Rapid continuous evolution of gene libraries towards arbitrary functions

Alexander (Olek) Pisera[1,2], Alireza Tanoori[1,2], Chang C. Liu[1-4]*

[1]Department of Biomedical Engineering, University of California; Irvine, CA, 92617, USA

[2]Center for Synthetic Biology, University of California; Irvine, CA, 92617, USA

[3]Department of Chemistry, University of California; Irvine, CA, 92617, USA

[4]Department of Molecular Biology & Biochemistry, University of California; Irvine, CA, 92617, USA

*Corresponding author. E-mail: ccl@uci.edu

Abstract

The emergence and evolution of new gene functions is central to biology, yet experimental tools that allow us to prospectively probe and understand this process are lacking. While systems for continuous directed evolution have successfully compressed gene evolution onto laboratory timeframes at scale, past work largely focused on the evolution of single genes towards single functions. In contrast, when nature searches for new gene functions, it does so by exposing a vast repertoire of genes to a multitude of complex selective pressures *in vivo*, multiplying the possibilities for innovation. To mimic natural gene evolution in laboratory time, we introduce OrthoRep Assisted Continuous Library Evolution (ORACLE). In ORACLE, high-diversity function-agnostic gene libraries (*e.g.*, open reading frames from *Saccharomyces cerevisiae* and *Escherichia coli*, cDNA from *Arabidopsis thaliana* and human cells) are encoded onto OrthoRep's p1 plasmid in yeast where library members durably mutate up to one-million times faster than the host genome. As a result, ORACLE provides an opportunity to quickly observe the evolutionary emergence and optimization of new gene functions that satisfy complex *in vivo* selection pressures from an abundance of sources. Using ORACLE, we discovered a wealth of novel gene variants conferring new cellular functions, each experimentally evolved in fewer than ~100 generations corresponding to less than one month of serial passaging. In several cases, genes that initially exhibited little to no observable phenotype rapidly evolved to confer transcription factor activity, chaperone activity, inhibition of specific targets, etc. These findings demonstrate the relative ease with which existing genetic material can enter the view of selection and evolve new function *in vivo*. Moreover, our work establishes ORACLE as an effective system for exploring the fundamental mechanisms of gene innovation and discovering bespoke biomolecules for diverse applications.

Introduction

Life necessitates frequent biomolecular innovations to keep up with changing environments. How new gene functions emerge and evolve is therefore a fundamental question in biology. Analysis of the evolutionary record has provided abundant evidence that biomolecular innovation largely occurs through gene duplication and divergence[1–4], for example following the IAD (innovation, amplification, divergence) model[5,6] wherein a gene's secondary activities confer a weak fitness benefit when encountering a new environment (innovation). The gene then undergoes duplication (amplification) to increase its secondary activities by higher expression and subsequently evolves unconstrained by its primary function due to the functional redundancy afforded by duplication (divergence). Another model for biomolecular innovation is *de novo* gene birth[7–11]. Omics data suggests that intergenic sequences, classically considered non-coding, can be transcribed and translated[12–14]. Similarly, intronic sequences or alternative frame segments of existing ORFs[11,15] may be aberrantly expressed. Such "non-coding" sequences may

thus be a source of new traits that result in their evolution into *bona fide* new genes. Gene birth and IAD can of course both occur, and other mechanisms such as horizontal gene transfer (HGT), pseudogenization, and evolution from pseudogenes into new genes augment these models.

Such models explaining the origination and evolution of new gene functions have motivated an important but limited and challenging set of prospective laboratory evolution experiments aimed at observing key steps in action. For example, Näsvall *et al.* showed that a histidine biosynthesis enzyme (HisA) could evolve weak tryptophan biosynthesis activity (normally achieved by TrpF) to yield a bifunctional enzyme[6]. Under continuous selection for histidine and tryptophan biosynthesis, the bifunctional enzyme then evolved *via* amplification/duplication and divergence, generating two distinct genes – one with strong HisA activity and one with strong TrpF activity – by the end of a 3000-generation evolution experiment that captured all stages of IAD. However, HisA and TrpF carry out the same isomerization reaction (on different substrates) and likely naturally evolved from a generalist ancestor[16], so the ability for HisA to find TrpF activity is not entirely surprising. To explore whether existing genes can evolve arbitrary functions, Soo *et al.* subjected a synthetic library of *Escherichia coli* genes to selections for antibiotic and toxin resistance and found a surprisingly high number of examples where an overexpressed gene was able to directly or indirectly confer some degree of resistance[17]. Likewise, experiments that test the evolutionary productivity of classically “non-coding” or *de novo* generated genes of varying degrees of randomness[18–27] also support the idea that perhaps more genetic matter than previously thought can be the origin of new genes. However, only a subset of such experiments demonstrated the evolutionary improvement of newly discovered genes through Darwinian cycles of mutation and selection[22,23,26]. Nevertheless, an overall model is emerging where 1) the wealth of genetic material that can evolve after duplication, HGT, or gene birth, 2) the commonness of secondary activities in extant genes (*e.g.*, moonlighting[28–30] or promiscuous[31–34] activity), and 3) the systems level complexity of cells that can afford many ways to achieve new selectable phenotypes combine to make the evolution of novel gene function frequent.

Inspired by this growing appreciation for the versatility of gene evolution in systems contexts, the difficulty of investigating extensive gene evolution experimentally (especially at scale), and the development of continuous gene evolution platforms that operate *in vivo* (where cellular systems translating biomolecular activity to organismal phenotypes remain intact)[35,36], we present OrthoRep Assisted Continuous Library Evolution (ORACLE) (**Figure 1**). In ORACLE, highly diverse, multi-length libraries of genes (*e.g.*, comprehensive collections of *E. coli*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, and *Homo sapiens sapiens* protein-coding genes) are encoded onto the p1 plasmid of OrthoRep[37–39], which accelerates the mutation and evolution of genes of interest (GOIs) at rates up to 1-million-fold higher than genomically encoded genes in the yeast, *S. cerevisiae*[39]. ORACLE was developed to be scalable, flexible and easy to use, enabling dozens of parallel evolution experiments from gene libraries. When the resulting populations were exposed to cell-based selections for new phenotypes, ORACLE quickly yielded multiple examples of gene evolution at scale within <100 generations, corresponding to only <1 month of serially passaging yeast cultures. In the most striking cases, a gene without a known function and whose wild type (WT) sequence had little or no effect on fitness rapidly evolved to satisfy new selection pressures. Experiments characterizing discovered evolutionary outcomes suggest their function as novel transcription factors, protein-protein interaction inhibitors, metal binders, *etc*.

Taken together, ORACLE provides the first platform capable of subjecting diverse, function agnostic, variable-length, high complexity genetic libraries to continuous directed evolution *in vivo*, resulting in the rapid emergence and evolution of novel functions on short laboratory timescales. With this capacity, ORACLE should advance the experimental investigation of evolutionary mechanisms leading to new

genes, the directed evolution of new biomolecules with arbitrary desired functions, and the study of gene evolution's flexibility in general.

Results

Basic template for ORACLE experiments with "subroutine" names

We first amplified complex gene libraries (**PCR**$_{lib}$) from various plasmid gene library collections and subcloned those PCR products into OrthoRep integration plasmids[40]. All ORACLE experiments began from these integration plasmids, which were used to integrate their contained gene libraries onto OrthoRep's hypermutating p1 in *S. cerevisiae* stains (INTEGRATION$_{p1}$ or **INT**$_{p1}$). This was done either directly into the selection strain for evolving the gene library or into a general donor strain that was abortively mated with a selection strain to transfer the cytoplasmic p1s *en masse* (**MATE**). Once a gene library was in the selection strain, continuous growth-based evolution experiments were carried out (**EVO**). Throughout **EVO** from time to time, the p1 plasmids comprising the population of evolving genes were amplified (**PCR**) and transferred into fresh selection strains by **INT**$_{p1}$ or directly transferred by **MATE** to prevent the accumulation of any genomic mutations that could confound gene adaptation. At the end of each evolution experiment, the population of outcome genes was isolated by **PCR**, sequenced, and integrated into a fluorescently marked selection strain at a non-hypermutating genomic locus (**INT**$_{genome}$). The non-mutating population was then competed against a different fluorescently marked strain that had a nonsense gene encoded (COMPETITION$_{population/NS}$ or **COMP**$_{pop/NS}$) to confirm that the evolved gene population was responsible for adaptation. If so, competition of the evolved gene population, which was always descended from a single gene from the initial gene library at this endpoint of evolution, was repeated against the unevolved WT gene (**COMP**$_{pop/WT}$) to assess whether the population of evolved genes gained enhanced function relative to their unmutated ancestor. A sample of mutant clones were isolated, sequenced, and subjected to competitive growth experiments to confirm their individual fitnesses against control strains expressing nonsense (**COMP**$_{mut/NS}$) or corresponding WT genes (**COMP**$_{mut/WT}$). Finally, custom experiments for mutant clones were carried out to dissect mechanisms by which they achieved their new functions.

The common "subroutines" of **PCR**$_{lib}$, **PCR**, **INT**$_{p1}$, **INT**$_{genome}$, **MATE**, **COMP**$_{pop/NS}$, **COMP**$_{pop/WT}$, **COMP**$_{mut/NS}$, and **COMP**$_{mut/WT}$ required development of customized genetic technologies, which we describe in **Box 1** and **Fig. S1-4**. In the main text to follow, we "call" these subroutines when appropriate.

Box 1: ORACLE subroutines

PCR$_{lib}$

PCR required amplifying diverse multi-sized libraries of genes, which risked introducing size biases and chimeric recombinants among genes. We therefore adopted emulsion PCR to minimize these risks, optimizing the process by titrating BSA to balance emulsion stability against PCR inhibition (**Fig. S1A**), titrating DNA to verify an approximate micelle number, and ensuring emulsion integrity by differential amplification of short and long amplicons[41] (**Fig. S1B**). During gene library preparation where gene libraries were amplified from plasmid library sources[42–44], DNA titration was used to find a concentration where template-containing droplets were abundant and where most template-containing droplets contained a single template plasmid, reducing size-based PCR biases relative to pooled PCR (**Fig. S1C**). After PCR of gene libraries, we used Gibson assembly[45] to subclone into a p1 integration vector redesigned with I-Ceu-I sites (to avoid common restriction sites) for linearization and exposure of TP901 attachment sites for integration onto p1[40]. In the current work, five DNA libraries were each subcloned into two p1 integration vectors, one of which encodes HIS3 and the other of which encodes LEU2 for selection of p1's after linearization and integration, for a total of ten p1 integration vector libraries. The five distinct libraries represented are 1) a yeast ORF library[42], 2) an *E. coli* ORF library[43], 3) an *E. coli* ORF library where each member is fused to a GFP tag[43], 4) a cDNA library from *A. thaliana*, and 5) a cDNA library from human cells.

INT$_{p1}$

We previously reported a high-efficiency method for integrating genes onto p1 landing pads[40]. The method consists of a p1 landing pad with attachment sites for TP901, donor DNA that is flanked by corresponding attachment sites (*i.e.*, linearized integration vectors), and the transient introduction of TP901, which carries out the high-efficiency recombination of donor DNA onto p1. The product of integration was selected for using markers contained in the donor DNA, specifically HIS3 or LEU2. The high transformation efficiency of TP901-mediated integration using fast and routine protocols ensured that genetic diversity of libraries and throughout evolution experiments was almost certainly oversampled during **INT$_{p1}$** steps.

INT$_{genome}$

Similar to **INT$_{p1}$**, routine integration into a genomic locus is possible using TP901. Additionally, because of yeast's efficient nuclear recombination mechanisms, overlapping fragments can be transformed, where they will assemble into the genomic locus *in vivo*. To enable rapid characterization of evolved gene populations at a non-mutating genomic locus at scale, we leveraged both TP901 and yeast nuclear recombination (**Fig. S2A**). For every evolved gene population we wished to characterize, we used **PCR** for amplification from p1, attaching ~60 bp overlaps to common flanks in the process. The 5' gene overlap was homologous to the 3' overlap of a synthetic DNA fragment that encoded a constitutively active nuclear promoter (usually pRPL18B[46]). The 3' gene overlap was homologous to the 5' overlap of a synthetic DNA fragment that encoded a selection marker (*e.g.*, *HIS3*). The 5' end of the nuclear promoter fragment contained the 5' TP901 attachment site and the 3' flank of the selection marker fragment contained the 3' TP901 attachment site. These three fragments were transformed into strains engineered with a genomic landing pad locus with corresponding TP901 attachment sites and containing a plasmid expressing TP901 recombinase[47]. Fragments reliably recombined and integrated at the TP901 attachment sites when transformed, resulting in ~100,000 successful transformants per routine experiment where each transformant is a single gene mutant encoded at a copy number of one.

COMP$_{X/Y}$ (where X and Y are what are being competed against each other)

Given the large scale at which we ran ORACLE evolution experiments, we needed streamlined characterization pipelines to accurately measure the relative performance of evolving populations against appropriate controls. To satisfy this need, we developed a high-resolution competition assay. In this assay, variable gene fragments from p1 are amplified by **PCR** and integrated by **INT$_{genome}$** into two otherwise isogenic yeast selection strains marked with two different fluorescent markers, GFP and RFP (**Fig S2A**). One strain (*e.g.*, GFP-tagged strain) then receives evolved gene populations or an evolved gene mutant, while the other (*e.g.*, RFP-tagged strain) is transformed with a control fragment (*e.g.*, a short nonsense sequence or the WT ancestor of an evolved gene). Mixed cultures (1:1 ratio) are then grown under selection for 24h per passage (**Fig. S2A-B**). The initial ratio and final ratio of GFP versus RFP cells are determined by flow cytometry (representative flow plots in **Fig. S2C**) and population sizes are measured by optical density (OD_{600}). Given time, change in OD_{600}, and GFP versus RFP ratios, we calculate an apparent growth rate constant for each strain (see **Methods**). We call this an apparent growth rate constant, because growth may not be continuously exponential over 24h. Since these growth rate measurements are done in pairwise competition against a control, comparison of an evolved gene population or evolved gene mutant against the control is expected to be maximally reliable. **INT$_{genome}$** was confirmed to generate pure populations of yeast overexpressing the gene of interest (**Fig. S2D**) ensuring that competition is among pure populations. Additionally, we often also compete strains containing nonsense fragments against each other, *i.e.* **COMP$_{X/Y}$**. This is to ensure that the two differently marked fluorescent strains have the same growth rate (*i.e.*, GFP or RFP expression affects both strains equivalently).

MATE

We developed an abortive mating-based system to efficiently transfer p1 plasmids into new hosts, for example to move gene libraries into selection strains or refresh p1's host genome background during evolution experiments. OrthoRep's p1 plasmid is maintained cytoplasmically. Abortive mating allows yeast cytoplasms to mix while avoiding nuclear fusion due to a defective Kar1 allele[48–50] (**Fig. S3**). Abortive mating is expected to produce three cell types containing the p1 plasmids of the donor: the original donor haploid, spurious diploids, and the recipient haploid, approximately in a 2:1:2 ratio[50]. We selected for the desired recipient haploid population containing p1 plasmid using a combination of positive selections on auxotrophic and/or antibiotic resistance markers along with layered counterselections (*e.g.*, thialysine, canavanine, 5-FC, 5-FOA, 5-FAA) to eliminate donor cells and diploid "cheaters"[51–55], as diagramed in **Fig. S3** and detailed in **Methods**. Note that the donor and recipient strains are opposite mating types such that each mating round changes the mating type of the host cell.

MATE may also be useful for evolution experiments where multiple p1 species encoding different genes are maintained by distinct markers for gene coevolution. This is because **MATE** allows a mixture of p1s from one cell to transfer into another cell without changing the composition of the mixture. Although we do not carry out coevolution of multiple p1-encoded genes in this work, we have validated that cells can stably maintain and express multiple p1 plasmids selected using different markers where sequential **MATE** is used to generate populations containing p1 mixtures in each cell (**Fig. S4**) at high enough efficiency where all pairwise combinations of ORF libraries can be constructed (**Fig. S5**). For detection of extremely rare events (*e.g.*, a specific pair of genes) it was necessary to apply a technique called MERGE[56], wherein the p1 destination strain contains a Cas9 plasmid and guide targeting the locus to be converted to match the destination strain (**Fig. S5**).

PCR + INT$_{p1}$

An alternative to **MATE** as a means of refreshing p1's host genome background during **EVO** is to rely on **PCR** + **INT$_{p1}$** (**Fig. S6A**). By **PCR** amplifying p1's evolving gene locus with the addition of flanks that include attachment sites for TP901-mediated integration, it was straightforward to then use **INT$_{p1}$** to directly reintegrate the amplified gene population into fresh selection strains containing 1) a p1 landing pad with TP901 attachment sites and 2) a transiently maintained plasmid expressing TP901. PCR bias (especially against longer genes) and formation of recombinants were minimized by reducing PCR cycle numbers[41]. An advantage of **PCR** + **INT$_{p1}$** is that nonfunctional p1 genes that can be generated from and hitchhike with evolving functional genes in the same cell[57]– p1 is multicopy – may be rapidly purged after **INT$_{p1}$**, as only one or a few copies of donor DNA are integrated into each cell during TP901-assisted integration onto p1[40]. The use of **PCR** + **INT$_{p1}$** from time to time during **EVO** may therefore not only afford a procedurally straightforward way to refresh p1's host genome background but may also enable faster selective sweeps by beneficial mutants.

Optimizations that streamlined **PCR** + **INT$_{p1}$** so that it could be carried out rapidly at scale included using GC-prep isolation to extract p1 populations[58], quarter-scale transformations (~100,000 CFU per transformation), and frozen competent cell preparations[59]. Together, these optimizations allowed for up to 96 **EVO** campaigns that included **PCR** + **INT$_{p1}$** steps to be carried out concurrently by one researcher.

Testing of the ORACLE experimental pipeline

To test whether the setup steps leading to **EVO** in the ORACLE experimental pipeline were reliable, we first carried out a mock ORACLE experiment where we simply sought to enrich a known selectable gene from a p1-encoded yeast ORF library (**Fig. 1C**). Specifically, we used **PCR$_{lib}$** to construct a comprehensive plasmid library of ~5,200 yeast ORFs and dubious ORFs pooled from a corresponding arrayed library originally cloned by Gelperin et al.[42] *ADE2*, *TRP5*, *URA3*, *HIS3*, *KAR1*, *LYS2*, *CAN1*, *LEU2*, *LYP1*, *MET15* were omitted from the pool due to our reliance on these genes as auxotrophic markers for OrthoRep components, as selection systems for various **EVO** experiments, or for various subroutines such as **MATE**. **INT$_{p1}$** was used to install the gene library into an OrthoRep strain that uses a WT orthogonal DNAP (wt TP-DNAP1) to replicate p1. The resulting population was then transferred into a selection OrthoRep strain using **MATE**. This strain encoding an error-prone DNAP (Trixy)[39] for p1 replication had its *ARO7* gene deleted, rendering it a phenylalanine auxotroph[60] that could be used to guide the evolution of the hypermutating p1-encoded genes to complement the auxotrophy. Since the *ARO7* gene is in the yeast ORF library installed onto p1, we expected that **EVO** in the absence of phenylalanine would enrich p1-*ARO7* library members. We found this to quickly occur simply after plating cells in media lacking phenylalanine where almost all surviving colonies contained p1-*ARO7*. This proved the integrity of the **PCR$_{lib}$**, **INT$_{p1}$**, and **MATE** subroutines and demonstrated that a known gene function could be selected by **EVO**.

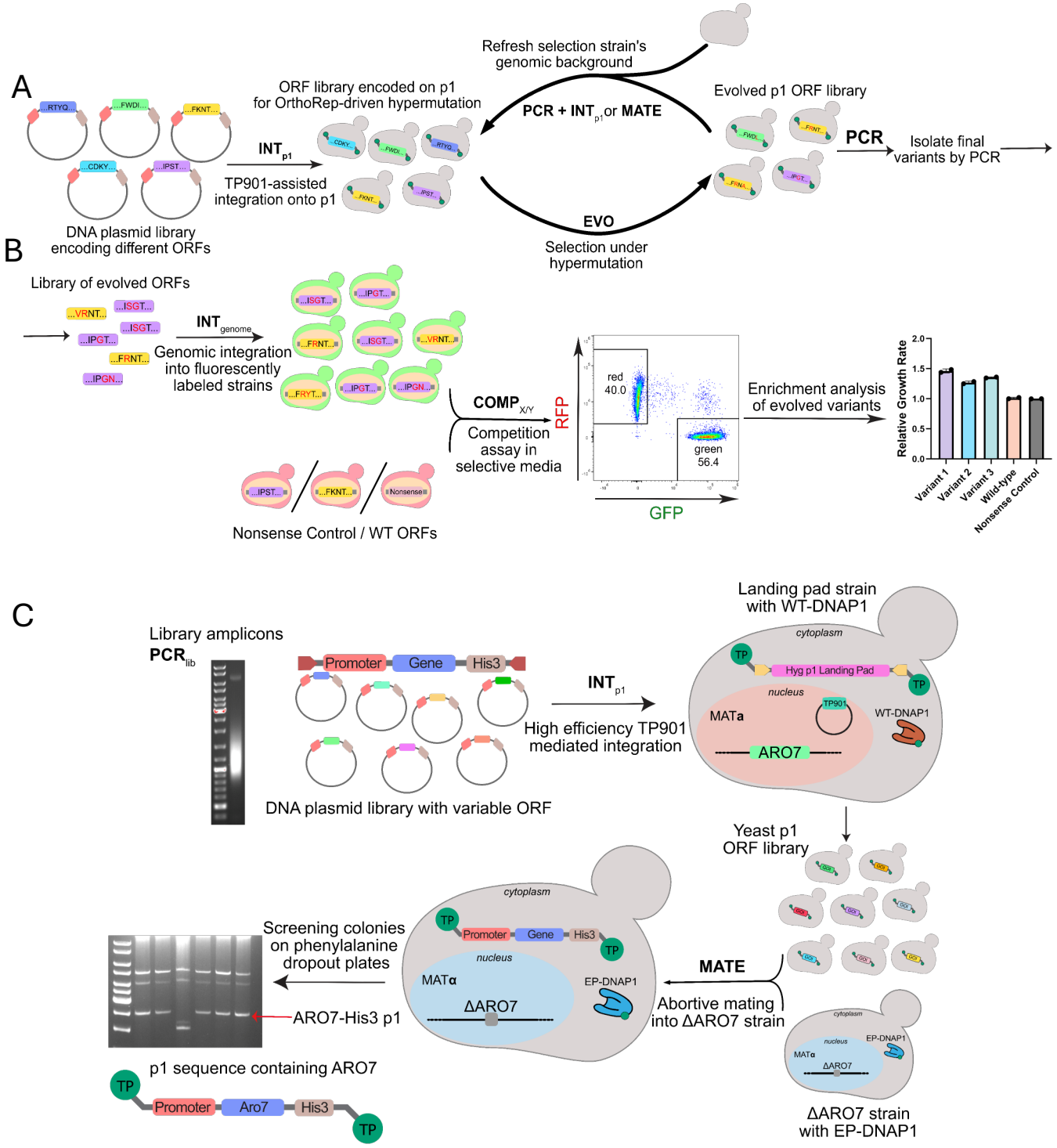


Figure 1. ORACLE expands the capabilities of OrthoRep to library evolution. and demonstration of rapid enrichment of a positive control gene. A) Schematic of ORACLE evolution. In ORACLE, gene libraries are first transformed onto p1 with **INT**$_{p1}$ and subjected to an arbitrary evolutionary pressure under hypermutation. Gene libraries are evolved with frequent refreshing of genome, utilizing either **MATE** or **PCR+INT**$_{p1}$. After several rounds of evolution and refreshing the genome, the genes are taken to the **COMP** stage for assessment. **B)** In **COMP**, the genes evolved by ORACLE are assessed for their fitness effect in a pooled outgrowth experiment where the ratio of cells containing test genes and control constructs are monitoring by flow cytometry. Genes are first assessed relative to a nonsense integration control (NS). If a gene increases fitness relative to NS, it is assessed for its fitness impact relative to the wild type gene (WT). The change in ratio and change in population size is used to calculate the fitness impact of the gene. **C)** Control experiment for the isolation of a positive control gene in ORACLE.

A plasmid library of yeast ORFs[42] was subjected to **PCR**$_{lib}$ followed by **INT**$_{p1}$ into a strain that expresses a WT OrthoRep DNAP for p1 replication. Using **MATE**, the p1-encoded library was then transferred into an *Δaro7* selection strain that encoded an error-prone OrthoRep DNAP, resulting in a yeast ORF library where the ORFs continuously hypermutate to support **EVO**. To test whether *ARO7* from the p1-encoded hypermutating yeast ORF library can be isolated by selection, we plated the **MATE** produced library on media lacking phenylalanine (-F) and screened colonies by p1 miniprep and sequencing. Over the course of two experiments, 9/10 colonies screened for their p1 encoded gene contained the expected *ARO7*, indicating that the **MATE** regime produced >99.99% correct mating events, as any diploid or incorrect haploid (**Fig. S3**) would have contained the genomic copy of *ARO7* from the donor cell. Shown is an agarose gel analysis of six minipreps[38] of DNA from yeast able to grow on -F. 5/6 contain the expected p1-*ARO7* plasmid.

Overview of ORACLE experiments

In this work, we carried out dozens of ORACLE experiments, where an experiment is defined as a particular gene library evolved towards an arbitrary function. Each was carried out in at least two replicates, one with the error-prone DNAP Trixy (error rate of 4 x 10^{-5} substitutions per base (s.p.b.)) driving hypermutation of the p1-encoded gene library and the other with the error-prone DNAP BadBoy3 (error rate of 10^{-4} s.p.b.)[39]. The attempted evolution experiments focused on three gene libraries: 1) the yeast ORF library described above[42], 2) a comprehensive *E. coli* ORF library where each ORF's C-terminus was fused to GFP[43,44], and the same *E. coli* ORF library without the GFP fusion, which we generated by subcloning (see **Methods**). **Table 1** summarizes the 13 ORACLE experiments where improved or novel gene function was clearly observed and characterized, including descriptions of the starting library used, the selection pressure imposed for **EVO**, and the evolutionary outcomes discovered. Nucleotide sequences of both mutants and WT genes can be found in **Table S1**. Additional ORACLE experiments with partially successful or uncharacterized evolutionary outcomes are also described throughout.

Experiment #	Gene library	Selection	Gene identified	Nonsilent Mutations	Mechanistic hypothesis
1	Yeast	Pantothenate auxotrophy	FMS1 Variant 1	V10A, K428R, I481V, A488V	Enhanced panothenate synthesis
2	e.coli-GFP	Pantothenate auxotrophy	UxuR-GFP	A32T, R67H, R150C, G215S, *GFP* Y343H, H358Y, K415E, I429T	Transcription of low expression yeast panthothenate synthesis genes
3	Yeast	Manganese tolerance	CCC1 Variant 1	S68P, F92L, T274A, A313T	Increased activity, specifity for mangenese, and overexpression better tolerated
3	Yeast	Manganese tolerance	CCC1 Variant 2	S68P, F256S, T274A, A313T	Increased activity, specifity for manganese, and overexpression better tolerated
4	Yeast	Lithium chloride tolerance	IPP1 Variant 1	K11D, I69V, F190L, K199R, K228E, T251A	Increased relief of unknown stress caused by high LiCl concentration
4	Yeast	Lithium chloride tolerance	IPP1 Variant 2	K11D, I69V, P167T, A232V, T251A	Increased relief of unknown stress caused by high LiCl concentration
5	Yeast	Copper sulfate tolerance	RTS3 Variant 1	C307frameshift	Copper binding and sequestration in nontoxic form
5	Yeast	Copper sulfate tolerance	RTS3 Variant 2	V8A, A19T, V39I, K74R, C307frameshift	Copper binding and sequestration in nontoxic form
6	e.coli GFP	Driving repressed promoter (pPHO5-Ade2)	YdfC-GFP Variant	D12V, A54V, *GFP* R91H, R165H, H173Y, V204I, R260H, I263T, A298V, V311I, L313H	Transcription factor / chromatin regulating activity
7	e.coli	Driving repressed promoter (pPHO5-Ade2)	YeiG Variant 1	E27G, G89R, A189V, P268S	Weak transcription factor activity
8	Yeast	Inhibiting degradation of medium strength N' degradation signal (type II UBR1 substrate) (pPOP6-UbiY-Ade2)	PET18 Variant 1	Y41L, F49S, S196N	Stabilization of N' terminal UbiX by direct binding to UbiX
9	Yeast	Inhibiting degradation of medium strength N' degradation signal (type II UBR1 substrate) (pPOP6-UbiY-Ade2)	PET18 Variant 2	K74E, A142T, E143D, R148K	Stabilization of N' terminal UbiX by direct binding to UbiX
10	Yeast	Inhibiting degradation of high strength N' degradation signal (type I UBR1 substrate) (pTDH3-UbiR-Ade2)	ROQ1 Variant 1	L2S Y88H	Increased stability in binding activity to UBR1
10	Yeast	Inhibiting degradation of high strength N' degradation signal (type I UBR1 substrate) (pTDH3-UbiR-Ade2)	RPS31 Variant 1	R74G	Unknown, inhibition of N' end rule substrate degradation
11	Yeast	Inhibiting degradation of high strength N' degradation signal (type I UBR1 substrate) (pTDH3-UbiR-Ade2)	ROQ1 Variant 2	I65T	Increased stability in binding activity to UBR1
12	e.coli	Inhibiting degradation of high strength N' degradation signal (type I UBR1 substrate) (pTDH3-UbiR-Ade2)	YjiE Variant 1	F143S	Global degradation inhibition
12	e.coli	Inhibiting degradation of high strength N' degradation signal (type I UBR1 substrate) (pTDH3-UbiR-Ade2)	AaT Variant 1	L19P, I62V	Transfer of longer lived (-LF) N' amino acids to arginine and lysine, and evolution of activity in yeast
13	e.coli	Inhibiting degradation of high strength N' degradation signal (type I UBR1 substrate) (pTDH3-UbiR-Ade2)	YdaG Variant 1	A17M, V18M, C30Y, V38M	Specific inhibition of UBR1
13	e.coli	Inhibiting degradation of high strength N' degradation signal (type I UBR1 substrate) (pTDH3-UbiR-Ade2)	SecB Variant 1	A14T, A41T, Q49L, D59N, V80I, A145T	Chaperone activity for UbiR-ADE2
14	Yeast	Preventing BAX cytotoxicity	NNT1 Variant 1	D19N, K49R, G60S, G90D, N135D, A219T	Unknown, inhibition of BAX mediated toxicity

Table 1. Summary of 14 key ORACLE experiments carried out along with 20 evolved gene variants representing evolutionary outcomes demonstrating clear enhanced or new gene functions.

ORACLE for tolerance part 1

In a first set of ORACLE experiments, we aimed to evolve tolerance to conditions known to reduce the growth rate of yeast, including high concentrations of manganese, NaCl, ethanol, clotrimazole, and growth media lacking B vitamins (pantothenate or biotin)[61,62]. For this first set of challenges, we carried out **EVO** from the yeast ORF library[42], as we predicted it was most likely to yield improved genes in yeast.

We used **PCR**$_{lib}$, **INT**$_{p1}$, and **MATE** to create two populations encoding the yeast ORF library on p1. In one population, p1 was replicated by the error-prone DNAP, Trixy (error rate of 4 x 10^{-5} substitutions per base (s.p.b)), and in the other population, p1 was replicated by the error-prone DNAP, BadBoy3 (error rate of 10^{-4} s.p.b.)[39]. Given six selection conditions, two starting populations for **EVO** to each condition, and our choice to carry out experiments in duplicate, 24 concurrent **EVO** experiments were performed.

Each **EVO** experiment underwent four passages in selective media with periodic adjustment of the population size during serial transfer to allow all 24 evolution experiments to be passaged on the same schedule. After the four passages, we used **MATE** to transfer evolving p1 populations into fresh strains that use a WT TP-DNAP1 to replicate p1. After three additional passages under selection conditions, we used **MATE** to transfer p1 populations into fresh strains that encode error-prone orthogonal DNAPs to replicate p1 for further **EVO**. This complete set of passages and **MATE**s were repeated once and evolution was stopped (**Fig. 1A-B**).

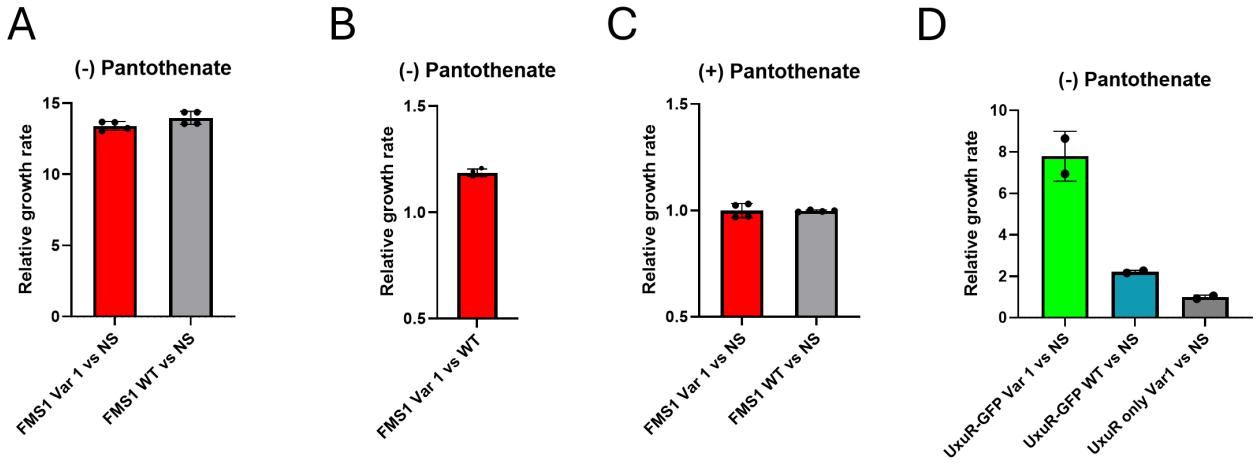


Figure 2. Characterization of evolution for pantothenate prototrophy from two gene libraries. A-D) Results of pantothenate prototrophy evolution. **A)** Relative growth rate of isolated high activity *FMS1 Variant 1* (V10A, K428R, I481V, A488V) in competition against NS control in media lacking pantothenate alongside WT FMS1. Genes were expressed under the control of an pRPL18B promoter (see **Box 1**, **INT**$_{genome}$ subroutine) and at this expression strength, the activity of Variant 1 and FMS1 were indistinguishable. **B)** Relative growth rate from competition of FMS1 *Variant 1* in direct competition against WT FMS1 in media lacking pantothenate under control of a weaker pREV1 promoter[46] (see **Box 1**, **INT**$_{genome}$ subroutine) demonstrating increased activity if *Variant 1* over WT, measured in quadruplicate. **C)** Relative growth rate from competition of *Variant 1* and WT FMS1 against NS control under control of a weaker pREV1 promoter in media containing pantothenate, demonstrating lack of fitness difference in the presence of pantothenate versus in the absence (**B**). **D)** Relative growth rate from competition of evolved bacterial transcription factor UxuR-GFP *Variant 1* (A32T, R67H, R150C, G215S, Y343H, H358Y, K415E, I429T) against NS control, alongside WT UxuR-GFP against NS control, and UxuR *Variant 1* without GFP against NS control in media lacking pantothenate. This shows evolutionary improvement of UxuR-GFP *Variant 1* over WT and the lack of activity improvement without the GFP tag.

OD-based monitoring of growth rates during **EVO** revealed that NaCl and biotin dropout tolerance did not improve and that clotrimazole resistance occurred too quickly to be explainable by an initially low frequency library gene enriching and evolving – instead we suspected clotrimazole resistance was evolved by easily accessible genomic adaptations such as efflux pump upregulation[63]. Thus, we considered **EVO** for tolerance to NaCl, biotin dropout, and clotrimazole unsuccessful from the yeast ORF library used. In contrast, tolerance of manganese, pantothenate dropout, and ethanol improved throughout **EVO**, suggesting successful adaptation.

COMP$_{pop/NS}$ was carried out for the presumptively successful **EVO** experiments. For the ethanol tolerance evolution experiments, **COMP**$_{pop/NS}$ did not confirm that the evolved p1 gene population conferred a growth benefit over the nonsense control, though we note that sequencing of all 4 ethanol tolerance **EVO** experiments showed near-complete enrichment of *DDR2* gene lineages for 3 of the 4. *DDR2* may warrant further investigation. For the pantothenate dropout tolerance and manganese tolerance **EVO** experiments, **COMP**$_{pop/NS}$ confirmed that the evolved p1 gene populations conferred a clear growth advantage under selection conditions.

Pantothenate dropout tolerance was achieved by improvements to an alternative pantothenate biosynthesis enzymes

For the pantothenate dropout tolerance **EVO** experiments, sequencing revealed the fixation of lineages descended from two genes, *FMS1* and *SPE4*, which is consistent with their prior implication in an alternative pantothenate synthesis pathway[64]. **COMP**$_{pop/WT}$ revealed that the population fitness of evolved *FMS1* variants was lower than that of WT *FMS1* (**Fig 2B**, **Fig. S7A**), but when repeated in a format where *FMS1* variants were expressed using a weaker promoter (pREV1 instead of pRPL18B)[46], **COMP**$_{pop/WT}$ enriched an *FMS1* mutant containing four mutations (V10A, K428R, I481V, A488V) that rendered it superior to WT *FMS1* in conferring tolerance to pantothenate dropout (**Fig. 2A Fig. S7B**), as assayed by **COMP**$_{mut/WT}$. In fact, the *FMS1* V10A, K428R, I481V, A488V (*FMS1 Variant 1*) mutant was able to support a level of growth in pantothenate dropout media comparable to that in complete media. Notably, there was no fitness cost for this mutant in pantothenate-rich media, indicating minimal metabolic burden. (**Fig 2C, Fig. S7C)**

Gene function can arise from unexpected sources

Given this early success of **EVO** for pantothenate prototrophy, we wished to test whether other libraries could originate and evolve similar tolerance outcomes. We describe a standout case here. Using the same processes of **PCR**$_{lib}$, **INT**$_{p1}$, **MATE**, and **EVO** (with rounds of **MATE**) starting from an *E. coli* ORF library where each ORF was fused to GFP[43,44], we found that an evolved version of a transcriptional repressor-GFP fusion, UxuR-GFP, dramatically increased yeast fitness in pantothenate dropout compared to WT UxuR-GFP, which only modestly increased yeast fitness in pantothenate dropout (**Fig. 2D, Fig. S7D**), as determined by **COMP**$_{mut/NS}$ and **COMP**$_{WT/NS}$. The evolved UxuR-GFP *Variant 1*, containing A32T, R67H, R150C, G215S, Y343H, H358Y, K415E, I429T mutations, required the GFP portion of the fusion for its functionality, likely due to the stability that GFP may confer (**Fig. 2D, Fig. S7E**). Although we do not know how the evolved UxuR-GFP mechanistically achieves tolerance to pantothenate dropout, this experiment demonstrates that non-native gene collections can be the source of novel adaptations that then evolve through mutational pathways, highlighting the potential role of HGT in providing the evolutionary origins of unexpected traits.

Manganese tolerance was achieved by modulating vacuolar transport

For the manganese tolerance **EVO** experiments, sequencing revealed the fixation of lineages descended from the *CCC1* gene. *CCC1* encodes a vacuolar transporter whose overexpression is known to mediate tolerance to toxic manganese concentrations[65]. Therefore, it was not surprising that our **EVO** experiments enriched *CCC1* lineages. Nevertheless, the actual evolved *CCC1*'s mutants, specifically variants carrying mutations S68P, T274A, A313T, with an additional F92L (*Variant 1)* or F256S (*Variant 2*), were superior to WT *CCC1*. First, these variants enhanced manganese tolerance (**Fig. 3A, Fig. S8A,B**) more than WT *CCC1*, as determined by **COMP**$_{mut/WT}$. Second, in normal conditions lacking manganese, the evolved *CCC1* mutants carried no fitness defect, as determined by **COMP**$_{mut/NS}$, whereas WT *CCC1* did confer a

growth defect, as determined by **COMP**$_{WT/NS}$ (**Fig. 3B, Fig. S8C**). Therefore, *CCC1* evolved an important new overall function: greater activity and specificity for conferring manganese tolerance.

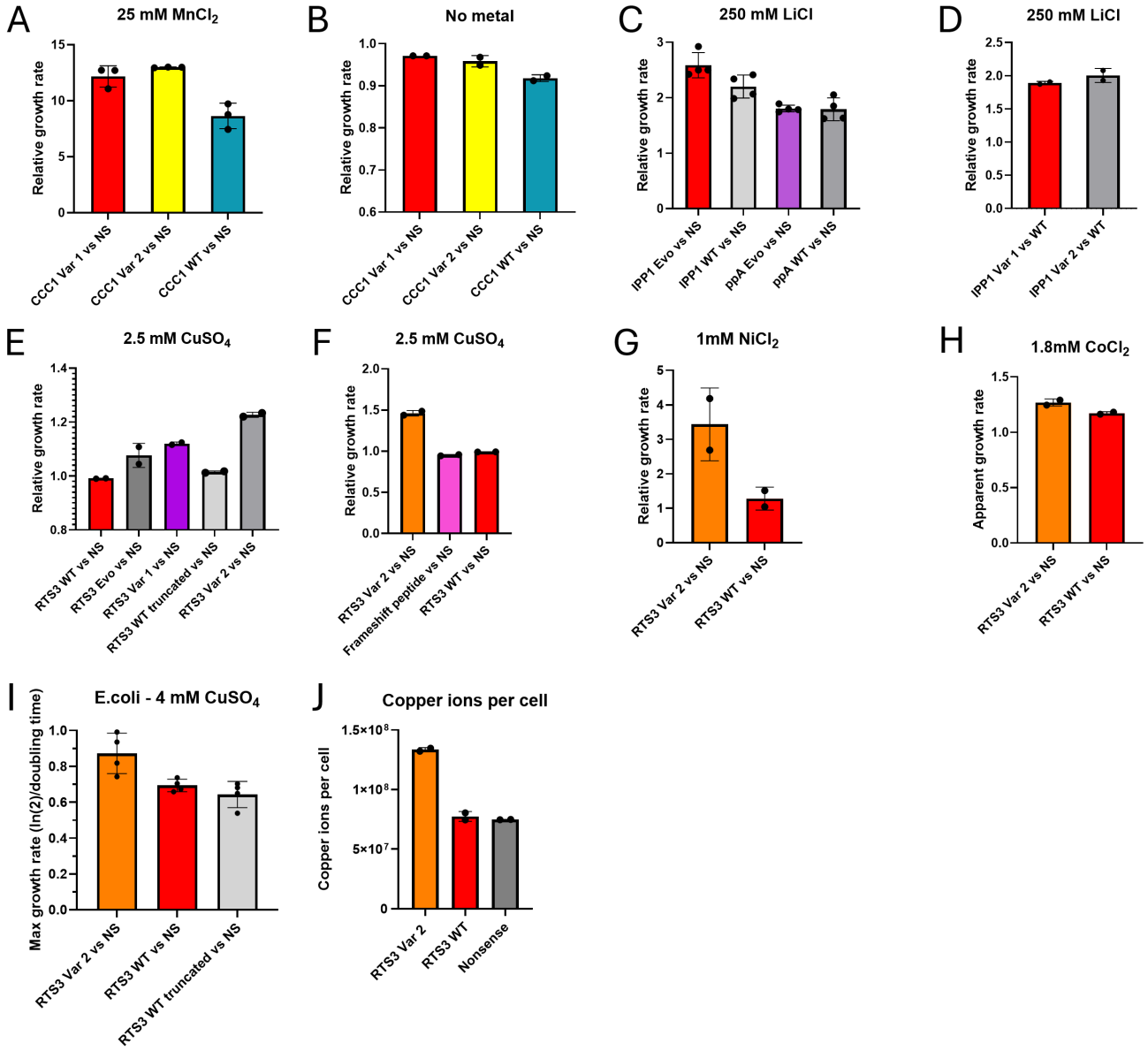


Figure 3. Assessment of evolved genes produced in metal tolerance experiments. A-B) *CCC1* mutants confer improved manganese tolerance. **A)** Relative growth rates from competition assay against NS control, demonstrating fitness improvement in media containing toxic manganese concentrations (25 mM, Manganese II chloride) of two variants (S68P, T274A, A313T, with additional F92L for *Variant 1* or F256S for *Variant 2*) over WT *CCC1*. **B)** Relative growth rates from competition assay against NS control, in synthetic complete media without metal demonstrating fitness cost of overexpression of *CCC1* (right bars), and reduced toxicity by evolved variants, note the Y-axis scale change as differences in growth rate were small absent metal. **C-D) Overexpression of pyrophosphatases increases resistance to toxic concentrations of LiCl**. **C)** Relative growth rates from competition assay against NS control by activity enriched yeast *IPP1* and *E. coli* ppA demonstrating improvement of *IPP1* relative to WT, but lack of improvement in *E. coli* ppA relative to WT, in synthetic media containing a toxic concentration of LiCl (250 mM). **D)** Comparison of relative growth rates vs NS control in toxic LiCl concentration with overexpression of two variants with shared mutations K11D, I69V, T251A and additional unshared mutations: F190L, K199R, K228E for *Variant 1* and P167T, A232V for *Variant 2*. **E-J) A frameshift mutation in RTS3 increases metal tolerance in yeast with *de novo* creation of a metal binding sequence. E)** Comparison of relative growth rates enabled by evolved *RTS3* variants and WT *RTS3* in 2.5 mM $CuSO_4$ against NS control. The evolved *RTS3* variants assessed here were cloned to omit the

AAs downstream of the premature stop codon. WT *RTS3* has no copper resistance phenotype nor does the WT version truncated to the same length as *Variant 1,* which only contains the frameshift mutation. Thus, the peptide sequence resulting from frameshift confers the phenotype rather than just the abbreviated length. The mutations in *Variant 2* (V8A, A19T, V39I, and K74R) further increase the copper resistance. **F)** Overexpression of the peptide resulting from the frameshift and prior to the stop codon (18 AA) alone confers no phenotype, demonstrating that fusion to the upstream *RTS3* sequence is required for copper resistance. **G)** The evolved *RTS3 Variant 2* also confers resistance to nickel, whereas WT *RTS3* has no activity. **H)** Evolved *RTS3* also confers resistance to cobalt, but in this case WT *RTS3* also has a small amount of activity. Apparent growth rates are in units per hr. **I)** The evolved *RST3 Variant 2* also increased copper tolerance in *E. coli*, while the WT *RTS3* did not. Max growth rate was extracted from a 24-hour growth curve at 37 °C, and the doubling time was calculated as described in methods for two bioreplicates and two technical replicates. Growth rates are in units per hr. **J)** *RTS3 Variant 2* overexpression increases the concentration of intracellular copper in high copper conditions, as measured by ICP-MS, indicating that the evolved variant likely causes sequestration of copper.

ORACLE for tolerance part 2

In the next set of ORACLE experiments, we once again aimed to evolve tolerance, including toxic concentrations of lithium and copper, but with the modification that instead of **MATE** to transfer p1 into fresh selection strain backgrounds from time to time during **EVO**, we used **PCR** + **INT**$_{p1}$. This effectively accomplishes the same operation as **MATE** and is a similarly streamlined process (see **Box 1**), with two advantages. First, we suspected that **MATE** strains could create minor selection pressures unrelated to the selection conditions for **EVO**, which we wished to avoid all else equal. Second, we note that p1 is a multicopy plasmid so each cell during **EVO** could contain a mixture of gene mutants. We suspected that **PCR** + **INT**$_{p1}$, which results in the installation of only one or a few copies of evolving genes into p1, would reduce the intracellular competition for fixation of gene variants encoded on different p1s competition and afford faster selective sweeps of superior genotypes[57]. We therefore used **PCR** + **INT**$_{p1}$ instead of **MATE** in all subsequent ORACLE experiments described.

For each subsequently described ORACLE result, the same **PCR**$_{lib}$, **INT**$_{p1}$, and **EVO** processes and passaging schedules were used as for the experiments described above (***ORACLE for tolerance part 1***), except with the replacement of **MATE** with **PCR** + **INT**$_{p1}$ and where noted. Also, as before, **EVO** experiments were done in two strains differing by the OrthoRep DNAP replicating p1: Trixy (error rate of 4 x 10^{-5} s.p.b.) and BadBoy3 (error rate of 10^{-4} s.p.b.)[39], but unlike before, each **EVO** was done as a single replicate.

Lithium tolerance emerged via an unknown mechanism involving pyrophosphatase activity

From the yeast ORF library and the *E. coli* ORF library, **EVO** under inhibitory lithium chloride conditions resulted in successful adaptation, as assayed by **COMP**$_{pop/NS}$, and the fixation of lineages derived from two inorganic pyrophosphatase genes, yeast *IPP1* and *E. coli ppa* (**Fig. 3C, Fig. S9A**). Despite their low overall sequence identity (22–27%), the active sites of these two genes are highly conserved (82–94% identity)[66], suggesting that their independent fixation from different **EVO** experiments and libraries was due to their primary enzymatic function as inorganic pyrophosphatases. The **COMP**$_{pop/WT}$ assays showed substantial adaptation beyond the WT *IPP1* ancestral gene for the yeast ORF library **EVO** outcomes. We therefore sampled two variants of *IPP1*s at the end of **EVO** for further testing. These two variants shared mutations K11D, I69V, T251A, but had additional unshared mutations (*Variant 1*: F190L, K199R, K228E, *Variant 2*: P167T, A232V). Both indeed conferred high lithium tolerance exceeding the activity of the WT *IPP1* sequence (**Fig. 3D, Fig. S9B**). Since the WT *IPP1* sequence still conferred some tolerance (**Fig. 3C, Fig. S9A**), the **EVO** experiment can be explained as a case where the WT *IPP1* gene first enriched from the library through its weak lithium tolerance activity and then evolved through a multi-mutational

pathway towards high lithium tolerance. To the best of our knowledge, the overexpression of pyrophosphatases has never been linked to lithium tolerance.

Copper tolerance was achieved through the discovery of a frameshift mutation that led to a de novo peptide appendage contributing to function

Two compelling outcomes emerged from copper tolerance **EVO** experiments starting from the *E. coli* ORF library and the yeast ORF library. The first outcome was the fixation of lineages descended from the *E. coli essQ* gene. EssQ is a predicted class II holin of the cryptic DLP12 prophage[67]. While overexpression of EssQ increased copper tolerance (**Fig. S10A**), the mutant variants produced through **EVO** did not improve upon the WT EssQ function.

The second outcome was the fixation of lineages descended from *S. cerevisiae RTS3*. *RTS3* encodes a largely unstructured protein previously linked to caffeine tolerance[68]. Strikingly, the evolved variant of *RTS3* that largely fixed after **EVO** contained a nucleotide deletion that resulted in a truncated protein with 120 amino acids (instead of *RTS3*'s original 263 amino acid size). Notably, the single nucleotide deletion caused a frameshift that introduced 18 amino acids prior to the stop codon, such that the last 18 amino acids of the 120 amino acid evolutionary outcome can be considered a *de novo* peptide. Furthermore, the frameshifted *RTS3* conferred copper resistance while the wild-type *RTS3* conferred no detectable resistance (**Fig. S10B-C**), suggesting that ORACLE drove the full origination and evolution of a new gene function in this case.

Several interesting observations about the frameshifted *RTS3* are of note. First, simply truncating WT *RTS3* to create a 120 amino acid fragment was not sufficient to observe any copper resistance, demonstrating that the 18 amino acid *de novo* peptide in the frameshifted *RTS3* was critical for function (**Fig. 3E**) (*Variant 1* contains only the frameshift mutation). Second, expression of only the 18 amino acid *de novo* peptide portion of the frameshifted *RTS3* demonstrated no increased copper resistance (**Fig. 3F, Fig S10B**), suggesting that the 18 amino acid *de novo* peptide appendage needs the remainder of *RTS3* for its discovered function, if only for stabilization as short peptides are typically quickly degraded. Third, the most active version we assayed among frameshifted *RTS3*s was found to contain additional mutations – V8A, A19T, V39I, and K74R (*Variant 2)* (**Fig. 3E, Fig. S10C**). This is consistent with an evolutionary pathway where WT *RTS3*, with no detectable initial activity for copper resistance, randomly sampled the single nucleotide frameshift mutation that introduced the 18 amino acid *de novo* peptide appendage and origination of weak copper tolerance, which was then improved by evolution through ORACLE's rapid hypermutation and continued selection.

To investigate mechanism, the frameshifted *RTS3* was tested for its ability to confer tolerance to other divalent ions, such as nickel (**Fig. 3G, Fig. S10D**) and cobalt (**Fig. 3H, Fig. S10E**), in **COMP**$_{mut/NS}$ assays, showing resistance to the additional metals as well. Additionally, the frameshifted *RTS3* was tested in *E. coli* where it was also found to increase tolerance to copper (**Fig. 3I**). The generality of contexts in which the frameshifted *RTS3* confers divalent metal ion tolerance suggests it may be acting to sequester the ions to reduce their toxicity.

To test this hypothesis, we measured intracellular copper concentrations using ICP-MS under high copper stress in yeast. Yeast strains expressing either WT *RTS3*, *RTS3 Variant 2*, or a NS control were passaged twice in media containing 2.5 mM $CuSO_4$. The intracellular copper content in cells was substantially elevated in the samples expressing *RTS3 Variant 2* relative to both WT *RTS3* and the NS control (**Fig. 3J)**. We therefore presume that the copper tolerance is mediated by copper sequestration inside the

cells, limiting the potential of copper to bind sensitive targets, supporting previous literature which found that *de novo* proteins have a tendency to bind metals.[69]

ORACLE for the modulation of expression and degradation mechanisms

A major mechanism by which evolution changes the phenotypes of organisms is by modifying the steady state level of proteins in the proteome. We wanted to test whether ORACLE could generate novel gene function that could increase the expression or decrease the degradation of a conditionally essential enzyme acting as the selection. The following three sections detail results of ORACLE experiments in yeast strains where the genomically encoded *ADE2*, responsible for a key step in adenine biosynthesis[70], was modified to decrease its expressed steady state protein level below the threshold needed for robust growth in the absence of adenine. As before, the same pipeline for ORACLE experiments was used, consisting of **PCR**$_{lib}$, **INT**$_{p1}$, and **EVO** with intermittent **PCR** + **INT**$_{p1}$ steps to refresh the host genome and prevent the accumulation potential genomic mutations confounding evolution of library genes. Passaging processes and schedules were as before such that each **EVO** was completed in <1 month involving ~14 total passages under selection. **EVO** experiments were done in two strains differing by the OrthoRep DNAP replicating p1: Trixy (error rate of 4 x 10^{-5} s.p.b.) and BadBoy3 (error rate of 10^{-4} s.p.b.)[39].

ORACLE generated presumptive new transcription factors from a cryptic prophage gene and from the moonlighting activity of a bacterial enzyme

We were interested in testing whether ORACLE could discover and evolve genes with novel transcription factor activity. In selection strains, we replaced the native promoter for *ADE2* with the phosphate responsive *S. cerevisiae* PHO5 promoter (pPHO5). In normal growth conditions, the pPHO5 promoter is covered by nucleosomes positioned to block access to the TATA box and other cis-regulatory elements, resulting in low transcription[71,72]. Upon phosphate starvation the PHO4 transcription factor can enter the nucleus, evict the nucleosomes and recruit transcriptional machinery[72]. Therefore, the selection strains did not grow without supplemental adenine, allowing us to carry out an ORACLE experiment to discover genes that either mimicked the PHO4 transcription factor or induce a cell state mimicking phosphate starvation whereby the natural PHO4 would enter the nucleus at relevant concentrations.

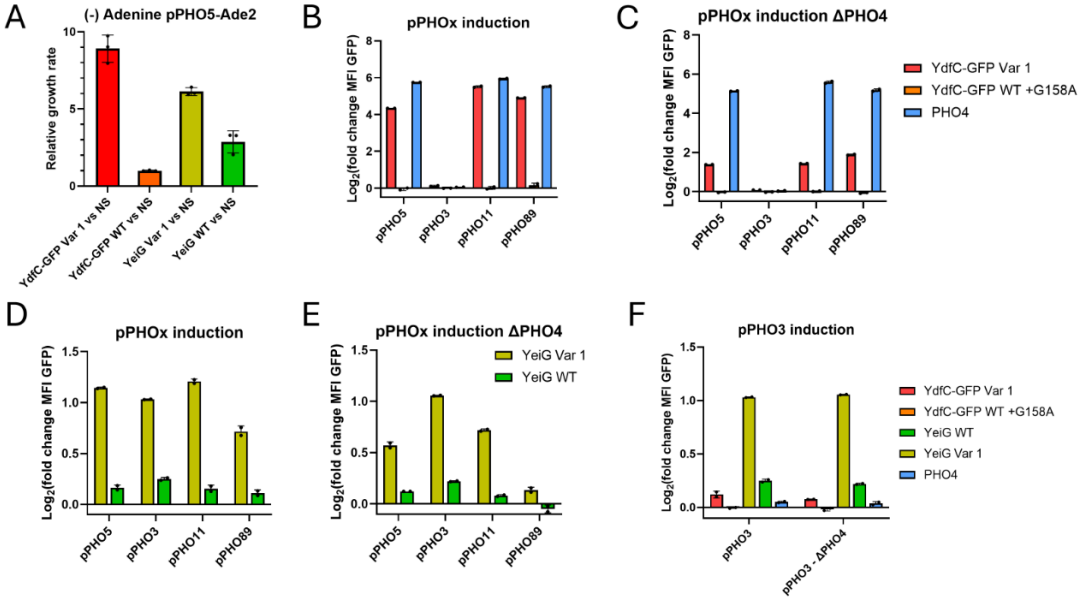


Figure 4. Cryptic prophage peptide YdfC-GFP evolved from a nonactive sequence to increase transcription of phosphate responsive genes and *E. coli* serine hydrolase YeiG evolved novel transcriptional activity. A) Relative growth rate in media

lacking adenine of evolved *YdfC-GFP Variant 1* (D12V, A54V, *GFP* R91H, R165H, H173Y, V204I, R260H, I263T, A298V, V311I, L313H), unevolved *YdfC-GFP*, evolved *YeiG Variant 1* (E27G, G89R, A189V, P268S), and unevolved *YeiG* in the pPHO5-Ade2 selection relative to NS control, demonstrating high activity of *YdfC-GFP Variant 1*, and undetectable activity of unevolved sequence. Also demonstrated is the low basal activity of WT *YeiG*, which was improved by evolution. **B-F)** Induction of phosphate responsive promoters. Data reported is the fold change in geomean GFP fluorescence compared to overexpression of nonsense control carried out alongside, measured in biological duplicate. **B)** *YdfC-GFP Variant 1* overexpression induces high levels of transcription from phosphate responsive promoters comparable to the activity of the natural transcription factor, *PHO4*, whereas overexpression of WT *YdfC-GFP* (with G158A mutation to negate background fluorescence) does not induce expression. **C)** *YdfC-GFP Variant 1* overexpression induces transcription of phosphate responsive promoters even in the context of cells lacking a genomic copy of *PHO4* (*Δpho4*), but at a reduced level, demonstrating that the induction is partially but not entirely dependent on the natural *PHO4* transcription factor. **D)** *YeiG Variant 1* evolved to increase induction of phosphate responsive promoters relative to WT *YeiG*. **E)** *YeiG Variant 1* overexpression still increases transcription from the phosphate responsive promoters in the context of *Δpho4*. **F)** Comparison of induction for natively downregulated PHO3 gene with or without the genomic *PHO4*, demonstrating distinct induction profile caused by overexpression of *YeiG Variant 1*.

From the *E. coli* ORF library fused to GFP, we found that **EVO** resulted in the fixation of lineages derived from a cryptic prophage peptide from the Qin prophage called YdfC[73]. Specifically, we isolated a *YdfC-GFP* variant containing mutations (D12V, A54V)-GFP(R91H, R165H, H173Y, V204I, R260H, I263T, A298V, V311I, L313H), *YdfC-GFP Variant 1*, which substantially outperforms the WT *YdfC-GFP* fusion (**Fig. 4A**, **Fig. S11A-B**). Notably, WT *YdfC-GFP* had no activity as determined by $\mathbf{COMP}_{mut/WT}$, making this an example where a gene with no activity presumably discovered an innovation through a new mutation and optimized that innovation further *via* a mutational pathway.

To dissect mechanism, we carried out a set of experiments aimed at characterizing the evolved *YdfC-GFP* gene's transcription factor activity. We built a series of yeast reporter strains where various phosphate response promoters, pPHO5, pPHO3, pPHO11, and pPHO89[74], were cloned upstream of a GFP reporter gene and integrated into the genome. We verified that GFP expression followed expectation in phosphate starvation conditions – induction for pPHO5, pPHO11, and pPHO89; repression for pPHO3 (**Fig. S11C**)[74]. We then tested the evolved *YdfC-GFP Variant 1* activity in these reporter strains in normal conditions without phosphate starvation. We observed that the evolved *YdfC-GFP* induced the same expression profile as phosphate starvation, copying the effect of overexpressing the endogenous PHO4 transcription factor that mediates the phosphate starvation program (**Fig. 4B**). In contrast, WT *YdfC-GFP* (containing a synthetically installed G158A mutation in GFP that makes it nonfluorescent without affecting folding so as not to confound the fluorescent readout from the reporter assay)[75] did not induce any expression of the phosphate starvation program (**Fig. 4B, Fig. S11D**). To determine whether the evolved *YdfC-GFP Variant 1* was indirectly inducing a phosphate starvation state or more directly interacting with pPHO5, we repeated our reporter assay with the endogenous copy of *PHO4* knocked out. Without PHO4, the evolved *YdfC-GFP Variant 1* still drove expression of the phosphate starvation program, albeit at a reduced level (**Fig. 4C, Fig. S11D**). This is consistent with the *YdfC-GFP Variant 1* acting as a transcription factor that has similar promoter binding and nucleosome eviction activity as PHO4. In short, ORACLE was able to originate and evolve a new gene function that replicates PHO4 transcription factor activity from an *E. coli* cryptic prophage gene-GFP fusion with no known function and no initial activity.

In the same set of **ORACLE** experiments, we discovered another interesting outcome from *YeiG*. The WT *YeiG* showed modest activity in inducing expression from pPHO5 and evolved mutants collected at the end of **EVO** showed high activity (**Fig. 4A, Fig. S11A-B**). In *E. coli*, *YeiG* is a serine hydrolase with primary activity on the substrate, S-formylgutathione[76]. To the best of our knowledge, no transcription factor activity for *YeiG* has been described. In reporter assays characterizing the ability of WT *YeiG* and

evolved *YeiG Variant 1* mutant to drive phosphate starvation response promoters, we found that the WT *YeiG* could induce a slight increase in expression from a panel of phosphate starvation response promoters while the evolved sequence containing mutations (*Variant 1*, E27G, G89R, A189V, P268S) drove much higher expression (**Fig. 4D, Fig. S11E**). As with the evolved *YdfC-GFP*, the transcription factor activity of the evolved *YeiG Variant 1* was partially reduced, but not eliminated, in strains where the endogenous *PHO4* was knocked out (**Fig. 4E, Fig. S11E**). Interestingly, the pPHO3 promoter is slightly downregulated in phosphate starvation, but the evolved *YeiG Variant 1* increased expression from pPHO3, in contrast to overexpression of PHO4 (**Fig. 4F**). Therefore, unlike the evolved *YdfC-GFP* that mimics the transcription factor activity and preference of *PHO4*, the evolved *YeiG Variant 1* induces a modified phosphate starvation response program. From a gene evolution perspective, this example of *YeiG* evolution could be considered a case where a moonlighting function of an enzyme beings its evolution of new regulatory gene function.

In service of completeness, we note that in similar ORACLE experiments starting from yeast libraries, we found the fixation of lineages derived from the genes *SPL2*, *PHO4*, *YDL129W*, and *YGR035C*. In **COMP**$_{pop/NS}$ experiments, we confirmed that the evolved populations deriving from these genes increase fitness under selection for Ade2 expression from pPOP5. However, in **COMP**$_{pop/WT}$ experiments, the evolved populations deriving from these genes did not surpass the fitness conferred by their WT ancestors (**Fig. S11F**), so we did not follow up with further characterization experiments in these cases.

ORACLE evolves new function from a potential pseudogene

In another configuration of Ade2-based selection, we reduced the steady-state concentration of Ade2 protein by replacing its native promoter with the relatively weak promoter, pPOP6[46], and by attaching an N-terminal ubiquitin fused to the Ade2 in order to generate defined N-terminal amino acids upon ubiquitin hydrolysis[77]. The identity of the N-terminal amino acid can dictate the half-life of proteins – the so-called N-end rule – by recruiting ubiquitin ligases to mark the protein for degradation[77–79]. In our case, ubiquitin was fused to an N-terminal tyrosine (UbiY) followed by a short peptide fused to Ade2[46,77]. Upon translation of the UbiY-Ade2 fusion, the ubiquitin is hydrolyzed by a deubiquitinating enzyme (DUB) and the new N-terminal amino acid (*i.e.*, tyrosine) is revealed. The half-life of the resulting protein is then controlled by the ubiquitin ligase, Ubr1, which recognizes hydrophilic N-terminal residues in its type I site, or hydrophobic residues in its type II site and has additional activity for misfolded substrates[80]. In short, in this selection, which we call pPOP6-UbiY-Ade2, selection pressure for higher Ade2 stability is achieved by continuous passaging in media lacking adenine and could be satisfied by, for example, inhibiting the recognition of the N-terminal Y by Ubr1, perturbing any number of associated processes that mediate degradation through N-terminal Ys, etc.

ORACLE experiments starting from both the yeast ORF library and the *E. coli* ORF libraries yielded clear adaptation from only one gene, which proved to be quite interesting. Specifically, we observed fixation of lineages derived from a yeast gene of unknown function, *PET18*. As far as we know, this gene has previously only been studied for potential activity in a thiamine salvage pathway due to its homology to Thi20, but in that study, no activity could be identified[81]. **EVO** in a selection strain that replicates p1 with Trixy and in a selection strain that replicates p1 with BadBoy3 both resulted in fixation of *PET18* lineages, *Variant 1*: Y41L, F49S, S196N and *Variant 2*: K74E, A142T, E143D, R148K, respectively; while the WT *PET18* showed very little activity in the pPOP6-UbiY-Ade2 selection context in **COMP**$_{WT/NS}$ assays (**Fig 5A, Fig. S12A)**. An evolved *PET18* variant from each of the two successful campaigns was subjected to further characterization (**Fig 5A, Fig. S12A)**. Notably, no mutations were shared between the two variants.

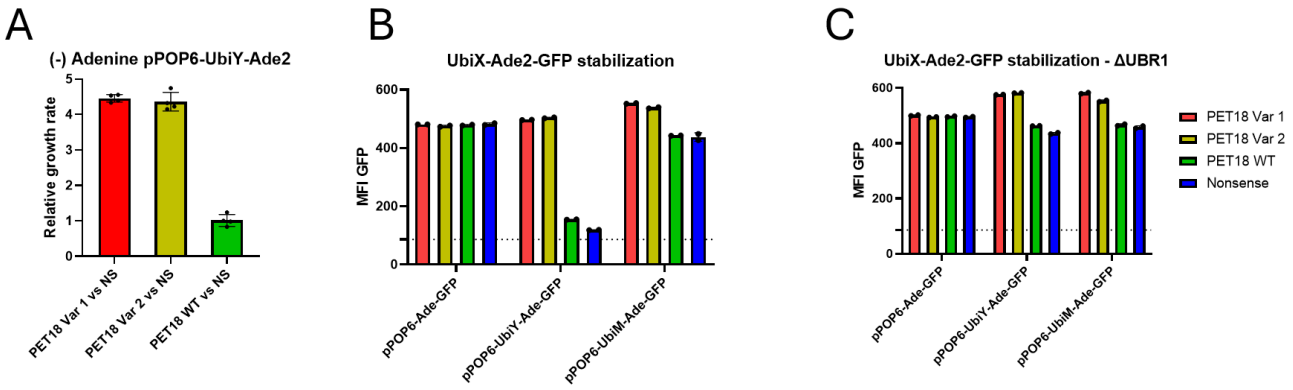


Figure 5. Evolved PET18 variants increase growth rate through Ubr1 independent mechanisms, but UbiX dependent mechanisms. A) Relative growth rate vs. NS from competition assays in media lacking adenine with cells with pPOP6-UbiY-Ade2 casette. *Variant 1* contains mutations Y41L, F49S, and S196N and *Variant 2* contains mutations K74E, A142T, E143D, and R148K. **B-C)** Effect of overexpression on protein level of pPOP6-Ade2-GFP, pPOP6-UbiY-Ade2-GFP, and pPOP6-UbiM-Ade2-GFP. **B)** WT *PET18* is barely distinguishable from nonsense control, and both evolved variants increase protein level, data reported is geomean GFP fluorescence measured in biological duplicate. **C)** Stabilization of UbiX tagged genes by evolved *PET18* variants is still detectable in cells lacking *UBR1 (Δubr1)*. Data reported is geomean GFP fluorescence measured in biological duplicate.

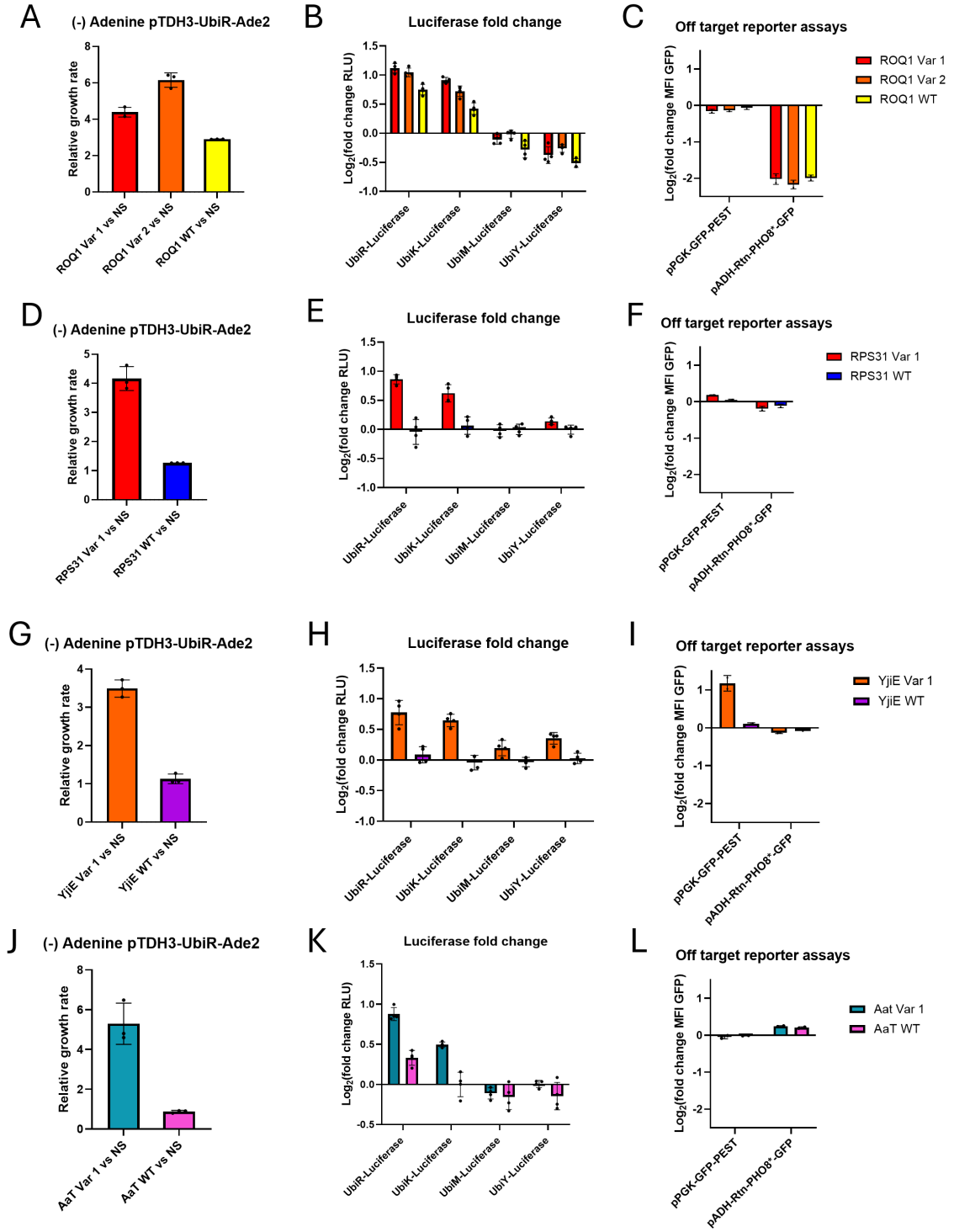
We developed reporter assays to measure the effects of overexpressing the evolved *PET18* variants on degradation of an Ade2-GFP fusion. These revealed that the evolved variants increased the level of protein only with UbiY attached, rather than by increasing transcription from pPOP6. The effects were not specific for UbiY as they reproduced with UbiF attached to Ade2-GFP (**Fig. S12B**), and slight increases in fluorescence could also be seen with UbiM N-terminally fused to Ade2-GFP (**Fig. 5B, Fig. S12A**). The evolved *PET18*s also increased fitness in the context of pRAD27 and pRNR2 driving the UbiY-Ade2 construct (**Fig. S12B**) and increased fluorescence of a pPOP6-UbiX-GFP reporter lacking Ade2 in the fusion (**Fig. S12C**).

To test whether the mechanism was specific to attenuating degradation after deubiquitylation by DUBs first revealed a specific N-terminal amino acid client for Ubr1, we knocked out the *UBR1* gene and measured the effect of evolved *PET18* variants on various Ade2-GFP reporters. Knockout of the *UBR1* gene resulted in a near fluorescence match between pPOP6-Ade2-GFP and pPOP6-UbiY-Ade2-GFP. Interestingly, overexpression of the evolved *PET18*s actually increased the fluorescence level of pPOP6-UbiY-Ade2-GFP to above the level of pPOP6-Ade2-GFP (**Fig. 7C**). The growth rate of samples in **Fig. 5A-C** was also measured in the absence of supplemental adenine in a separate growth curve assay (**Fig. S12D**) and confirmed that the evolved *PET18* variants increased growth rate under adenine selection in pPOP6-UbiX-Ade2-GFP containing cells even in the absence of genomic *UBR1*.

Our most parsimonious mechanistic hypothesis is that the evolved *PET18* evolved to be a "chaperone" that both prevents recognition of UbiY by DUBs and directly stabilizes the UbiY-Ade2-GFP fusion. This hypothesis is supported by a predicted AlphaFold 3 complex structure of evolved *PET18*s with UbiY-Ade2-GFP, showing an interaction between the evolved *PET18* and the UbiY portion of the fusion[82] (**Fig. S12E**). *PET18* shares structural homology to many aminopyrimidine aminohydrolases[81], and given its lack of known activity, may be a pseudogenized version of a thiaminase. The evolution of presumptive chaperone activity from a potential pseudogene with no known function is a strong example of how evolutionary innovation can spring from unexpected sources.

ORACLE yields several modulators of protein degradation pathways

In the final configuration of Ade2-based selection, we again selected for stabilizing a UbiX-Ade2 fusion, but replaced UbiY with UbiR, which results in the release of an N-terminal arginine-Ade2 protein that should be more quickly degraded as N-terminal arginines are stronger clients for *UBR1*[80]. To compensate, we drove UbiR-Ade2 with a stronger pTDH3 promoter[46]. We therefore refer to this selection as pTDH3-UbiR-Ade2.



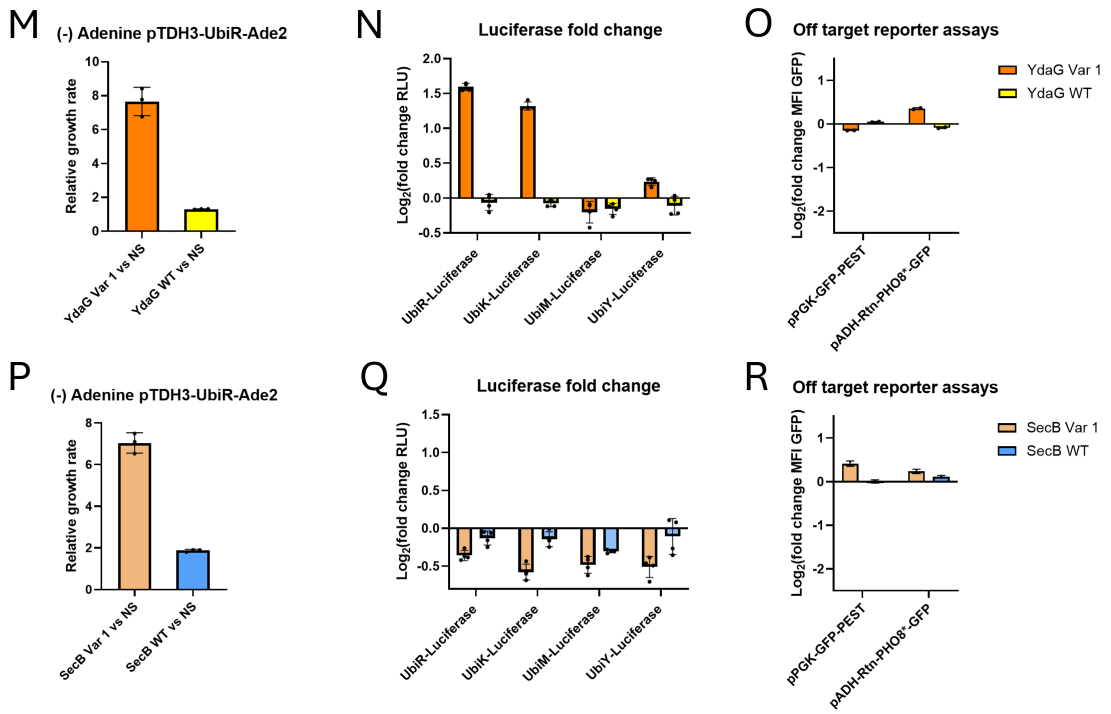


Figure 6. A-C) Natural Ubr1 ligand, *ROQ1*, evolved greater activity for Ubr1 regulation. A) Relative growth rate from competition assay of evolved *ROQ1* variants and nonsense control demonstrating slight increase in growth rate by evolved variants. **B)** Both evolved and WT *ROQ1* sequences increased level of UbiX-Nanoluc reporter. Luciferase signal was measured in biological duplicate, and technical duplicate, and the fold change relative to a nonsense control transformed alongside is reported. **C)** *ROQ1* variants redirect Ubr1 to increase degradation of cytoplasmic misfolded reporter. Data reported is the fold change in geomean GFP fluorescence compared to overexpression of nonsense control carried out alongside, measured in biological duplicate. **D-F) *RPS31* evolved novel UbiX stabilization activity with a single mutation. D)** *RPS31* Variant 1 containing one mutation (R74G) increased growth rate under selection, whereas WT *RPS31* had undetectable activity in competition assay. **E)** Novel stabilization activity by *RPS31 Variant 1* is supported further by luciferase assays where WT *RPS31* had undetectable activity. Luciferase signal was measured in biological duplicate, and technical duplicate, and the fold change relative to a nonsense control transformed alongside is reporter. **F)** Evolved *RPS31* Variant 1 did not substantially impact degradation mediated by a PEST degradation tag or degradation of a misfolding reporter. Data reported is the fold change in geomean GFP fluorescence compared to overexpression of nonsense control carried out alongside, measured in biological duplicate. **G-I) *E. coli* gene *YjiE* evolved novel activity for regulating a common degradation component through a single mutation. G)** *YjiE* containing one mutation, F143S (*Variant 1*), demonstrated increase in growth rate by evolved variant in competition against NS control, while WT *YjiE* had undetectable activity. **H)** *YjiE Variant 1* increased the level of UbiX-Nanoluc reporters while WT *YjiE* had undetectable activity. Luciferase signal was measured in biological duplicate, and technical duplicate, and the fold change relative to a nonsense control transformed alongside is reported. **I)** Evolved *YjiE Variant 1* decreased degradation of GFP-PEST indicating a decrease in global degradation. Data reported is the fold change in geomean GFP fluorescence compared to overexpression of nonsense control carried out alongside, measured in biological duplicate. **J-L) *E. coli* gene *AaT* evolved improved aminotransferase activity. J)** Evolved *AaT Variant 1* containing two mutations (L19P, I62V) increases growth rate more than WT in competition assay. **K)** Increased activity by *AaT Variant 1* is supported further by luciferase assays, though WT *AaT* had measurable activity. Luciferase signal was measured in biological duplicate, and technical duplicate, and the fold change relative to a nonsense control transformed alongside is reported. **L)** *AaT Variant 1* did not substantially impact degradation of PEST degradation tag, or degradation of misfolded reporter. Data reported is the fold change in geomean GFP fluorescence compared to overexpression of nonsense control carried out alongside, measured in biological duplicate. **M-O) Cryptic prophage peptide YdaG evolved novel UBR1 inhibition. M)** *YdaG* peptide *Variant 1* containing four mutations A17M, V18M, C30Y, V38M, conferred an increase in growth rate but WT *YdaG* activity was indistinguishable from NS control. **N)** Stability of UbiX-Nanoluc reporter was dramatically increased by *YdaG Variant 1*, while WT *YdaG* activity was undetectable, supporting fitness competition data. Luciferase signal was

measured in biological duplicate, and technical duplicate, and the fold change relative to a nonsense control transformed alongside is reported. **O)** Evolved *YdaG Variant 1* only slightly decreased the degradation of the misfolding reporter and didn't impact GFP-PEST, supporting a model of specific UBR1 inhibition rather than inhibition of global degradation. Data reported is the fold change in geomean GFP fluorescence compared to overexpression of nonsense control carried out alongside, measured in biological duplicate. **P-R) *E. coli* SecB increased fitness through undetermined means. P)** Evolved SecB *Variant 1* containing six mutations (A14T, A41T, Q49L, D59N, V80I, A145T) increased growth rate in competition assays, more than WT SecB. **Q)** *SecB Variant 1* and *WT SecB* both slightly decreased the level of the NanoLuc reporter. Luciferase signal was measured in biological duplicate, and technical duplicate, and the fold change relative to a nonsense control transformed alongside is reported. **R)** *SecB Variant 1* did not substantially impact degradation mediated by the PEST degradation tag nor degradation of misfolding reporter. Data reported is the fold change in geomean GFP fluorescence compared to overexpression of nonsense control carried out alongside, measured in biological duplicate.

EVO from yeast ORF libraries under selection for adenine biosynthesis yielded successful lineages descended from a recently discovered *UBR1* modulator, *ROQ1*[80,83]. **COMP**$_{mut/WT}$ assays on evolved *ROQ1* variants showed slightly improved activity over WT *ROQ1* (**Fig. 6A, Fig. S13A-B**) through mutations that have not been previously reported. *ROQ1* is known to alter the substrate specificity of *Ubr1* away from hydrophilic N-terminal amino acids towards misfolded clients by acting as a pseudosubstrate and allosterically regulating Ubr1 with multiple cooperating motifs[80,83]. The two characterized evolved *ROQ1* variants contained nonsynonymous mutations L2S, Y88H (*Variant 1*) and I65T (*Variant 2*). Given that the I65K and I65A mutations were previously identified to make *ROQ1* less active[83], it was surprising that an additional activity-increasing mutation could be identified at I65. The evolved *ROQ1* variants were shown in reporter assays (where a luciferase was fused to various UbiXs) to more strongly inhibit degradation of UbiR-tagged and UbiK-tagged NanoLuc compared to WT *ROQ1* (**Fig. 6B**), confirming the evolution of a more potent *ROQ1*. In a separate GFP assay utilizing a misfolded reporter known to be a substrate for Ubr1 after *ROQ1* mediated reprogramming[80], the evolved *ROQ1*s and WT *ROQ1* redirected Ubr1 comparably (**Fig. 6C**).

ORACLE additionally generated an evolved yeast *RPS31* containing only one mutation, R74G, that could increase the stability of Ade2 in the UbiR-Ade2 construct. Notably, overexpression of the WT *RPS31* gene produced no detectable fitness benefit in COMP$_{WT/NS}$ assays (**Fig. 6D**) and no increase in luciferase activity in UbiX-NanoLuc assays (**Fig. 6E**), while the evolved *RPS31* (R74G) showed substantial activity in both. Interestingly, when *RPS31* (R74G, *Variant 1*) was tested for its effect on global protein degradation using reporter assays where GFP was tagged with a PEST degron or a misfolded protein, no effect was observed (**Fig. 6F**). This suggests that *RPS31* (R74G) specifically stabilized proteins with various N-terminal amino acid variants but did not broadly affect protein degradation (**Fig 8D-F, Fig. S13A-B**). This is further supported by the stabilization activity of *RPS31* (R74G) on a UbiY-mRuby reporter (**Fig. S13C**). The WT *RPS31* is expressed as a ubiquitin S31 fusion, which is cleaved after residue G77 by DUBs[84]. It is possible that the R74G mutation prevents this cleavage and thus impacts the proper formation of the 40S subunit, though it is difficult to speculate how that might directly relate to the observation that *RPS31* (R74G) stabilizes various N-end rule substrates. From a gene evolution perspective, this is an example where a single mutation originates a novel detectable functional output.

ORACLE also produced an interesting set of outcomes from the pTDH3-UbiR-Ade2 selection when **EVO** was carried out from the *E. coli* ORF library. One outcome was *YjiE*, which has been described as a hypochlorite-specific transcription factor in *E. coli*[85]. Similar to *RPS31*, WT YjiE had undetectable activity, while a variant containing only one amino acid substitution potently inhibited degradation of N-end rule substrates (**Fig. 6G-H, Fig. S14A-C**). Notably, the evolved *YjiE* variant also inhibited degradation of a GFP-PEST reporter (**Fig. 6I**), which is ubiquitinated by the SCF complex[86]. Therefore, the evolved *YjiE*

Variant 1 likely inhibits a component of global protein degradation or alters expression of components common between the SCF complex and Ubr1-based degradation pathways.

Another outcome was *E. coli AaT*. *AaT* naturally attaches leucine and phenylalanine to N-terminal arginines and lysines[78,87]. **EVO** yielded an improved variant of *AaT*, with mutations L19P and I62V (*AaT Variant 1)*. The improvement over WT AaT as determined by **COMP**$_{mut/WT}$ was recapitulated in the luciferase reporter (UbiX-NanoLuc) assay (**Fig. 6J-K, Fig. S14A-B**). The mechanism of evolved *AaT* is obvious: leucine is made to be the new N-terminal amino acid after attachment to arginine, creating a longer half-life to the protein given the preferences of *UBR1*[78,80], and ORACLE experiments simply enriched and evolved AaT for higher activity in yeast (**Fig. 6J-K, Fig. S14A-B**).

Perhaps the most interesting example of all evolution experiments selecting for extended Ade2 half-life from the UbiR-Ade2 fusion was the fixation of lineages descended from a short *E. coli* cryptic prophage gene called *YdaG*. The evolved variant we isolated contained mutations A17T, V18M, C30Y, V38M (*Variant 1)* and substantially outperformed WT *YdaG*, which conferred no activity in selection for Ade2 stabilization (**Fig 6M, Fig. S15A-B**). Strikingly, the evolved *YdaG* variant inhibited the degradation of N-end rule substrates even more potently than the endogenous *ROQ1* gene in the UbiX-NanoLuc luciferase reporter assay (**Fig. 6M** versus **Fig. 6A**). Unlike *ROQ1*, the evolved YdaG variant increased the level of N-terminal tyrosine tagged mRuby2 as well and slightly increased the misfolded protein degradation reporter, PHO8*-GFP (**Fig. 6N, Fig. S15C**). The evolved *YdaG* variant had no detectable effect on degradation of the GFP-PEST reporter (**Fig. 6O**). Taken together, these data suggest that the evolved *YdaG* variant specifically inhibits *UBR1*.

The final example from selections to extend Ade2's half-life from the UbiR-Ade2 fusion was the emergence of a mutated chaperone gene from *E. coli*, *SecB*. The evolved *SecB* variant we isolated contained the mutations A14T, A41T, Q49L, D59N, V80I, and A145T (*Variant 1) (***Fig 6P**, **Fig S15A-B**). Reporter assays on the evolved *SecB* variant's effect on N-end rule substrates and other degradation tags did not reveal much other than a slight decrease in degradation of the GFP-PEST and PHO8*-GFP misfolding reporters. However, the effect of the evolved *SecB* variant on growth rate in **COMP**$_{mut/NS}$ assays was high (**Fig. S15A**). The WT *SecB* also had activity (**COMP**$_{WT/NS}$) but was much less potent compared to the evolved *SecB* variant (**COMP**$_{mut/WT}$). The evolved *SecB* variant also slightly increased mRuby stability from the UbiY-mRuby fusion (**Fig. S15D**). We therefore speculate that this *SecB* chaperone primarily evolved to use Ade2 protein as a specific client, but do not have direct evidence for this hypothesis.

Expanding ORACLE's scope to heterologous targets

We were interested in exploring whether we could evolve regulators of therapeutically relevant human signaling proteins, which should be possible with ORACLE because human regulatory proteins often maintain the same function when transferred into yeast[88] and yeast models have been used to study human disease[89,90]. As demonstration, we chose to evolve inhibition of the human protein *BAX*, a pore-forming pro-apoptotic protein of the *Bcl-2* protein family that plays a crucial part in the mitochondrial pathway of apoptosis. While reduction of *BAX* activity has been implicated in certain cancers and neurodegenerative diseases,[91] *BAX* depletion can also be protective in diseases of uncontrolled cell death[92–94]. It is known that in *S. cerevisiae*, *BAX* expression induces cell death by cytochrome c release, as it does in as mammalian cells[89,90], which should allow for the growth-based selection of *BAX* inhibition.

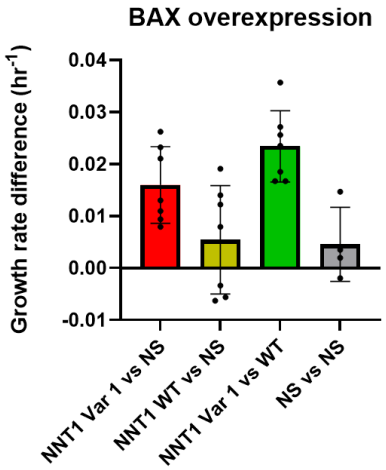


Figure 7. **Rescue of *BAX*-induced cell death**. Excess growth rate over nonsense control achieved by the evolved *NNT Variant* 1 (D19N, K49R, G60S, G90D, N135D, A219T) and WT *NNT1* under *BAX* expression. While *NNT1* WT did not improve growth over the NS control, *NNT1 Variant 1* bearing 6 nonsynonymous mutations improved growth rate significantly. *BAX-2A-URA3* was expressed under an inducible *GAL1 promoter* and growth in uracil-deficient medium selects for the expression of *BAX-2A-URA3*.

In an ORACLE strain, we genomically encoded *BAX* under the inducible *GAL1* promoter and also coupled the expression of *BAX* to a sortable marker, *GFP*, using a 2A peptide fusion. This allowed for the selection of genes that inhibit *BAX* activity under galactose induction while preventing simple genomic mutations that disable *BAX* expression from fixing, since those would reduce the expression of GFP which we selected for with intermittent rounds of fluorescence activated cell sorting (FACS). **EVO** under the overexpression of the *BAX-2A-GFP* fusion through 4 intermittent rounds of *BAX* induction, growth, and FACS enriched for an evolved *NNT1* variant (*NNT1 Variant 1*) with mutations D19N, K49R, G60S, G90D, N135D, and A219T from the yeast ORF library. *NNT1* is a protein methyltransferase known to methylate elongation factor 1α. The evolved *NNT1 Variant 1* promoted survival in the presence of *BAX* induction while the WT *NNT1* did not provide any protective effect and was indistinguishable from the nonsense control (**Fig. 7, Fig. S16**). While we did not investigate the mechanism for protection, this experiment suggests the potential of ORACLE in discovering and evolving therapeutically relevant phenotypes and biomolecules.

Conclusion

In this work, we have presented an abundance of examples where ORACLE quickly yielded new gene functions from often unexpected sources, including cases where the parental gene from which novel traits originated had no known function and no observed activity towards the function it eventually gained. These results support a growing appreciation that new gene function can spring from unexpected and possibly inert sources, providing motivation to probe the generally held assumption that genes primarily evolve new activities similar to their original activities. Taken together, our experiments show that gene evolution is highly versatile when the diversity of genes available as sources of new function is large and when the systems level complexity of cells is available to afford multiple routes to new selectable phenotypes. These conditions are of course met in nature. ORACLE, which begins continuous gene evolution from large gene libraries *in vivo*, also meets these conditions, while providing a time acceleration through continuous hypermutation such that we are able to observe evolutionary trajectories and outcomes emerge on laboratory timeframes at scale. We therefore suggest that ORACLE, especially with further technological expansion (see **Supplementary Text**), provides a unique and powerful

experimental approach to answering the fundamental biological question, how do new gene functions originate and evolve.

Methods

Emulsion PCR (PCR_{lib}). The commercially available Micellula emulsion PCR kit (E3600-01, Roboklon, Poland) was used for preparation of emulsion PCRs. We followed the recommended ratios and volumes for each reaction. For each reaction, 220 µL emulsion component 1, 20 µL emulsion component 2, 60 µL emulsion component 3 were mixed by pipetting. When preparing multiple reactions, the volumes were scaled up and distributed in 300 µL aliquots in 1.5 mL Eppendorf tubes. The oil emulsion mix was cooled on wet ice while the aqueous PCR component of the reaction is prepared. 55 µL of a PrimeSTAR GXL DNA polymerase (TaKaRa) reaction was prepared with the modifications of addition of a final concentration 0.025 mg/mL acetylated BSA. For each library, four emulsions were prepared with a serial 1:2 dilution of template DNA to ensure that the desired occupancy of micelles was hit, starting from 10 ng of plasmid. 50uL of the PCR mix was added to the cooled emulsion mix, and the remaining 5 µL of the PCR was saved as an open PCR control. The cooled emulsions were manually vortexed in a cold room at 4C for 5 minutes. The 350 µL reaction volume was distributed into an 8-tube PCR strip with 50 µL / well, and a well-formed emulsion makes an off-white “creamy” mix. 25x cycles of PCR were run with a TM of 60C, and 2-minute extension time, alongside the open controls. Post reaction the emulsion reactions were pooled, and emulsions broken and purified using the Supplied purification reagents and columns. 1/10 of the reaction was run on an agarose gel alongside the open control, and the emulsion reaction which yielded the lowest DNA output (lowest micelle occupancy) was selected for gibson assembly. Primer sequences and plasmids used can be found in **Table S2, S3** respectively.

Frozen competent yeast cell preparation. Desired strain is inoculated into selective media two days prior to frozen competent cell preparation. The day prior, cells are back diluted at 30 ºC with shaking at 200 rpm to be in late exponential phase the morning of. Note that for error-prone, P1-containing strains, several passages may be required for yeast to acclimate and grow well. The OD_{600} of the yeast culture was measured using a spectrophotmeter. Approximately 2.5 x 10^9 cells were inoculated into 500 mL of YPD, and the culture was grown in the shaking incubator at 30 ºC for 4-6 hours until reaching a cell density of 2 x 10^7 cells/mL. Yeast cells were harvested by centrifugation at 3,000 x g for 5 minutes and washed with 0.5 volumes of sterile water, followed by a second wash with 0.01 volumes of sterile water. The cell pellet was resuspended in 0.01 volumes of sterile-filtered frozen competent cell solution (5% glycerol, 10% DMSO, both from ThermoFisher), and 50 µL aliquots were dispensed into 1.5 mL microcentrifuge tubes. These tubes were placed in a styrofoam container and stored at -80 ºC.

Yeast transformation. Frozen competent cells were thawed in a 37 ºC water bath for 30 seconds and pelleted by centrifugation at 13,000 x g for 2 minutes. Frozen competent cell transformation mix (260 µL of 50% w/v PEG 3350 (ThermoFisher), 36 µL of 1 M LiAc (ThermoFisher), 50 µL of 2.0 mg/mL single-stranded carrier DNA (ThermoFisher), 14 µL of DNA plus sterile water) was added to the cell pellet, and the mixture was vortexed vigorously to resuspend the cells. The tube was then incubated in a 42 ºC water bath for 30 minutes. After incubation, the cells were pelleted by centrifugation at 13,000 x g for 30 seconds, and the supernatant was removed. Cells are either resuspended in selective media for outgrowth, or streaked on selective plates for clonal isolation. When genomic manipulation was carried

out to construct a new strain, the modified locus was confirmed by PCR. Yeast strains generated in this study and their genotypes can be found in **Table S4.**

Yeast culture and growth monitoring. Yeast strains were grown in standard media, including yeast extract peptone dextrose (YPD) (10 g/L bacto yeast extract; 20 g/L bacto peptone; 20 g/L dextrose) or the appropriate synthetic drop-out media (yeast nitrogen base w/o amino acids (US Biological), drop-out mix synthetic minus the appropriate nutrients (US Biological), and dextrose).

OD_{600} was regularly measured in a 96-well format. 100 µL of culture was measured per well using a Tecan Spark. To convert to a standard 1 cm path length, background was subtracted and the value multiplied by 10. This factor was determined empirically to match readings from a standalone spectrophotometer, corresponding roughly to 1.5E7 cells/mL at OD = 1.

cDNA Plasmid library extraction. Takara normalized Arabidopsis (Cat #630487) and Human (Cat #630481) cDNA libraries were obtained as 2 micron plasmids pre-transformed into yeast. Plasmids were extracted by first spinning down one full tube of yeast and resuspending in a 100 U/ml Zymolyse solution in 0.9M sorbitol and 0.1M EDTA. This was incubated at 37C for 1 hr with rotation. This was spun down at max speed for one minute, washed with 0.9% NaCl, and then resuspended in 600 µL 0.9% NaCl. This was then lysed with a standard plasmid miniprep kit (Zymo research #D4209), and DNA purified with the Supplied column. Mixed genomic and plasmid DNA was then transformed into commercial DH5b electromax cells (Invitrogen #18290015) according to manufacturer protocol. 1/1000 of the electroporation was plated on antibiotic media for CFU count, and the rest was grown in liquid selection. DNA was prepared to be used as a plasmid template by miniprep.

Plasmid library gibson assembly. DNA libraries amplified as described by $\mathbf{PCR}_{lib}$ were cloned into a backbone plasmid containing chloramphenicol resistance, I-Ceu-I restriction digest sites flanking the integration cassette, TP901 integration flanks, a P1 cytoplasmic promoter, and auxotrophic markers His3 or Leu2. The open vector for each library was prepared by PCR with primers which add homology to the amplicon generated in emulsion PCR and a stop codon where appropriate, DpnI digested (NEB) and column purified. Gibson assembly[45] was carried out at an approximately 1:1 molar ratio of insert to backbone. Reaction was column purified and electroporated into ElectroMAX Dh5b cells, and 1/10,000 was plated for titer, and rest was inoculated into liquid selective media.

Library P1 integration ($\mathbf{INT}_{P1}$). 2 µg of assembled library was digested with I-Ceu-I for 3 hours. Note that I-Ceu-I is sensitive to carryover of salt in the digest, so reactions are scaled to ensure a max of 20% of the volume of the digest was from miniprepped library DNA. Depending on the cell line being transformed, two transformations of frozen competent cells prepared as described[40] are sufficient to yield millions of CFU. Given the libraries contain around 5k genes this covers the libraries adequately. When transforming into cells with a lower-copy, error-prone polymerase, it is necessary to passage the cells in landing pad selective media 3–4 times before preparing competent cells, as they initially grow slowly and improved growth rate increases transformation efficiency. Post transformation, 1/5,000 of the total culture is plated on selective plates for p1 integration, and the rest (two transformations worth) is transferred to 15 mL selective culture immediately (no YPD recovery). OD should be measured at this point. The following day, a successful transformation will be saturated, and library can be banked as a starting point to be returned to. Libraries were banked in duplicate by spinning down 12 mL, and resuspending in 2 mL of 10% DMSO 5% glycerol and freezing in a isopropanol freezing chamber. Cells should be passaged 3x prior to selection as it takes approximately this many passages to stabilize expression from the integrated p1 cassettes. A GFP only control cloned into the same backbone is included alongside the library integrations to monitor expression and as an evolution control.

Yeast abortive mating (MATE). One day prior to abortive mating p1 donor library containing cells should be passaged to media only selecting for the linear plasmid (-H or -L). p1 receiver cells which contain only genomic modifications should be passaged to YPD to a final OD of 0.025. On the day of mating, the OD of both the p1 donor and receiver cells is measured, and 1 mL of OD 1, p1 receiver, and 0.5 mL of OD=1 p1 donor are mixed, spun down, and resuspended in 25 µL of 0.9% NaCl. This mix of cells is spotted onto a YPD plate and allowed to dry before the plate is flipped and incubated at 30 °C for 6 hours. At this scale, the yield of desired cells (correct haploid with p1 transfer) can be expected to be around 500k CFU. The mating can be scaled up or down as desired. For abortive mating in a 96-well plate, we scaled down 5-fold and spotted 4 columns of 8 rows using a multichannel pipette to fit 32 mating reactions on a single YPD plate. After 6 hrs the mated cells should be scraped into selective media. In my setup, for one direction this media is composed of -HR + Nurseothricin + Canavanine (no ammonium sulfate), mating into the opposite type is -HUK + Thialysine. The following day the second layer of selection is added in (-HR+Nat+Can+5FOA or -HUK+Thia+5-FC). Including 5-FOA immediately after mating will reduce the CFU by 10-fold, as even haploids formed as desired will still contain residual Ura3 in the cytoplasm, and vice-versa with 5-FC. Canavanine and Thialysine are not problematic for immediate selection as they are membrane bound transporters, and membranes are completely resynthesized with cell budding. After mating the cells are passaged in full counter selection 2x, then once in -H or -HU, then into selective media.

Antibiotic concentrations and media conditions for counterselection. Thialysine is used at a final concentration of 50 µg/mL in media lacking lysine (-K). Canavanine was used at 50 µg/mL in media lacking arginine (-R). 5-FOA was used at 500 µg/mL in SC media containing uracil. 5-FC was used at a concentration of 25 µg/mL in liquid SC.

Library evolution with retransformation (PCR + INT_{P1}). For the configuration of ORACLE utilizing the retransformation, frozen competent cells are prepared as described but banked as 12.5 µL aliquots rather than 50 µL aliquots. This is done to make the transformation compatible with recovery in a 24 well block. After a round of evolution (first round starting from the full library transformation), DNA is prepared by: measuring OD of each well and pipetting volume to reach an equivalent of 6E6 cells (1 mL of OD 0.4) into an Eppendorf tube. All are spun down at max speed for 15 seconds, media aspirated completely, washed with 1 mL of ddH2O, spun down again, and ddH2O aspirated completely. This is done to ensure consistency across wells as media carryover can impact PCR efficiency. Pellets are resuspended in 100 µL of 5% Chelex, and glass beads added to approximately ½ the volume. Tubes are vortexed for 6 minutes on a foam tube holder, and then boiled for 6 minutes on a heat block set to 105 °C. All tubes are spun down at max speed for 10 seconds, and 40 µL of DNA solution is transferred to a 96 well for storage. PCRs are conducted with Platinum Superfi 2x PCR mix (Thermo #12369010).

For PCR: a master mix of 10 µL 2x PCR mix, 6.75 µL of ddH2O, 0.125 µL 100 µM primers, (scaled up to reaction number) is prepared and distributed to the appropriate number of PCR tubes (16 uL/rxn). 3 µL of GC prepped DNA is distributed to prepared master mix. 17x cycles of PCR with a 4 min extension time (theoretically sufficient for 16kb) are run. 2 µL of PCR is used to assess PCR on an agarose gel. 13 µL of unpurified PCR mix is used as DNA template for transformation at 0.25x scale the regular transformation. With well prepared competent cells this yields about 100k transformants / transformation. Post transformation the cells are recovered in 2 mL -H overnight and passaged twice prior to selection. The number of passages in selective evolution media should be determined according to how quickly a genomic adaptation can take over the culture.

Evolution conditions (EVO). For all evolution conditions, the libraries are initially passaged to a starting OD of 0.1 in 2 mL (approximately 3E6 cells), and as the active sequences enrich, the amount of cells passaged is decreased, first to 0.05, and then to 0.025 OD. Cells are passaged every day. Reducing the cells passaged allows for faster enrichment at the cost of diversity.

- **Manganese:** Cells are passaged in 25 mM MnCl(II) media in -H. After the first round of evolution the concentration is increased to 30 mM, and then 35 mM.
- **Pantothenate dropout:** Media lacking pantothenate was prepared with YNB lacking calcium pantothenate (Formedium CYN3401).
- **Copper:** Media containing 2.5 mM Copper Sulfate, is gradually increased to 2.75, then 3.0 mM.
- **Lithium chloride:** Media containing 250 mM Lithium Chloride
- **Adenine dropout:** Transformations and regular culture are conducted in -H++A, media containing 80 mg/L adenine. For selective conditions, media lacking histidine and adenine is prepared, and gradual dropout of adenine is conducted. First 2.5 mM, then 1.25 mM, then 0 mM of supplemental adenine.

Fluorescent competition (COMP$_{X/Y}$). Competition strains which contain a Zeocin TP901 genomic landing pad, GFP or RFP fluorescence cassette, and TP901 2uM plasmid are prepared as frozen competent cells. DNA is prepared from evolution campaign by amplifying gene with primers which add 45 bp of homology to integration flanks. Integration flanks are prepared by PCRing a plasmid containing TP901 site-promoter, and -terminator-His3-TP901 site. Integration fragments are specific to the library as the cloning scar leftover from the original library differs based on the vector the library was cloned from. PCR of genes is done as for preparation of the re-transformation based evolution. 5 µL of each flank, and 5 µL of the variable gene are mixed, and transformed in a ½ scale frozen comp cell transformation (**INT**$_{genome}$). For example, the green cells are transformed with the genes from evolution, and the red cells transformed with a nonsense fragment.

After 2 days recovery, cells are passaged once 1:50 dilution. The following day the cells of interest are mixed with the control (nonsense integration or WT) 1:1 by volume. The OD of the mixture is noted to use for calculation of the starting OD in competition. The mixture of cells is inoculated at a 1:100 dilution (or depending on media condition) into selective media, and the ratio at day 0 is measured. The following day, OD is remeasured and flow is used to measure the cell ratio again, and the cells are again passaged noting the OD to monitor population expansion, and so on. A nonsense vs. nonsense control is run alongside the competition as eventually the competition becomes unstable due to genomic adaptation unrelated to the transformed gene. 20k cells were measured per well by flow for GFP/mRuby ratio in competition.

Calculation of relative growth rate and enrichment values from competition. We report relative growth rate values from the competition experiments. These can be simply calculated by using the population expansion of the two species in the same well. We measure the fractions of cells that are GFP/mRuby in the first timepoint, the starting OD600, and final fractions of GFP/mRuby in the end timepoint, and OD. As the OD is directly proportional to the # of cells in the culture, we calculate:

$$Generations(GFP) = \log_2 \frac{(\%GFP_f * OD_f)}{(\%GFP_i * OD_i)}$$

The same calculation is done for the mRuby species within the same well. This is converted to growth rate by:

$$Relative\ growth\ rate = \frac{Generations(GFP)}{Generations(mRuby)}$$

In the supplementary figures we report the enrichment which is calculated as below for the full competition experiment.

$$Enrichment\ (Green) = log2 \frac{\left(\frac{Fraction\ Green_{T_x}}{Fraction\ \ Red_{T_x}}\right)}{\left(\frac{Fraction\ Green_{T_0}}{Fraction\ \ Red_{T_0}}\right)}$$

ICP-MS for determination of intracellular copper concentration. Yeast strains expressing Wild-type or the evolved version of *RTS3* gene under RPL18b promoter from the HO locus were built. These strains were passaged twice in media containing 2.5 mM copper sulfate. 25 ml of saturated yeast culture was collected by centrifugation at 3000 g for 10 mins at 4C. The harvested yeast pellet was washed twice with 0.9% sodium chloride containing 5 mM EDTA. Then, the cells were washed 4 times with 0.9% sodium chloride to remove EDTA and the remaining extracellular copper. After each wash the supernatant was collected using a micropipette without disturbing the pellet. The harvested cells were resuspended with 250 ul of milliQ water and the optical density was measured using a spectrophotometer. The remaining sample was autoclaved for 20 mins at 121 degrees and sent to Solvias, Kaiseraugst, Switzerland for ICP-MS. The samples were decomposed by addition of oxidizing acids and heating in a closed-vessel, microwave-heated, high-pressure autoclave. The concentration of Cu was determined using an Agilent 7900 instrument. The quantitation limit/ reporting limit for the measurement was 0.1 mg/kg.

96 well frozen competent cell preparation and transformation. Frozen competent cells are prepared as described above with a couple key changes to facilitate easier transformation with no centrifugation. The same outgrowth conditions are used, but one extra water wash step is conducted. Cells are frozen at the same concentration of competent cell mix, but 4 µL is aliquoted per well of a 96 well plate and frozen in a Styrofoam box. On day of transformation the DNA / PEG / LiAc / Master mix is prepared to 45 µL / well containing 32.5 µL PEG, 4.5 µL LiAc, 1.25 µL ssDNA, 2.25 µL H2O, 1.5 µL of the PCR gene amplicon, and 1.5 µL of each flank. For transformation 96 well plates are thawed by placing at 37C for 1 minute, and foil is removed. Thawed cells in DMSO / Glycerol mix are resuspended with 45 µL of transformation mix by pipetting (note, these are not spun down and frozen cell mix is left in the transformation), and incubated at 42 °C for 30 min in a thermocycler. Post heat shock, the mixture of cells / PEG is diluted to 1 mL in SC media + selection by pipetting (again note, the PEG / DNA mixture is not removed from the outgrowth). The 1 mL is split into two wells for bioreplicates as the transformed cells will be separate transformation events. These are outgrown for two days and 1/25 of desired wells are plated for titer. Cells will be saturated after two days and generally 1-5,000 transformants per well can be obtained with the genomic TP901 landing pad method.

Luciferase assay. Nanoluciferase containing strains transformed with the 96 well method were outgrown two days to saturation, passaged twice more 1:500 for 24 hrs each passage. Morning of luciferase measurement cells were passaged 1:20 in -H media. After 5 hours of outgrowth OD of wells was measured. OD of all wells was within a 10% range indicating all were in exponential growth. Cells were diluted 1:3 in 0.9% NaCl to a final OD of approximately 0.5-0.6. 8 µL of diluted cells was transferred to a

white, opaque 384 well plate, with no two wells adjacent to another occupied well to minimize cross talk. Sample was mixed with Promega nano luciferase kit (N1110) 1:1. Luminescence was read on Biotek Synergy H1 with integration time 1 sec, gain 135. Nonsense control was included as a sample of in plates to ensure normalization was accurate. Luminescence was first normalized by OD, and luminescence was reported as log2(RLU sample/RLU control) for the appropriate cell line.

GFP and mRuby degradation assays. Fluorescent reporter containing strains transformed with the 96 well method were outgrown two days to saturation, passaged twice more 1:500 for 24 hrs each passage. Cells in late log phase were diluted in 0.9% NaCl with propidum iodide and gated for viability, FSC/SSC and FSC-A / FSC-H for single cells. 20k cells per well were collected, log2(MFI Geomean sample/MFI Geomean control) is reported in these assays.

Yeast growth curves and max growth rate calculation. To measure the max growth rate yeast were grown to saturation, then passaged 1:500 twice (once per day). On the morning of the growth curve, yeast were back diluted 1:10 and grown for 4 hours to ensure that all cells were in exponential growth prior to inoculation. Cells in exponential growth were inoculated 1:100 (1 µL into 100 µL of -HA) of a clear bottom 96 well plate. Samples were measured in biological duplicate. Plate was covered with a breathable film and transferred to a biotek plate reader shaking at 500 rpm, 30 °C, with OD600 measured every 15 minutes. The growth curve was complete within 24 hours of growth. To calculate the growth rate the OD was ln transformed, and the slope of the linear range was fit using graphpad.

***E. coli* growth curves and max growth rate calculation**. To measure the max growth rate *E. coli* containing IPTG inducible plasmids were grown to saturation, then back diluted 1:20 in 2YT+IPTG and grown for 2 hours to ensure that all cells were in exponential growth and expressing the plasmid prior to inoculation in the Copper media. Cells in exponential growth were inoculated 1:100 (1 µL into 100 µL of 2YT + 4 mM copper chloride) of a clear bottom 96 well plate. Plate was covered with a breathable film and transferred to a biotek plate reader shaking at 500 rpm, 37 C, with OD_{600} measured every 15 minutes. The growth curve was complete within 24 hours of growth. To calculate the growth rate the OD was ln transformed, and the slope of the linear range was fit using graphpad.

Fluorescence-Activated Cell Sorting for *BAX* overexpression evolution: A base strain expressing *BAX*-2A-GFP from an inducible Gal promoter and a TP901 p1 landing pad was constructed. The yeast and E.coli ORF plasmid libraries were digested and integrated as discussed above. Library was passaged once to 0.2% Glucose, 2% Raffinose synthetic dropout media and then induced in 2% Galactose media lacking Glucose. All samples were OD normalized to 0.5 for all the passages. 2-3 days post-induction in Galactose media, cells were live-dead stained using the TO-PRO-3 Stain (Thermo-fisher scientific) and sorted for 100 thousand GFP+ cells. After each two rounds of Induction+Sorting, the p1 library was PCRed and re-integrated to fresh strains as mentioned above. This cycle was repeated for 4 sorts in total.

Supplementary Information Outline

Supplementary Text 1: Origins of libraries

Supplementary Text 2: Description of non-active gene phenomena from mating-based evolutions

Supplementary Text 3: Design of a good selection

Supplementary Text 4: Future technology developments for ORACLE

Figure S1. Optimization of emulsion PCR and demonstration of decreased size bias in **PCR**$_{lib}$

Figure S2. Diagram of assessment transformation method and competition setup

Figure S3. Visual representation of the abortive mating technique (**MATE**)

Figure S4. Validation of capacity of cells to maintain and express from multiple P1s

Figure S5. Demonstration of ability to fish out a specific combination of genes at a frequency of <1/10,000,000

Figure S6. Schematic of **PCR**$_{lib}$ + **INT**$_{p1}$

Figure S7. Full competition data from the pantothenate evolutions relating to Figure 2

Figure S8: Full competition data from the manganese tolerance evolutions supporting conclusions made in the main text relating to Figure 3

Figure S9. Full competition data from the lithium tolerance evolutions supporting conclusions made in the main text relating to Figure 3

Figure S10. Full competition data from the copper tolerance evolutions supporting conclusions made in the main text relating to Figure 3

Figure S11. Competition and reporter data supporting conclusions made in the main text relating to Figure 4

Figure S12. Competition and reporter data supporting conclusions made in the main text relating to Figure 5

Figure S13. Full competition and reporter data supporting conclusions made in the main text relating to Figure 6

Figure S14. Supplementary Fig 14: Full competition and reporter data supporting conclusions made in the main text relating to Figure 6

Figure S15. Full competition and reporter data supporting conclusions made in the main text relating to Figure 6

Figure S16. Evolved NNT1(D19N, K49R, G60S, G90D, N135D, A219T) significantly enriches over both NS control and WT NNT1 during BAX-induced apoptosis

Table S1. Expanded table of genes and variants discovered in evolution campaigns

Table S2. Important primers used in study

Table S3. Plasmids used in study

Table S4. Yeast strains and genotypes used in study

Zip: Annotated genbank plasmid files of all plasmids used in study

Supplementary References

Author Contributions

A.P. conceived the project ideas with input from C.C.L. A.P. and A.T. designed and carried out all experiments, collected and processed all data, and analyzed all data with input from C.C.L. A.P. and C.C.L. wrote the manuscript.

Acknowledgements

This work was funded by NIH R35GM136297 to C.C.L. A.P. is supported by a Paul & Daisy Soros Fellowship for New Americans. We thank Professor Wayne Patrick for his valuable contribution of the pooled ASKA library.

References

1. Baier, F., Copp, J. N. & Tokuriki, N. Evolution of Enzyme Superfamilies: Comprehensive Exploration of Sequence-Function Relationships. *Biochemistry* **55**, 6375–6388 (2016).
2. Jensen, R. A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409–425 (1976).
3. Kondrashov, F. A. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. R. Soc. B Biol. Sci.* **279**, 5048–5057 (2012).
4. Ohno, S. *Evolution by Gene Duplication*. (Springer-Verlag, 1970).
5. Bergthorsson, U., Andersson, D. I. & Roth, J. R. Ohno's dilemma: Evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci.* **104**, 17004–17009 (2007).
6. Näsvall, J., Sun, L., Roth, J. R. & Andersson, D. I. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* **338**, 384–387 (2012).
7. McLysaght, A. & Hurst, L. D. Open questions in the study of de novo genes: what, how and why. *Nat. Rev. Genet.* **17**, 567–578 (2016).
8. Montañés, J. C., Huertas, M., Messeguer, X. & Albà, M. M. Evolutionary Trajectories of New Duplicated and Putative De Novo Genes. *Mol. Biol. Evol.* **40**, msad098 (2023).

9. Parikh, S. B., Houghton, C., Van Oss, S. B., Wacholder, A. & Carvunis, A.-R. Origins, evolution, and physiological implications of de novo genes in yeast. *Yeast Chichester Engl.* **39**, 471–481 (2022).
10. Vakirlis, N. *et al.* A Molecular Portrait of De Novo Genes in Yeasts. *Mol. Biol. Evol.* **35**, 631–645 (2018).
11. Van Oss, S. B. & Carvunis, A.-R. De novo gene birth. *PLoS Genet.* **15**, e1008160 (2019).
12. Blevins, W. R. *et al.* Uncovering de novo gene birth in yeast using deep transcriptomics. *Nat. Commun.* **12**, 604 (2021).
13. Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet.* **9**, e1003569 (2013).
14. Ingolia, N. T. *et al.* Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Rep.* **8**, 1365–1379 (2014).
15. Ardern, Z. Alternative Reading Frames are an Underappreciated Source of Protein Sequence Novelty. *J. Mol. Evol.* **91**, 570–580 (2023).
16. Barona-Gómez, F. & Hodgson, D. A. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep.* **4**, 296–300 (2003).
17. Soo, V. W. C., Hanson-Manful, P. & Patrick, W. M. Artificial gene amplification reveals an abundance of promiscuous resistance determinants in Escherichia coli. *Proc. Natl. Acad. Sci.* **108**, 1484–1489 (2011).
18. Babina, A. M. *et al.* Rescue of Escherichia coli auxotrophy by de novo small proteins. *eLife* **12**, (2023).

19. Digianantonio, K. M. & Hecht, M. H. A protein constructed de novo enables cell growth by altering gene regulation. *Proc. Natl. Acad. Sci.* **113**, 2400–2405 (2016).
20. Donnelly, A. E., Murphy, G. S., Digianantonio, K. M. & Hecht, M. H. A de novo enzyme catalyzes a life-sustaining reaction in Escherichia coli. *Nat. Chem. Biol.* **14**, 253–255 (2018).
21. Fisher, M. A., McKinley, K. L., Bradley, L. H., Viola, S. R. & Hecht, M. H. De novo designed proteins from a library of artificial sequences function in Escherichia coli and enable cell growth. *PloS One* **6**, e15364 (2011).
22. Frumkin, I. & Laub, M. T. Selection of a de novo gene that can promote survival of Escherichia coli by modulating protein homeostasis pathways. *Nat. Ecol. Evol.* **7**, 2067–2079 (2023).
23. Hoegler, K. J. & Hecht, M. H. A de novo protein confers copper resistance in Escherichia coli. *Protein Sci. Publ. Protein Soc.* **25**, 1249–1259 (2016).
24. Keefe, A. D. & Szostak, J. W. Functional proteins from a random-sequence library. *Nature* **410**, 715–718 (2001).
25. Knopp, M. *et al.* A novel type of colistin resistance genes selected from random sequence space. *PLoS Genet.* **17**, e1009227 (2021).
26. Smith, B. A., Mularz, A. E. & Hecht, M. H. Divergent evolution of a bifunctional de novo protein. *Protein Sci. Publ. Protein Soc.* **24**, 246–252 (2015).
27. Tong, C. L., Lee, K.-H. & Seelig, B. De novo proteins from random sequences through in vitro evolution. *Protein-Carbohydr. Complexes Glycosylation ● Seq. Topol.* **68**, 129–134 (2021).
28. Huberts, D. H. & van der Klei, I. J. Moonlighting proteins: an intriguing mode of multitasking. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1803**, 520–525 (2010).
29. Jeffery, C. J. An enzyme in the test tube, and a transcription factor in the cell: Moonlighting proteins and cellular factors that affect their behavior. *Protein Sci.* **28**, 1233–1238 (2019).

30. Singh, N. & Bhalla, N. Moonlighting Proteins. *Annual Review of Genetics* vol. 54 265–285 (2020).

31. Copley, S. D. Shining a light on enzyme promiscuity. *Protein–nucleic Acid Interact. • Catal. Regul.* **47**, 167–175 (2017).

32. Copley, S. D., Newton, M. S. & Widney, K. A. How to Recruit a Promiscuous Enzyme to Serve a New Function. *Biochemistry* **62**, 300–308 (2023).

33. Khersonsky, O., Roodveldt, C. & Tawfik, D. S. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508 (2006).

34. Rix, G. *et al.* Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nat. Commun.* **11**, 5644 (2020).

35. Molina, R. S. *et al.* In vivo hypermutation and continuous evolution. *Nat. Rev. Methods Primer* **2**, 36 (2022).

36. Rix, G. & Liu, C. C. Systems for in vivo hypermutation: a quest for scale and depth in directed evolution. *Curr. Opin. Chem. Biol.* **64**, 20–26 (2021).

37. Ravikumar, A., Arrieta, A. & Liu, C. C. An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* **10**, 175–177 (2014).

38. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A. & Liu, C. C. Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell* **175**, (2018).

39. Rix, G. *et al.* Continuous evolution of user-defined genes at 1 million times the genomic mutation rate. *Science* **386**, eadm9073 (2024).

40. Pisera, A., Yu, Y., Williams, R. L. & Liu, C. C. Ultra-Efficient Integration of Gene Libraries onto Yeast Cytosolic Plasmids. *bioRxiv* 2024.11.29.626108 (2024) doi:10.1101/2024.11.29.626108.

41. Williams, R. *et al.* Amplification of complex gene libraries by emulsion PCR. *Nat. Methods* **3**, 545–550 (2006).

42. Gelperin, D. M. *et al.* Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev.* **19**, 2816–2826 (2005).

43. Kitagawa, M. *et al.* Complete set of ORF clones of Escherichia coli ASKA library (A Complete S et of E. coli K -12 ORF A rchive): Unique Resources for Biological Research. *DNA Res.* **12**, 291–299 (2005).

44. Patrick, W. M., Quandt, E. M., Swartzlander, D. B. & Matsumura, I. Multicopy Suppression Underpins Metabolic Evolvability. *Mol. Biol. Evol.* **24**, 2716–2722 (2007).

45. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

46. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).

47. Xu, Z. & Brown, W. R. Comparison and optimization of ten phage encoded serine integrases for genome engineering in saccharomyces cerevisiae. *BMC Biotechnol.* **16**, (2016).

48. Chakrabortee, S. *et al.* Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. *Cell* **167**, 369-381.e12 (2016).

49. Georgieva, B. & Rothstein, R. kar-mediated plasmid transfer between yeast strains: Alternative to traditional transformation methods. in *Methods in Enzymology* (eds. Guthrie, C. & Fink, G. R.) vol. 350 278–289 (Academic Press, 2002).

50. Vallen, E., Hiller, M., Scherson, T. & Rose, M. Separate domains of KAR1 mediate distinct functions in mitosis and nuclear fusion. *J. Cell Biol.* **117**, 1277–1287 (1992).

51. Boeke, J. D., La Croute, F. & Fink, G. R. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet. MGG* **197**, 345–346 (1984).

52. Erbs, P., Exinger, F. & Jund, R. Characterization of the Saccharomyces cerevisiae FCY1 gene encoding cytosine deaminase and its homologue FCA1 of Candida albicans. *Curr. Genet.* **31**, 1–6 (1997).

53. Larimer, F. W., Ramey, D. W., Lijinsky, W. & Epler, J. L. Mutagenicity of methylated N-nitrosopiperidines in Saccharomyces cerevisiae. *Mutat. Res. Mol. Mech. Mutagen.* **57**, 155–161 (1978).

54. Sychrova, H. & Chevallier, M. R. Cloning and sequencing of the Saccharomyces cerevisiae gene LYP1 coding for a lysine-specific permease. *Yeast* **9**, 771–782 (1993).

55. Toyn, J. H., Gunyuzlu, P. L., Hunter White, W., Thompson, L. A. & Hollis, G. F. A counterselection for the tryptophan pathway in yeast: 5-fluoroanthranilic acid resistance. *Yeast* **16**, 553–560 (2000).

56. Abdullah, M. *et al.* Rapid, scalable, combinatorial genome engineering by marker-less enrichment and recombination of genetically engineered loci in yeast. *Cell Rep. Methods* **3**, (2023).

57. Pisera, A. (Olek) & Liu, C. C. A System to Explore the Adaptive Dynamics of Multicopy Plasmids: The Role of Copy Number and Mutation Rate in Evolutionary Outcomes. *bioRxiv* 2024.12.13.628408 (2024) doi:10.1101/2024.12.13.628408.

58. Blount, B. A., Driessen, M. R. & Ellis, T. GC PREPS: Fast and easy extraction of stable yeast genomic DNA. *Sci. Rep.* **6**, (2016).

59. Gietz, R. D. & Schiestl, R. H. Frozen competent yeast cells that can be transformed with high efficiency using the LIAC/SS carrier DNA/peg method. *Nat. Protoc.* **2**, 1–4 (2007).

60. Ball, S. G., Wickner, R. B., Cottarel, G., Schaus, M. & Tirtiaux, C. Molecular cloning and characterization of ARO7-OSM2, a single yeast gene necessary for chorismate mutase activity and growth in hypertonic medium. *Mol. Gen. Genet. MGG* **205**, 326–330 (1986).

61. Bracher Jasmine M. *et al.* Laboratory Evolution of a Biotin-Requiring Saccharomyces cerevisiae Strain for Full Biotin Prototrophy and Identification of Causal Mutations. *Appl. Environ. Microbiol.* **83**, e00892-17 (2017).

62. Perli, T., Wronska, A. K., Ortiz-Merino, R. A., Pronk, J. T. & Daran, J.-M. Vitamin requirements and biosynthesis in Saccharomyces cerevisiae. *Yeast* **37**, 283–304 (2020).

63. Taylor, M. B. *et al.* yEvo: experimental evolution in high school classrooms selects for novel mutations that impact clotrimazole resistance in Saccharomyces cerevisiae. *G3 GenesGenomesGenetics* **12**, jkac246 (2022).

64. White, W. H., Gunyuzlu, P. L. & Toyn, J. H. Saccharomyces cerevisiae Is Capable of de Novo Pantothenic Acid Biosynthesis Involving a Novel Pathway of β-Alanine Production from Spermine*. *J. Biol. Chem.* **276**, 10794–10800 (2001).

65. Li, L., Chen, O. S., Ward, D. M. & Kaplan, J. CCC1 Is a Transporter That Mediates Vacuolar Iron Storage in Yeast *. *J. Biol. Chem.* **276**, 29515–29519 (2001).

66. Lahti, R. *et al.* Conservation of functional residues between yeast and E. coli inorganic pyrophosphatases. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* **1038**, 338–345 (1990).

67. Srividhya, K. V. & Krishnaswamy, S. Sub classification and targeted characterization of prophage-encoded two-component cell lysis cassette. *J. Biosci.* **32**, 979–990 (2007).

68. Hood-DeGrenier, J. K. Identification of phosphatase 2A-like Sit4-mediated signalling and ubiquitin-dependent protein sorting as modulators of caffeine sensitivity in S. cerevisiae. *Yeast* **28**, 189–204 (2011).

69. Wang, M. S., Hoegler, K. J. & Hecht, M. H. Unevolved De Novo Proteins Have Innate Tendencies to Bind Transition Metals. *Life Basel Switz*. **9**, (2019).

70. Dorfman, B.-Z. THE ISOLATION OF ADENYLOSUCCINATE SYNTHETASE MUTANTS IN YEAST BY SELECTION FOR CONSTITUTIVE BEHAVIOR IN PIGMENTED STRAINS. *Genetics* **61**, 377–389 (1969).

71. Crooijmans, M. E. *et al*. Cell-to-cell heterogeneity of phosphate gene expression in yeast is controlled by alternative transcription, 14-3-3 and Spl2. *Biochim. Biophys. Acta BBA - Gene Regul. Mech*. **1864**, 194714 (2021).

72. Korber, P. & Barbaric, S. The yeast PHO5 promoter: from single locus to systems biology of a paradigm for gene regulation through chromatin. *Nucleic Acids Res*. **42**, 10888–10902 (2014).

73. Ragunathan Preethi T., Ng Kwan Lim Evelyne, Ma Xiangqian, Massé Eric, & Vanderpool Carin K. Mechanisms of Regulation of Cryptic Prophage-Encoded Gene Products in Escherichia coli. *J. Bacteriol*. **205**, e00129-23 (2023).

74. Vardi, N., Levy, S., Assaf, M., Carmi, M. & Barkai, N. Budding Yeast Escape Commitment to the Phosphate Starvation Program Using Gene Expression Noise. *Curr. Biol*. **23**, 2051–2057 (2013).

75. Bartkiewicz, M. *et al*. Non–fluorescent mutant of green fluorescent protein sheds light on the mechanism of chromophore formation. *FEBS Lett*. **592**, 1516–1523 (2018).

76. Gonzalez, C. F. *et al*. Molecular Basis of Formaldehyde Detoxification: CHARACTERIZATION OF TWO S-FORMYLGLUTATHIONE HYDROLASES FROM ESCHERICHIA COLI, FrmB AND YeiG *. *J. Biol. Chem*. **281**, 14514–14522 (2006).

77. Hackett, E. A., Esch, R. K., Maleri, S. & Errede, B. A family of destabilized cyan fluorescent proteins as transcriptional reporters in S. cerevisiae. *Yeast* **23**, 333–349 (2006).

78. Bachmair, A., Finley, D. & Varshavsky, A. In Vivo Half-Life of a Protein Is a Function of Its Amino-Terminal Residue. *Science* **234**, 179–186 (1986).

79. Trulsson, F. *et al.* Deubiquitinating enzymes and the proteasome regulate preferential sets of ubiquitin substrates. *Nat. Commun.* **13**, 2736 (2022).

80. Szoradi, T. *et al.* SHRED Is a Regulatory Cascade that Reprograms Ubr1 Substrate Specificity for Enhanced Protein Quality Control during Stress. *Mol. Cell* **70**, 1025-1037.e5 (2018).

81. Onozuka, M., Konno, H., Kawasaki, Y., Akaji, K. & Nosaka, K. Involvement of thiaminase II encoded by the THI20 gene in thiamin salvage of Saccharomyces cerevisiae. *FEMS Yeast Res.* **8**, 266–275 (2008).

82. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

83. Peters, N. *et al.* Reprograming of the ubiquitin ligase Ubr1 by intrinsically disordered Roq1 through cooperating multifunctional motifs. *EMBO J.* **44**, 1774–1803 (2025).

84. Lacombe, T. *et al.* Linear ubiquitin fusion to Rps31 and its subsequent cleavage are required for the efficient production and functional integrity of 40S ribosomal subunits. *Mol. Microbiol.* **72**, 69–84 (2009).

85. Gebendorfer, K. M. *et al.* Identification of a Hypochlorite-specific Transcription Factor from Escherichia coli*. *J. Biol. Chem.* **287**, 6892–6903 (2012).

86. Berset, C. *et al.* Transferable Domain in the G1 Cyclin Cln2 Sufficient To Switch Degradation of Sic1 from the E3 Ubiquitin Ligase SCFCdc4 to SCFGrr1. *Mol. Cell. Biol.* **22**, 4463–4476 (2002).

87. A. Kuno, M. Taki, K. Taira, & T. Hasegawa. Leucyl/Phenylalanyl (L/F)-tRNA-protein transferase-mediated N-terminal specific labelling of a protein in vitro. in *Nucleic Acids Res. Supl.* vol. 3 259–260 (2003).

88. Kachroo, A. H. *et al.* Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* **348**, 921–925 (2015).

89. Manon, S., Chaudhuri, B. & Guérin, M. Release of cytochrome c and decrease of cytochrome c oxidase in Bax-expressing yeast cells, and prevention of these effects by coexpression of Bcl-xL. *FEBS Lett.* **415**, 29–32 (1997).

90. Zha, H. *et al.* Structure-function comparisons of the proapoptotic protein Bax in yeast and mammalian cells. *Mol. Cell. Biol.* **16**, 6494–6508 (1996).

91. Ionov, Y., Yamamoto, H., Krajewski, S., Reed, J. C. & Perucho, M. Mutational inactivation of the proapoptotic gene BAX confers selective advantage during tumor clonal evolution. *Proc. Natl. Acad. Sci.* **97**, 10872–10877 (2000).

92. Libby, R. T. *et al.* Susceptibility to Neurodegeneration in a Glaucoma Is Modified by Bax Gene Dosage. *PLOS Genet.* **1**, e4 (2005).

93. Spitz, A. Z., Zacharioudakis, E., Reyna, D. E., Garner, T. P. & Gavathiotis, E. Eltrombopag directly inhibits BAX and prevents cell death. *Nat. Commun.* **12**, 1134 (2021).

94. Perez, G. I. *et al.* Prolongation of ovarian lifespan into advanced chronological age by Bax-deficiency. *Nat. Genet.* **21**, 200–203 (1999).