differential expression level of each gene using DESeq2(7). Transcripts per Million (TPM) were calculated by DESeq2.

I-modulon decomposition. For independent component analysis (ICA), we combined the expression profiles generated in this study with a collection of 278 expression profiles previously generated in our research group. ICA was performed as described previously(8). Briefly, the expression compendium was centered using the WT *E. coli* MG1655 expression profile reported in this manuscript as the baseline condition. We executed FastICA 100 times with random seeds and a convergence tolerance of 10^{-7}. We constrained the number of components in each iteration to the number of components that reconstruct 99% of the variance as calculated by principal component analysis. The resulting components were clustered using DBSCAN to identify robust independent components. I-modulons were extracted from independent components by iteratively removing genes with the largest absolute value and computing the D'agostino K^2 test statistic(9) of the resulting distribution. Once the test statistic fell below a cutoff of 500 (identified through a sensitivity analysis(8), we designated the removed genes an i-modulon.

Differential i-modulon activity analysis was applied to identify statistically significant differences between i-modulon activities. We first computed the distribution of differences in i-modulon activities between biological replicates, and then fit a log-normal distribution to each distribution. To test for differential activity of an i-modulon between two different conditions, we first computed the average activity of the i-modulon between biological replicates. We then computed the absolute value of the difference in i-modulon activities between the two conditions. This difference was compared against the log-normal distribution for the i-modulon to calculate a p-value. I-modulons were designated as significant if the p-value was below 0.001.

RNA-seq data for paraquat treatment (250 μM) was obtained from GSE65711(10). RNA-seq data for iron starvation (0.2 mM 2,2′-dipyridyl) and iron supplementation (0.1 mM $FeCl_2$) were obtained from GSE54900(4).

Phenotype characterization. Culture density were measured at 600 nm absorbance with a spectrophotometer and correlated to cell biomass. Samples for the substrate uptake and secretion rate were filtered through a 0.22 μm filter (PVDF, Millipore) and measured using refractive index detection by HPLC (Agilent 12600 Infinity) with a Bio-Rad Aminex HPX87-H ion exclusion column. The HPLC method was the following: injection volume of 10 μL and 5 mM H_2SO_4 mobile phase set to a flow rate and temperature of 0.5 mL/min and 45°C, respectively.

The oxygen uptake rate of each aerobic culture was determined by measuring the rate of dissolved oxygen depletion in an enclosed respirometer chamber using YSI 5300A Biological Oxygen Monitor System that utilizes Clark type polarographic oxygen probes (Cole-Parmer Instruments, Vernon Hills, IL).

Quinone extraction and estimation. The respiratory quinones were extracted following a protocol standardized earlier in the lab (11, 12). 4 ml of cultures were quenched with 6 ml of ice-cold methanol. Then, 6 ml of petroleum ether was added rapidly and vortexed for 1 minute. Next, the mixture was centrifuged for 2 minutes at 900g. 3 ml of upper phase was transferred to a fresh 15 ml tube. A second round of extraction was performed from the lower phase using 3 ml of petroleum ether. The upper

phases were combined and dried under nitrogen gas. Dried extract was re-dissolved in 100 µl methanol and analyzed using an HPLC system fitted with a XBridge BEH C18 (2.5 µm) 2.1 x 50 mm column XP (Waters). Methanol with 0.1% formic acid was used as mobile phase at a flow rate of 0.3 ml/min. Detection of quinones was performed using a UV detector at 290 nm for UQ and 248 nm for NQ. Ubiquinone-8 (Avanti Polar Lipids: 900151P) and menaquinone-7 (Sigma-Aldrich: 1381119) were used as standards. Peaks were identified by UV/Vis spectral analysis and mass spectral analysis. The relevant peak area was used to estimate the amount of each quinone species. LC-MS grade methanol, petroleum ether, and formic acid was purchased from Sigma-Aldrich.

Proteome-constrained simulation. We used the genome-scale model of metabolism and protein expression enhanced by protein-folding network (FoldME) as it is shown to offer fine-grained descriptions of the proteome composition, and predict multi-scale cellular adaptation to the genetic changes(13). We incorporated the detailed experimental characterization of the *E. coli* strains into the model to infer the underlying metabolic changes between the WT, the NQ dependent strains.

First, we constrained the model with all the experimentally quantified exchange rates and, second, we simulated the constrained model at the measured growth rate. Finally, to further capture the metabolic shift in respiration system before and after evolution, we quantified the mass fraction of transcripts involved in related pathways using the RNASeq data (Table S3). Imposing these mass fractions directly into the model may generate infeasible solutions due to inconsistency with the measured growth rate and exchange rates. Hence, we considered the fact that mass fraction of the ribosomal protein (ϕ_r) correlated linearly with growth rate(14, 15), and formulated the constraints on the mass fraction of selected pathway relative to ϕ_r as the reference:

$$\phi_r.V_{pathway} \geq \phi_{pathway}.V_r \quad (1)$$

$\phi_{pathway}$ denotes the total mass fraction of proteins involved in the corresponding pathway as calculated from the transcriptomic profile.

$$V_{pathway} = \sum_i mw_i.V_i^{translation} \quad (2)$$

$$V_r = \sum_i mw_{r-protein_i}.V_{r-protein_i}^{translation} \quad (3)$$

where mw_i and $\mathrm{V}_i{}^{translation}$ denote the molecular weight and translation flux of the ith protein in the corresponding pathway.

Lag time estimation. For each replicate, we fit the absorbance measurements obtained from the Bioscreen C reader using non-linear least squares by running the nls command on R to estimate the growth parameters of the Baranyi growth model in each condition(16). The default starting values for lag phase duration, μmax , $\log_{10}(\mathrm{N}_0)$ and $\log_{10}(\mathrm{N}_{max})$ were initially set to 4, 0.8, 0.1, 0.6. In addition, because some strains exhibited biphasic growth patterns, we excluded data points past a selected time threshold. To select time thresholds and better starting values for the growth curve parameters in an unbiased fashion, we ran a sensitivity analysis in which we computed the sum of squared errors from fitting the non-linear growth model to the absorbance measurements from time t = 0 to time t = T (with T varying from 2 hours to the total duration of the experiment). We subsequently used Findpeaks from the Pracma package(17) to find the times at which the sum of squared errors minima occur. We then selected the latest time point at which a minimum occurs and used the estimated growth parameters as the starting values for a subsequent nonlinear squares regression run. We reported the estimated parameters from the second run. We subsequently ran ANOVA to test for the significance of the differences observed in relative lag phase between treated and untreated samples in R. Calculated relative lag phase durations were subjected to a two-way analysis of variance having two levels of treatment type (paraquat and hydrogen peroxide), and two levels of evolution treatment (evolved and pre-evolved), excluding the measurements from the wild-type cells. The main effect of treatment type yielded an F ratio $F(1,67) = 157.6$, p-val < 0.001 and that of the evolution treatment yielded an F ratio of $F(1,67) = 238.995$, p-val < 0.001. The interaction was significant (p-val < 0.001). We subsequently subdivided the data set into paraquat treated conditions and hydrogen peroxide treated conditions and subjected the measurements for relative lag phase to a one-way analysis of variance having two levels of evolution treatment (evolved and pre-evolved). The main effect of evolution treatment was significant in the paraquat treated cells ($F(1,40) = 414.04$, p-val < 0.001 but not in the hydrogen peroxide treated cells ($F(1,27) = 2.39$, p-val $= 0.13$).

Computing the cause of growth rate limitation. We investigated why $\Delta ubiC$ strains did not reach the wild type growth rate even after adaptive evolution. The consequence of *ubiC* deletion is increased use of naphthoquinone (NQ). As outlined in the

main text, NQ is susceptible to losing electrons to oxygen, thus generating superoxide. From this phenomenon, we hypothesized that two separate mechanisms are responsible for lower evolved growth rate: (i) the increased cost (metabolic and protein expression) of detoxifying reactive oxygen species (ROS), and (ii) the lowered efficiency of the electron transport chain due to diversion of electrons away from cytochromes and toward superoxide. (Note that a third hypothesis is that the superoxide generated causes downstream damage to cellular components. However, we assume that the detoxification capacity is high enough that the additional ROS is fully detoxified to oxygen and water.)

To quantitatively compute growth rate subject to these two mechanisms, we used a genome-scale model of metabolism and macromolecule expression (ME)(18). The model accounts for expression (including transcription, translation, and metal cofactor incorporation), and metabolism (including ATP usage and redox balancing requirements), all in the context of the ME network reconstruction of *E. coli*.

To test hypothesis (i). we performed the following steps:
1) Create an artificial superoxide source in the periplasm by adding the reaction: $e^- + O_2 \rightarrow {O_2}^{\cdot -}$
2) Constrain (force) the reaction flux between 0 and a max value (30 mmol/gDW/h)
3) Simulate maximum growth rate

At step 1), we do not actually take electrons away from another metabolite, since we are testing only the fitness cost of detoxifying superoxide.

To test hypothesis (ii). we performed the following steps:
1) Add the reaction: $NQH_2 + 2\ O_2 \rightarrow NQ + 2\ {O_2}^{\cdot -} + 2\ H^+$
(where NQH_2 is menaquinol 8, and NQ is menaquinone 8 in the model)
2) Constrain (force) the reaction flux between 0 and a max value (30 mmol/gDW/h)
3) Simulate maximum growth rate

In both simulations, we compute growth rate versus the "leak percent" defined as
Leak percent = v_{leak} / (v_{leak} + $v_{cytochrome}$) x 100%
where v_{leak} is the flux of electrons leaked to superoxide due to the added reaction in mechanism (i) or (ii), and $v_{cytochrome}$ is the sum of total flux of electrons directed to the three cytochrome oxidases.

Computation of the cost of naphthoquinone and ubiquinone biosynthesis . We computed the cost of synthesizing NQ and UQ using the genome-scale model of *E. coli* metabolism, iML1515(19). We defined cost as the moles of ATP, carbon, or oxygen required to synthesize one mole of NQ/UQ.

We performed the following steps:
1) Add an artificial sink reaction for NQ (menaquinone-8) or UQ (ubiquinone-8)
2) Set the max glucose uptake rate to 10 mmol/gDW/h
3) Compute maximum flux of the sink reaction, v_Q using parsimonious flux balance analysis (pFBA) (pFBA is used to compute the most efficient ATP, carbon, and oxygen usage to make NQ or UQ)
4) Compute costs (where v_Q is the maximum sink flux):
(a) ATP (turnover) cost = $\sum_{j \epsilon ATP_{Used}} v_j / v_Q$,
where ATP_{Used} is the set of reactions consuming ATP. (Note that due to mass balance, this sum of fluxes is equivalent to the sum of reactions of producing ATP, or half the sum of absolute fluxes that consume or produce ATP.)
(b) Carbon cost = 6 carbon x glucose uptake rate / v_Q
(c) Oxygen cost = 2 oxygen x oxygen uptake rate / v_Q
For all computations, we removed the non-growth associated ATP maintenance requirement by setting its lower bound to 0.

Data availability. Resequencing and expression profiling data that support the findings of this study have been deposited to NCBI Sequence Read Archive (SRA accession: PRJNA560068) and Gene Expression Omnibus (GSE135867) respectively.

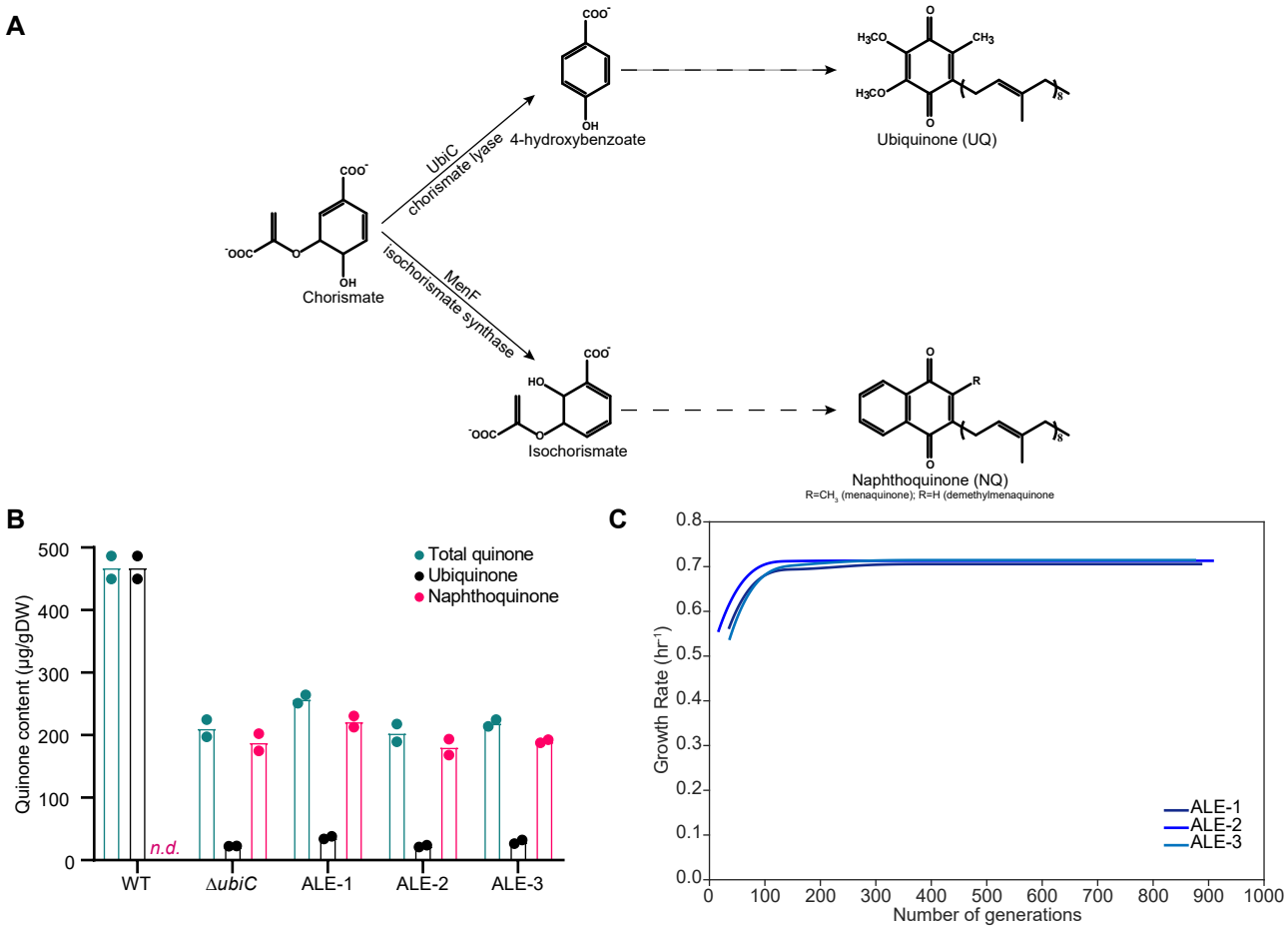

Fig. S1. (A) Part of respiratory quinone biosynthetic pathway highlighting the chorismate node. (B) Estimate of respiratory quinones. Bar height represents the average to replicates shown individually as dots. 'n. d.' stands for not detected. (C) The extended axis plot of the growth rate evolution trajectories of Δ*ubiC* replicates corresponding to figure 1B.

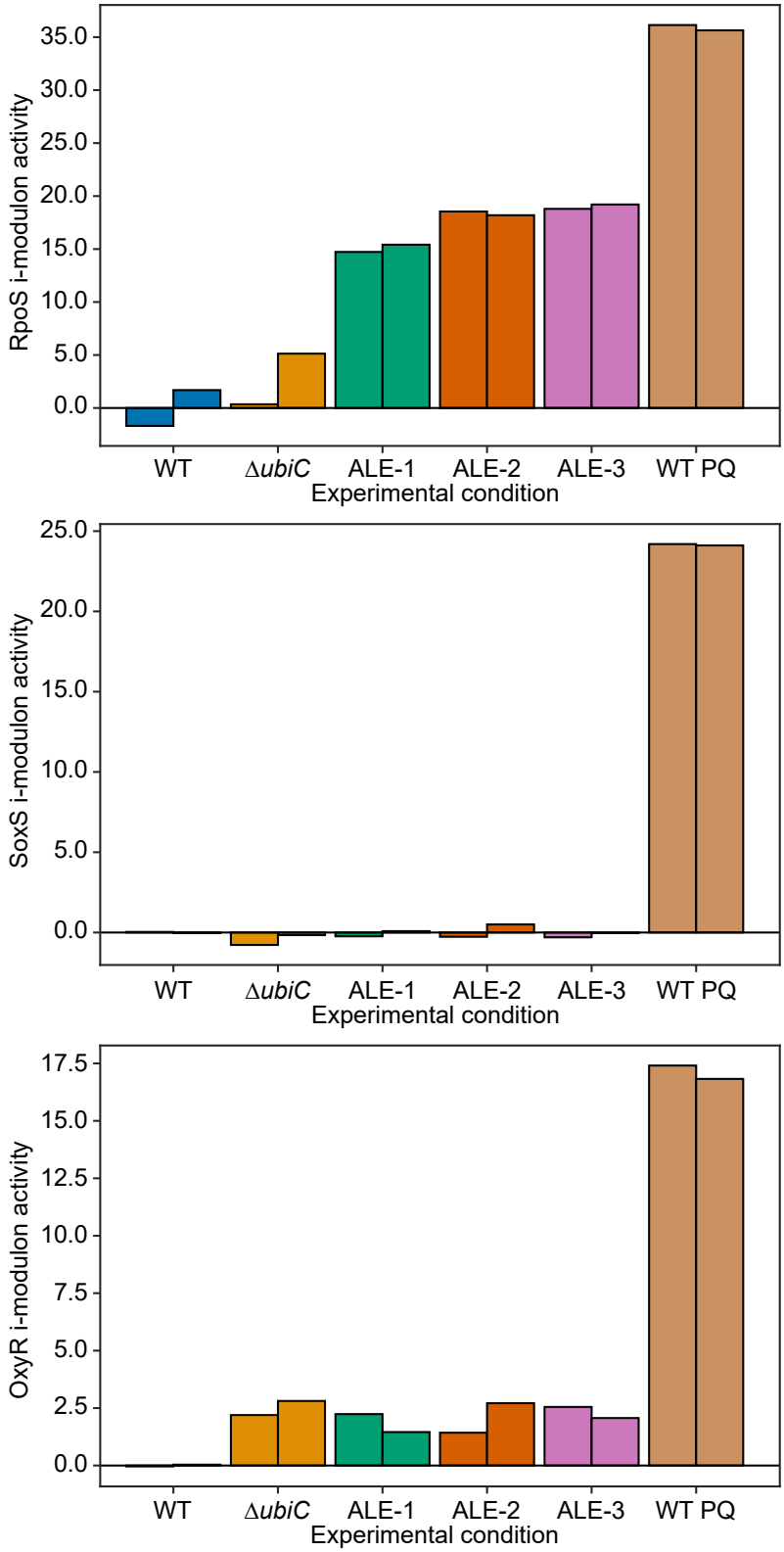

Fig. S2. ICA for the activity of ROS defense related i-modulons. The bars with identical colors represent biological replicates of the corresponding strain.

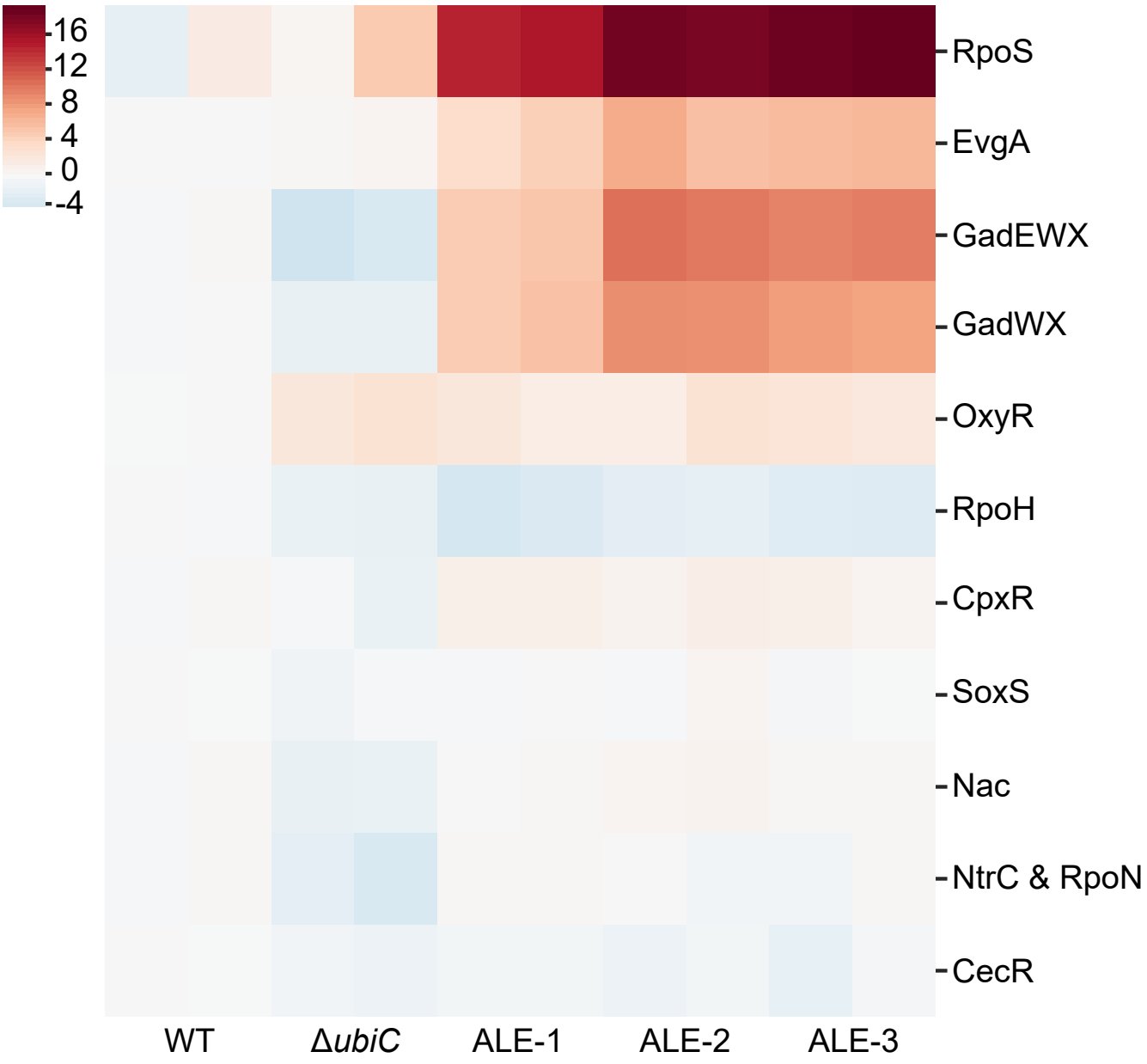

Fig. S3. Heatmap showing activities of stress i-modulons. Biological replicates are shown individually.

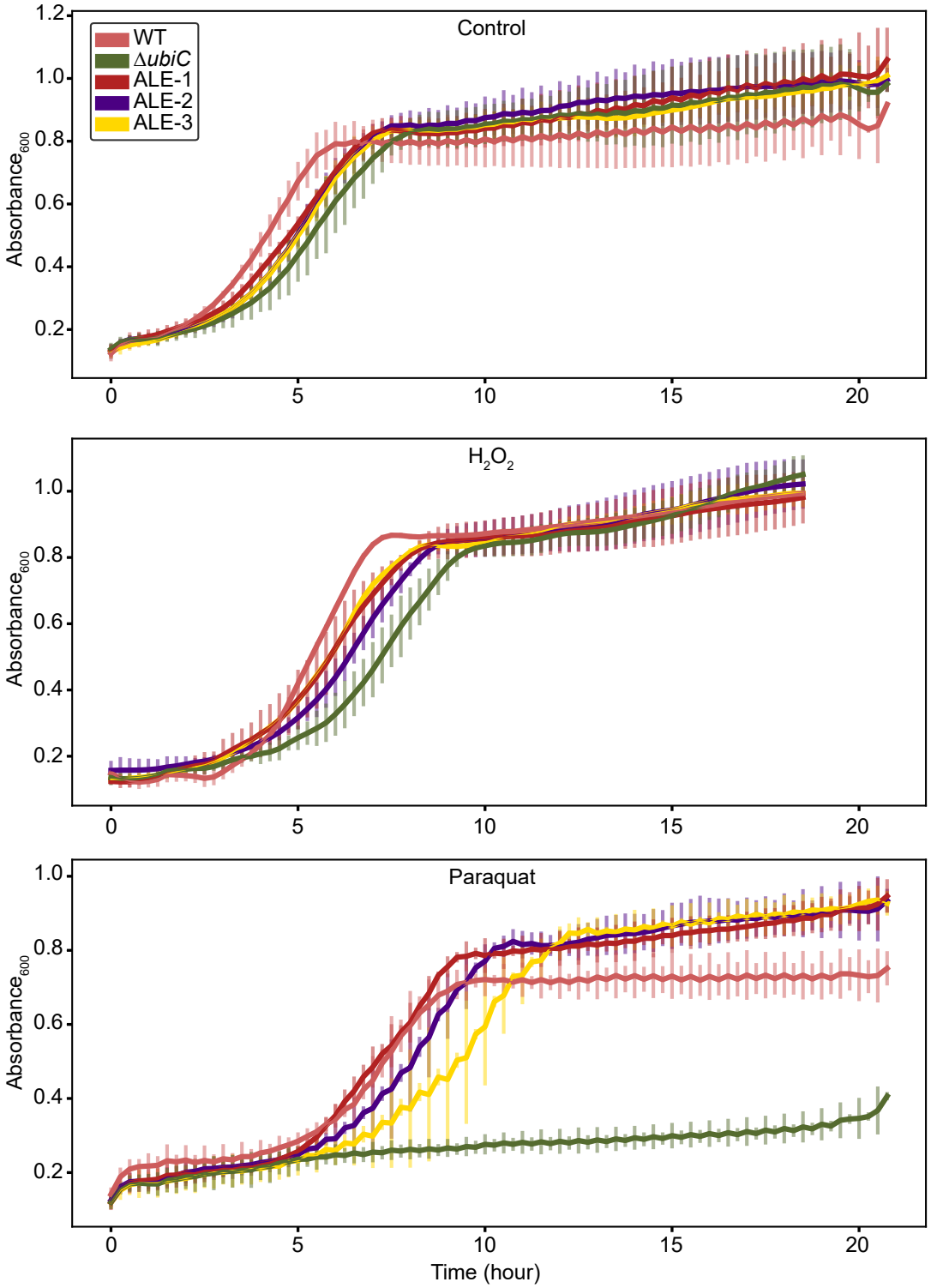

Fig. S4. Growth curve of the strains treated with peroxide (1 mM Hydrogen peroxide) and superoxide (1μM Paraquat) generating chemicals.

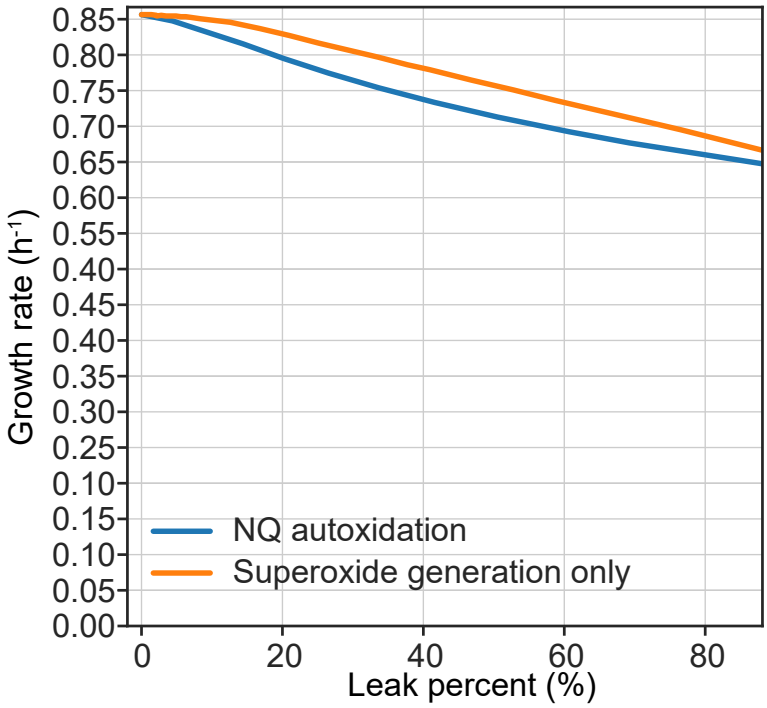

Fig. S5. The extended axis plot corresponding to figure 3D showing genome-scale model-based calculation of the impact of periplasmic non-productive electron leak on the growth rate of E. coli.

Table S1. List of mutations observed in evolved strains

Strain	Mutations		
ALE-1	*pdhR* (G→T E37*)	*ubiE* (T→G)	
ALE-2	*aroP, pdhR* (T→C)	*yoeI, yeeY* [(G)7→8]	*yfaY* (G→A)
ALE-3	*pdhR* (G→T G62C)	*mhpE* [(GCG)3→2]	

In the $\Delta ubiC$ strain, an 82-bp deletion between $pyrE$ and rph occurred during the strain construction and thus appeared in all ALE endpoint strains. This deletion is commonly found in ALE of *E. coli*(20). This metabolic mutation has been shown to relieve a defect in pyrimidine biosynthesis present in the WT strain that improves the growth rate(21, 22). This explains the difference between the reported(23) and observed growth rate of the pre-evolved $\Delta ubiC$ strain.

Table S2. Estimate of the biosynthetic cost of respiratory quinone

Cost parameter	Q = UQ	Q = NQ
mole ATP/mole Q	53.4996	42.5
mole Glucose/mole Q	13.0421	12.7549
mole O_2/mole Q	12.7524	8.52967
mole Carbon/mole Q	78.2524	76.5296

Table S3. Mass fraction of major metabolic pathways

Strain	r-protein	Glycolysis	oxPPP	TCA	oxPhos	Pyruvate
WT	0.1600	0.0489	0.0103	0.0282	0.0431	0.0082
$\triangle$*ubiC*	0.1321	0.0596	0.0114	0.0232	0.0367	0.0072
ALE-1	0.1318	0.0524	0.0115	0.0272	0.0415	0.0081
ALE-2	0.1292	0.0553	0.0126	0.0253	0.0420	0.0096
ALE-3	0.1214	0.0611	0.0127	0.0232	0.0394	0.0097

Table S4. List of primers used in this study

Kanamycin cassette specific primers	
k1	CAGTCATAGCCGAATAGCCT
k2	CGGTGCCCTGAATGAACTGC
Gene specific primers	
ubiC (U)	CTGGCATCCTGGACGGTGAT
ubiC (D)	CCGGCAGCGCGCATCAGCCA
pdhR (U)	GTGAATCGGTTCAATTCGGA
pdhR (D)	AACACCTTCTTCACGGATGA

(U: upstream primer; D: downstream primer)

Additional data table S1 (Supplementary_table.pdf)
i-modulon genes

References

1. LC Thomason, N Costantino, DL Court, E. coli genome manipulation by p1 transduction. *Curr. protocols molecular biology* **79**, 1–17 (2007).
2. T Baba, et al., Construction of escherichia coli k-12 in-frame, single-gene knockout mutants: the keio collection. *Mol. systems biology* **2** (2006).
3. A Anand, et al., Pseudogene repair driven by selection pressure applied in experimental evolution. *Nat. microbiology* **4**, 386 (2019).
4. SW Seo, et al., Deciphering fur transcriptional regulatory network highlights its complex role beyond iron metabolism in escherichia coli. *Nat. communications* **5**, 4910 (2014).
5. B Langmead, C Trapnell, M Pop, SL Salzberg, Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome biology* **10**, R25 (2009).
6. M Lawrence, et al., Software for computing and annotating genomic ranges. *PLoS computational biology* **9**, e1003118 (2013).
7. MI Love, W Huber, S Anders, Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology* **15**, 550 (2014).
8. AV Sastry, et al., The escherichia coli transcriptome consists of independently regulated modules. *bioRxiv*, 620799 (2019).
9. RB D'agostino, A Belanger, RB D'Agostino Jr, A suggestion for using powerful and informative tests of normality. *The Am. Stat.* **44**, 316–321 (1990).
10. SW Seo, D Kim, R Szubin, BO Palsson, Genome-wide reconstruction of oxyr and soxrs transcriptional regulatory networks under oxidative stress in escherichia coli k-12 mg1655. *Cell Reports* **12**, 1289–1299 (2015).
11. M Bekker, et al., Changes in the redox state and composition of the quinone pool of escherichia coli during aerobic batch-culture growth. *Microbiology* **153**, 1974–1980 (2007).
12. VA Portnoy, et al., Deletion of genes encoding cytochrome oxidases and quinol monooxygenase blocks the aerobic-anaerobic shift in escherichia coli k-12 mg1655. *Appl. Environ. Microbiol.* **76**, 6529–6540 (2010).
13. K Chen, et al., Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation. *Proc. Natl. Acad. Sci.* **114**, 11548–11553 (2017).
14. M Scott, CW Gunderson, EM Mateescu, Z Zhang, T Hwa, Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099–1102 (2010).
15. M Scott, S Klumpp, EM Mateescu, T Hwa, Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol. systems biology* **10**, 747 (2014).
16. J Baranyi, TA Roberts, A dynamic approach to predicting bacterial growth in food. *Int. journal food microbiology* **23**, 277–294 (1994).
17. HW Borchers, Pracma: practical numerical math functions. *R package version* **1** (2015).
18. L Yang, MA Saunders, JC Lachance, BO Palsson, J Bento, Estimating cellular goals from high-dimensional biological data. *arXiv* (2018).
19. JM Monk, et al., iml1515, a knowledgebase that computes escherichia coli traits. *Nat. biotechnology* **35**, 904 (2017).
20. RA LaCroix, et al., Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of escherichia coli k-12 mg1655 on glucose minimal medium. *Appl. Environ. Microbiol.* **81**, 17–30 (2015).
21. KF Jensen, The escherichia coli k-12" wild types" w3110 and mg1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyre expression levels. *J. bacteriology* **175**, 3401–3407 (1993).
22. TM Conrad, et al., 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 biology* **10**, R118 (2009).
23. A Nitzschke, K Bettenbrock, All three quinone species play distinct roles in ensuring optimal growth under aerobic and fermentative conditions in e. coli k12. *PloS one* **13**, e0194699 (2018).