

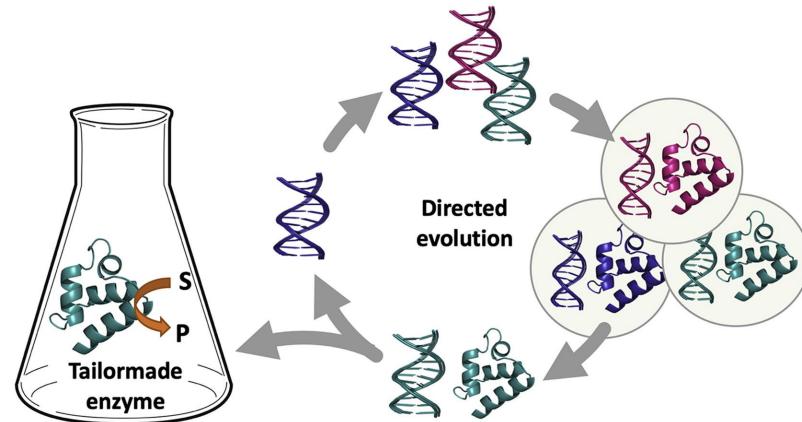
Lecture 8 - Evolutionary Engineering Methods for Biotechnology and Biomedical Applications

BENG168

**Instructor: Adam M. Feist, Assistant Professor, Shu Chien -
Gene Lay Department of Bioengineering**

Evolution vs. Rational Design

- **Structural Knowledge:** Rational design (Lecture 7) requires high-resolution blueprints, whereas directed evolution works without *a priori* mechanistic data.
- **Complexity Handling:** Directed evolution mimics natural selection on a laboratory timescale to solve problems too complex for human design.
- **Iterative Nature:** Success is achieved by repeatedly cycling through mutation and screening or selection to climb "fitness peaks" in a protein's functional landscape.



doi.org/10.1016/j.cbsa.2020.11.006

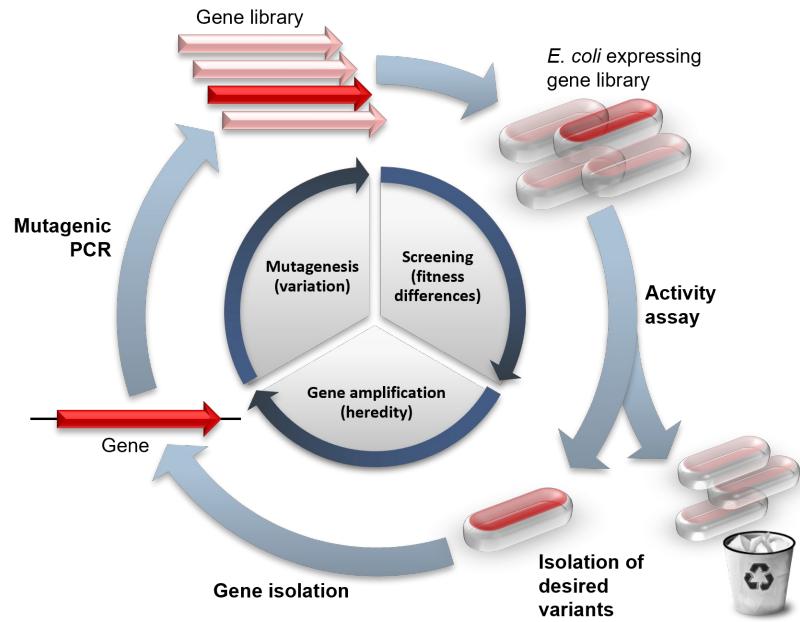
Module 1: Directed Evolution (In Vitro)

(Source Pages: Chapter 3: 161–164, 168–173)

- **Genetic Diversification:** Mutate genes outside living cells.
- Screen libraries for superior variants.
- Toolbox for targeted single-protein optimization.

The Directed Evolution Cycle

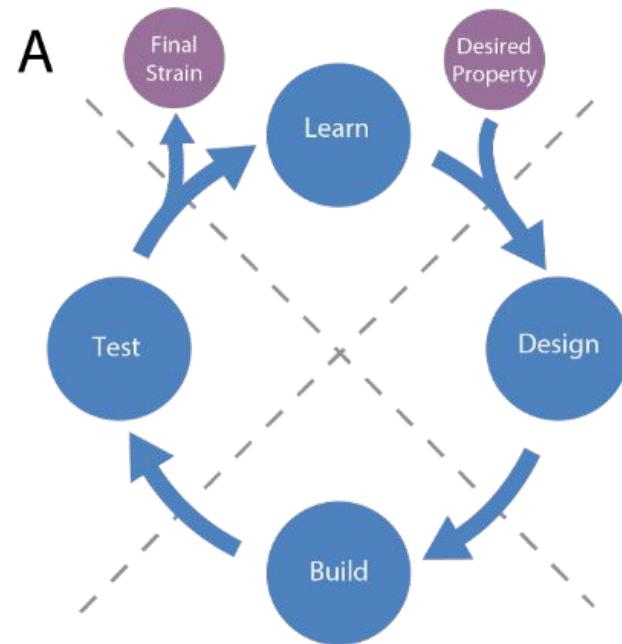
- **Library Creation:** The first step involves generating a large library of genetic variants using random or targeted mutagenesis.
- **Host Expression:** These genetic variants are expressed in a suitable host system to produce the actual proteins for testing.
- **Screening Winners:** Researchers assay the library for desired properties, using the best variants as templates for the next round of evolution.



By Thomas Shafee - Thomas, Shafee, (2014). PhD Thesis.
University of Cambridge.

The Design-Build-Test-Learn (DBTL) Loop

- **Design Phase:** Define biological targets and hypothesize genetic modifications using *in silico* models or other strategies.
- **Build Phase:** Physicalize genetic blueprints in a host organism using genome engineering tools (**molecular cloning** or **CRISPR-Cas9**).
- **Test Phase:** Measure strain performance through rigorous phenotypic assays and high-throughput **multi-omics analysis** against original design goals.
- **Learn Phase:** Analyze collected datasets to identify causal mutations and refine hypotheses for the next iteration (**bioinformatics** or **modeling**).



The Design, Build, Test, Learn cycle. Sandberg et al. and Feist.
Metab Eng. 2019; doi: 10.1016/j.ymaben.2019.08.004

Error-Prone PCR (epPCR)

- **Low-Fidelity Replication:** This method uses a DNA polymerase that lacks proofreading ability, intentionally introducing random mutations during amplification.
- **Reaction Tweaking:** Adding manganese ions or unbalancing the concentrations of dNTPs further increases the frequency of errors.
- **Diversity Scale:** This simple approach generates thousands of unique variants of a single gene, providing ample material for large-scale screening.

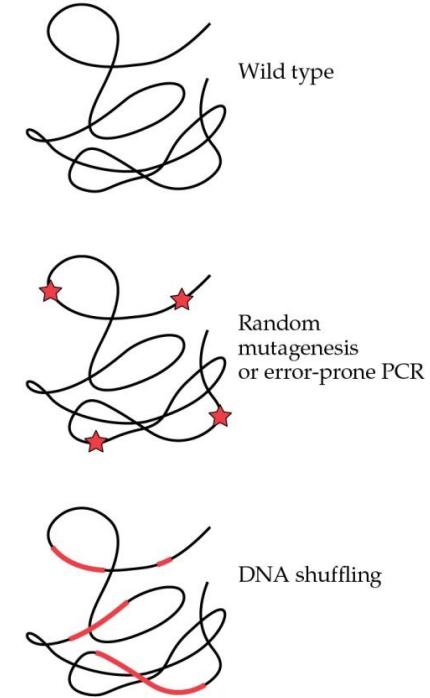


Figure 3.56 Amino acid changes may be introduced at random positions in a protein by either error-prone PCR or PCR using degenerate oligonucleotide primers, both of which cause single amino acid substitutions (red stars), and by DNA shuffling, in which genes are formed with large regions from different sources (red lines).

Random Mutagenesis: Degenerate Primers

- **Doped Synthesis:** Primers are synthesized with a mixture of all four nucleotides at specific sites to create sequence degeneracy.
- **Residue Randomization:** This allows every possible amino acid substitution at a specific target site without needing prior structural data.
- **Focused Mutations:** It is particularly effective when mutations need to be clustered in a defined, **critical region of the protein**.

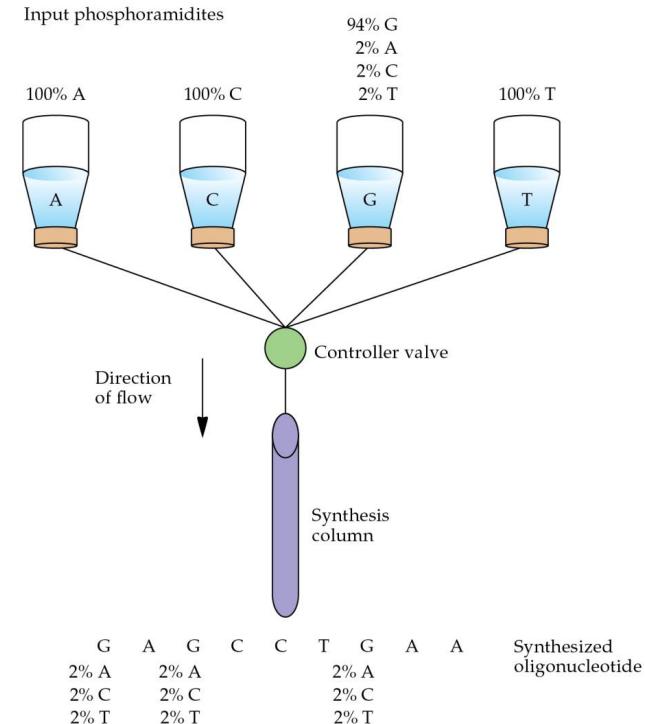


Figure 3.55 Chemical synthesis of oligonucleotide primers with any of the four nucleotides at defined positions. In this case, the flask with G phosphoramidite consists of a mixture of nucleotides, such as 94% G, 2% A, 2% C, and 2% T, leading to a mixture of oligonucleotides that may have A, C, or T at the sites where G is the specified nucleotide.

DNA Shuffling (Molecular Breeding)

- **Gene Fragmentation:** Related genes are fragmented by DNase I and then reassembled by PCR to create chimeric sequences.
- **Trait Recombination:** This "molecular breeding" creates hybrid genes that possess novel combinations of traits from different family members.
- **Synergistic Evolution:** Shuffling is more effective than single mutations for evolving complex traits like simultaneous thermostability and high activity.

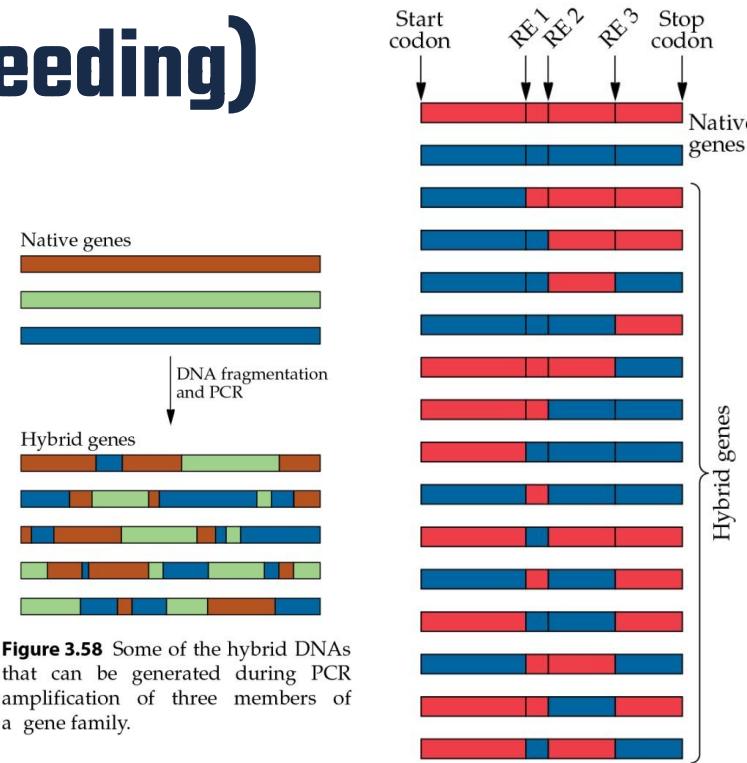


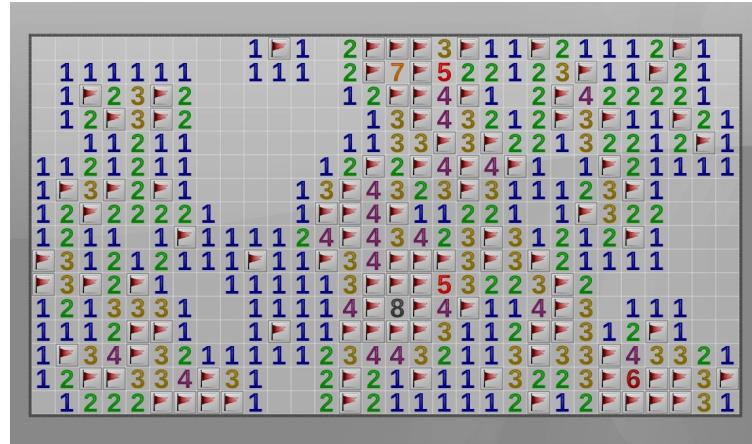
Figure 3.58 Some of the hybrid DNAs that can be generated during PCR amplification of three members of a gene family.

Figure 3.57 The 14 different hybrid genes that can be generated by combining restriction enzyme fragments from two genes from the same gene family that have three different restriction sites in common. RE, restriction enzyme.

Directed Evolution Parameters: Library Size & Screening

- **Variant Diversity:** Experiments generate massive libraries to maximize the chance of beneficial mutations.
- **Throughput Limits:** Assays are the main bottleneck; individual screening manages only thousands.
- **Statistical Probability:** Success requires a library size providing a 99% probability of discovery.

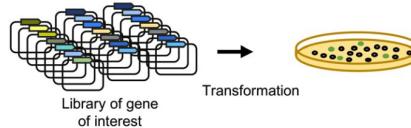
A Numbers Game



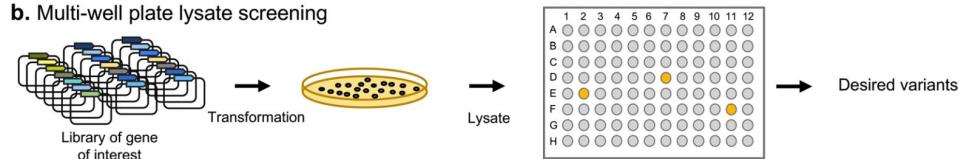
Screening Techniques

- **Defined Assays:** Spatially separate variants in plates for robotic activity measurements.
- **Flow Cytometry:** FACS sorts individual cells based on fluorescent signals with massive throughput.
- **Compartmentalization:** Emulsion droplets isolate variants for assays exceeding standard transformation limits.

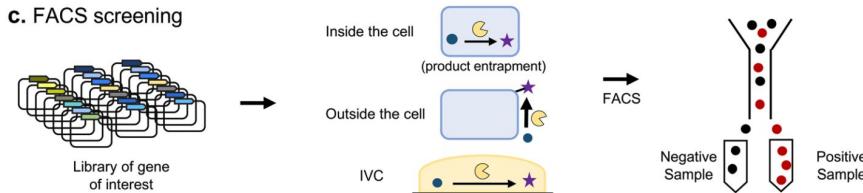
a. Fluorescence screening



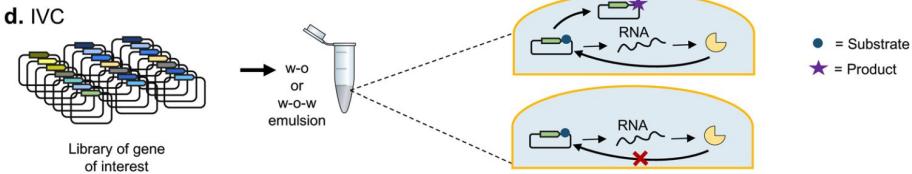
b. Multi-well plate lysate screening



c. FACS screening



d. IVC



Screening Techniques

Selles Vidal et al. RSC Chem. Biol., 2023, 4, 271 DOI:
[10.1039/d2cb00231k](https://doi.org/10.1039/d2cb00231k)

Case Study: Subtilisin Stability

- **Calcium Dependency:** Industrial subtilisins (biodegradable protease cleaning agents) require calcium for stability, but laundry detergents contain chelators that strip calcium away.
- **Loop Deletion:** Researchers deleted the calcium-binding loop and used random mutagenesis to restore the enzyme's structural integrity.
- **Restored Activity:** Identifying 7 stabilizing mutations resulted in a calcium-free enzyme that was even more stable than the native form.

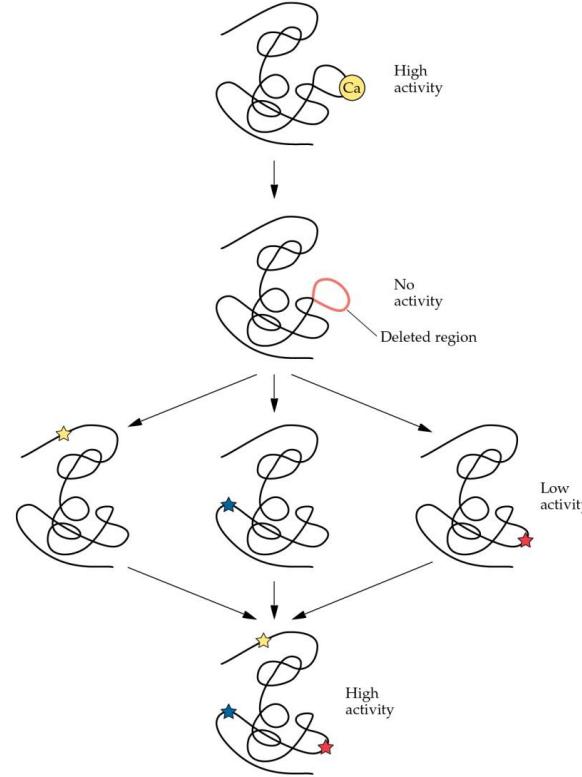


Figure 3.63 Genetic engineering of calcium-independent subtilisin. The native calcium-containing enzyme is highly active but lost almost all of its activity when the loop that binds calcium was deleted. The gene encoding the enzyme with the deleted calcium-binding region was subjected to random mutagenesis. After several rounds of mutagenesis, mutants with stabilizing mutations and a low level of activity were selected. Several of these mutations (indicated by colored stars) were combined into a single derivative to generate subtilisin that does not require calcium and has a high level of activity.

Case Study: Protease Sensitivity (Directed)

- **Degradation Challenge:** The cancer-fighting protein APPI is too rapidly degraded by host proteases.
- **Yeast Screening:** Error-prone PCR and yeast surface display were used to screen for variants with ultra-high binding affinity.
- **Metastatic Inhibition:** A triple mutant achieved 1,500-fold higher affinity, successfully blocking the invasiveness of aggressive prostate cancer cells.

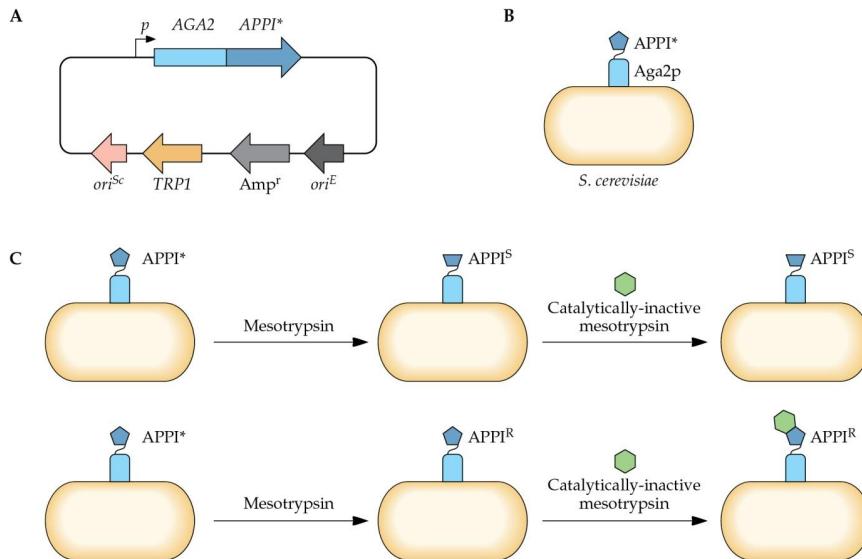
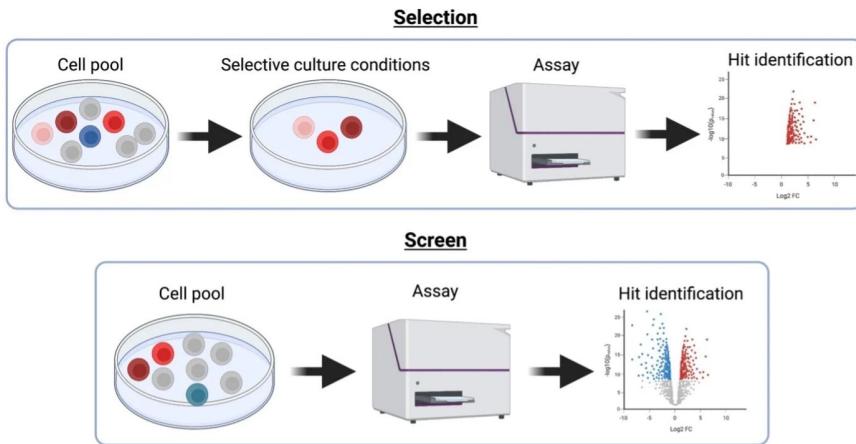


Figure 3.64 Random mutagenesis to decrease protease sensitivity of a protease inhibitor. (A) The gene encoding the protease inhibitor domain of the human amyloid precursor protein (*APPI*) was amplified by error-prone PCR and cloned into a vector in the same reading frame as the coding sequence for the yeast Aga2p agglutinin protein (*AGA2*) to generate a library of *APPI* variants (*APPI**) in *S. cerevisiae*. (B) The Aga2p fusion partner directed *APPI** to the surface of *S. cerevisiae* cells to facilitate screening of protease-resistant variants. (C) The *S. cerevisiae* library was incubated with the protease mesotrypsin to cleave susceptible *APPI* variants (*APPI^S*), including wild-type *APPI*, and then with a labeled, mutant form of mesotrypsin that does not have proteolytic activity but binds to intact *APPI* to identify protease-resistant *APPI* (*APPI^R*). *p*, *GAL1* promoter; *ori^{Sc}* and *TRP1*, origin of replication and selectable marker for *S. cerevisiae*, respectively; *ori^E* and *Amp^R*, origin of replication and selectable marker for *E. coli*, respectively. Data from Cohen et al., *Biochem. J.* **473**:1329–1341, 2016.

Screening vs. Selection

- **Individual Assays: Screening** involves testing every single variant individually, which is thorough but limits library size to thousands.
- **Survival Pressure: Selection** only allows desired variants to survive a specific challenge, such as growth or a toxic substrate.
- **Throughput Power:** Selection is vastly more powerful because it can interrogate libraries containing **millions or billions** of unique variants.



Workflow of screens and selections

Plasmids 101: Screens vs. Selections By Susanna Stroik

Video - Concepts in the module or a demonstration

- [Using Directed Evolution to Solve Problems | Beyond the Elements](#)
 - two videos, both are good.
 - 2nd video is an application.
- [Directed Evolution of Enzymes: Frances Arnold 2018 Nobel Prize Winner](#) (in class watch until 1:45)
- [Rolling Circle Amplification](#) (2min11s)

More Videos

- [Molecular Breeding & Genomics Technology for Plant Diversity](#) (3min)
- [Directed Evolution: Frances Arnold](#) (5min14s)

“Nature has explored only a tiny fraction of the life and life’s molecules that are possible. With evolution in our hands, with the ability to set genetic diversity and to tailor the forces of selection, we can now explore paths that Nature has left unexplored.”

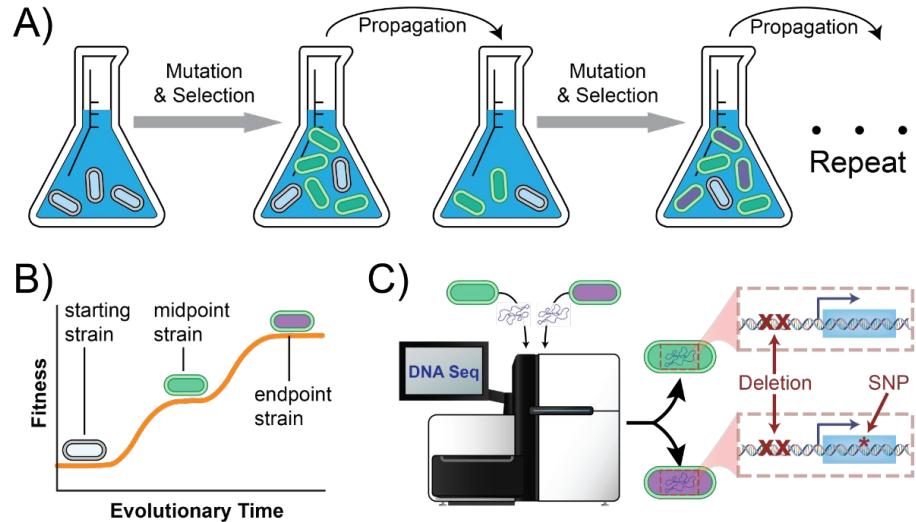
FRANCES ARNOLD

Module 2: Adaptive Laboratory Evolution (ALE) (ALE Review Article, Met. Eng. Journal 2019, doi: [10.1016/j.ymben.2019.08.004](https://doi.org/10.1016/j.ymben.2019.08.004))

- Harness natural selection *in vivo*.
- Optimize growth-coupled cellular phenotypes.
- Automated platforms enable rapid discovery and experimentation.

Whole-Cell (ALE) vs. Directed Evolution

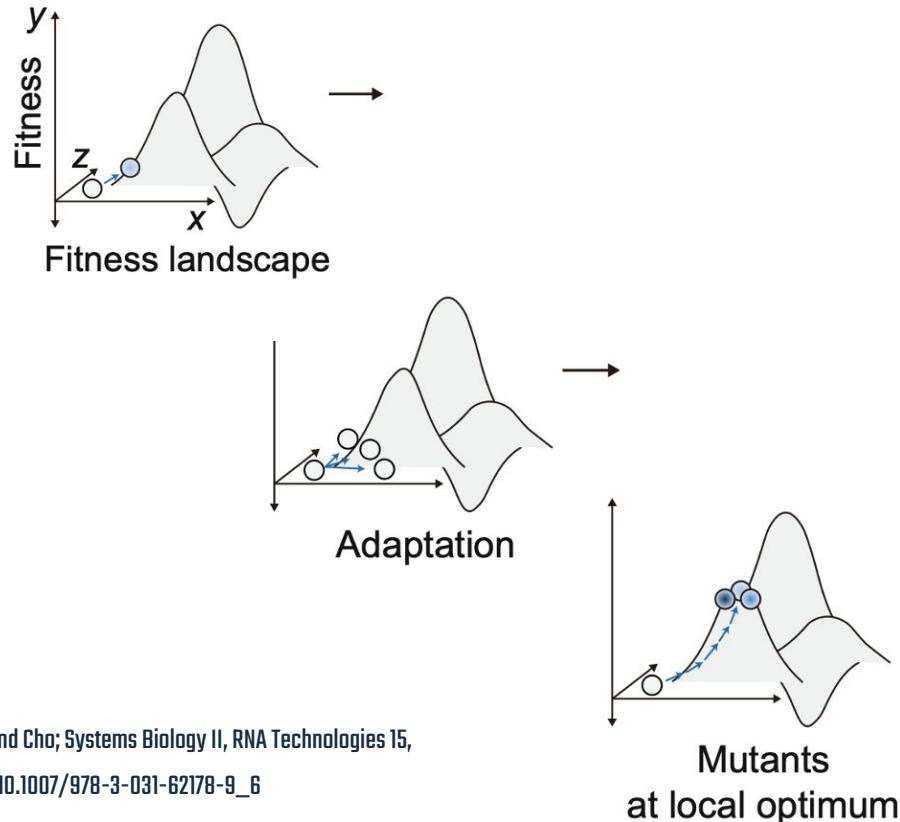
- **Genome Targets:** Unlike directed evolution (Module 1), ALE targets the entire genome within a living, functioning cell.
- **Selection Force:** ALE relies on biological fitness, specifically the growth rate, as the primary force for identifying beneficial mutants.
- **Autonomous Mutation:** The cell does the work of generating mutations and the method selects for those that improve survival in the environment.



Sandberg et al. and Feist. Metab Eng. 2019
doi: 10.1016/j.ymetben.2019.08.004

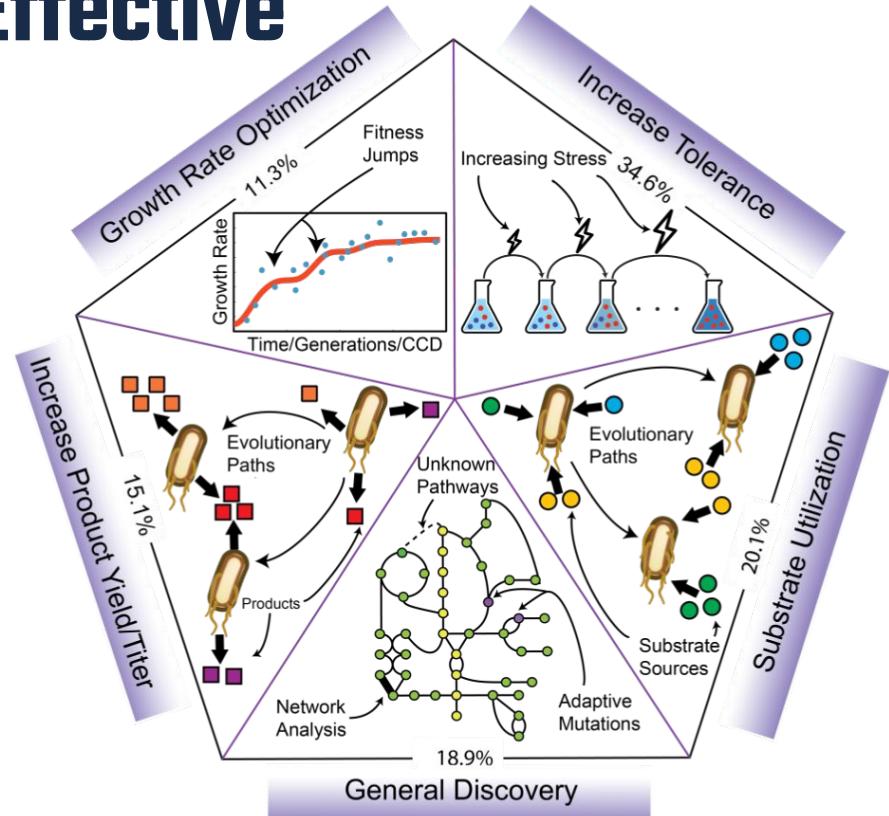
Key Parameters: Population Size

- **Mutation Supply:** Large populations (10^8 - 10^{10} cells) provide a high supply of random mutations for selection.
- **Landscape Access:** A high mutation supply rate increases the probability that the population will find a high fitness peak.
- **Numbers Game:** The sheer number of cell divisions is the fundamental engine that drives rapid laboratory evolution.



Where ALE is Extremely Effective

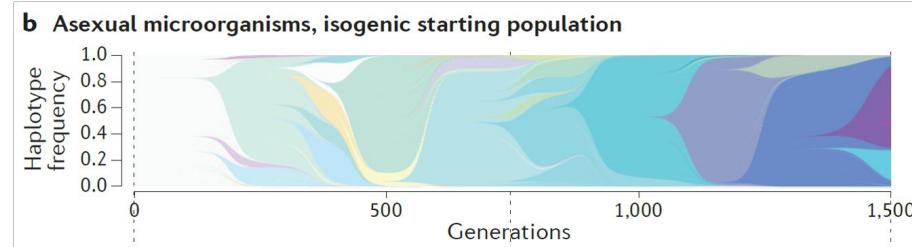
- Tolerant Host Strains
- Increasing Substrate Uptake
- Optimizing Engineered Strains
- Weaning Off Supplement on to New Substrates
- Finding Failure Modes and Escape Routes
- Altering / Improving Kinetic Properties of Enzymes
- Finding Transporters
- Identifying Degradation Pathways



Clonal Interference

- **Haplotype Competition:** Multiple beneficial mutations arise in different lineages and compete within the same asexual population.
 - **Haplotype** - a set of genetic determinants located on a single chromosome (in one strain).
- **Lineage Fixation:** Typically, the single most fit lineage will outcompete all others and reach "fixation," becoming the dominant genotype.
- **Diverse Solutions:** Competition explains why parallel replicates can arrive at different genetic solutions to the same industrial problem.

Muller plot

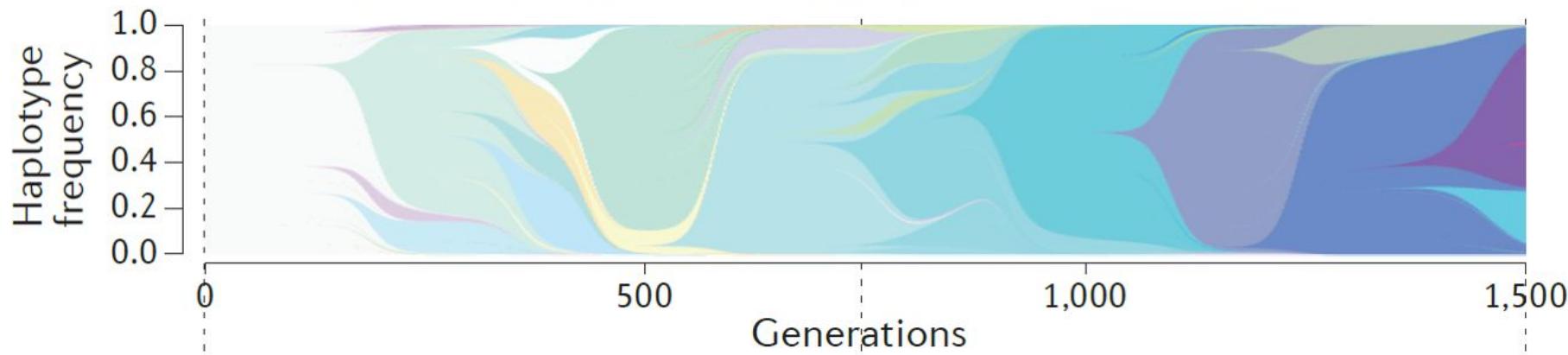


Nature Reviews Genetics 16, 567–582 (2015) doi:10.1038/nrg3937

Clonal Interference

Muller plot

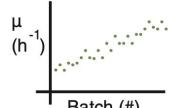
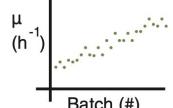
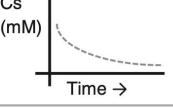
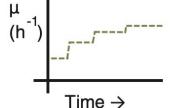
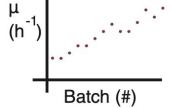
b Asexual microorganisms, isogenic starting population



Batch vs. Continuous Culture

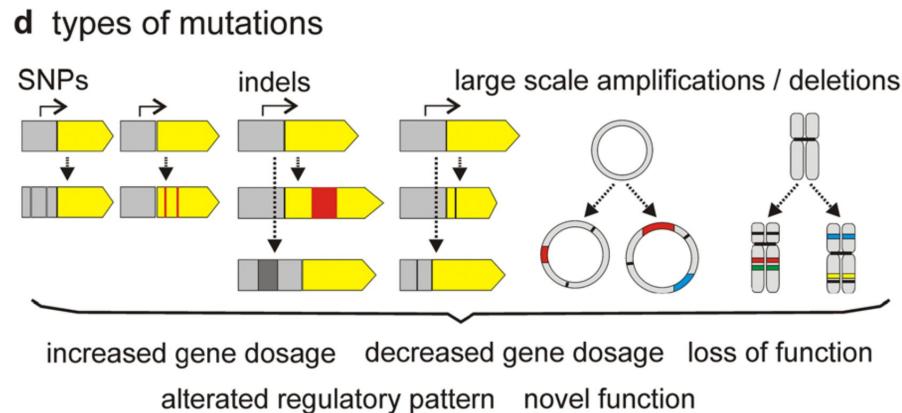
Mans, Daran and Pronk. - CoIB 2018, 50:47–56
10.1016/j.copbio.2017.10.011

- **Serial Batch:** Frequent dilution into fresh medium selects for cells with the highest maximum growth rate (μ_{\max}).
- **Chemostat Selection:** Providing nutrients at a constant, limiting rate selects for cells with high substrate affinity, with a low residual nutrient concentration (C_s).
- **Regime Outcome:** The specific environmental setup determines exactly which biological "bottleneck" the cells will evolve to overcome.

Strategy	Typical output	Characteristics
Serial shake flask (SF) cultivation		<ul style="list-style-type: none">• Simple• Cheap• Compatible with robotization
Sequential batch reactor (SBR)		<ul style="list-style-type: none">• Controlled cultivation• Empty-refill cycles easily automated• On-line analysis of e.g. CO2 production
Chemostat cultivation		<ul style="list-style-type: none">• No empty-refill cycles• Selection for substrate affinity• Selection for mixed substrate utilization
Accelero-/turbido-/auxostat		<ul style="list-style-type: none">• No empty-refill cycles• On-line feedback to control dilution (growth) rate• Selection for mixed substrate utilization
Dynamic selection pressure (SF/SBR)		<ul style="list-style-type: none">• Selection for constitutive improved phenotype• Selection for mixed substrate utilization

Types of Mutations Found

- **SNPs and Indels:** Point mutations and small insertions/deletions (Indels) account for the majority of genetic changes in evolved strains.
- **Structural Variation:** ALE often uncovers large-scale gene duplications or deletions that increase gene dosage for critical pathways.
- **Regulatory Changes:** Mutations frequently occur in global transcriptional regulators, which coordinate the expression of hundreds of different genes.



Automated ALE Platforms

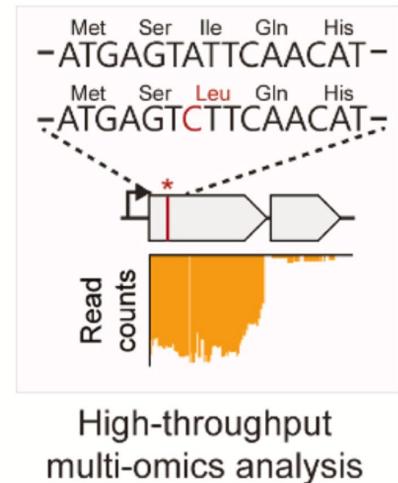
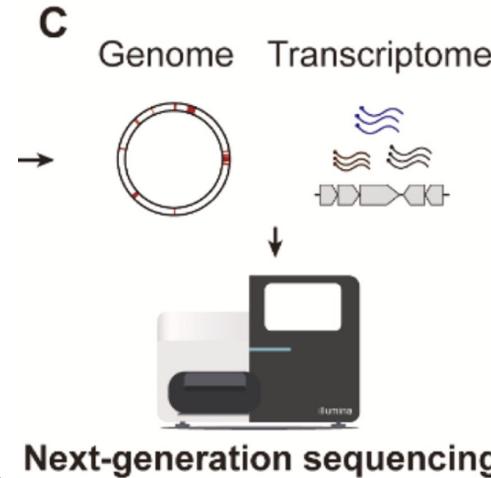
- **Robotic Propagation:** Systems like ALEbot handle 24/7 liquid handling, significantly reducing researcher fatigue and experimental error.
- **Tight Control:** Automation allows for high-frequency monitoring and strictly controlled environments to isolate true causal mutations.
- **Parallel Replicates:** Running dozens of experiments simultaneously provides the statistical power to identify convergent evolution across different lineages.



ALEbot v3.0 - Feist Lab - UC San Diego

Multi-Omics Validation

- **Causality Identification:** Recurrent mutations across independent lineages strongly imply they are the "drivers" of the new trait (Lecture 2 - sequencing)
- **Regulatory Landscapes: Transcriptomics** (RNA-seq) reveals how the cell's internal priorities and pathways have been globally remodeled (Lecture 3).
- **Reverse Engineering:** Scientists prove causality by re-introducing discovered mutations into the ancestor using CRISPR-Cas9 to recreate the evolved phenotype (Lecture 2).

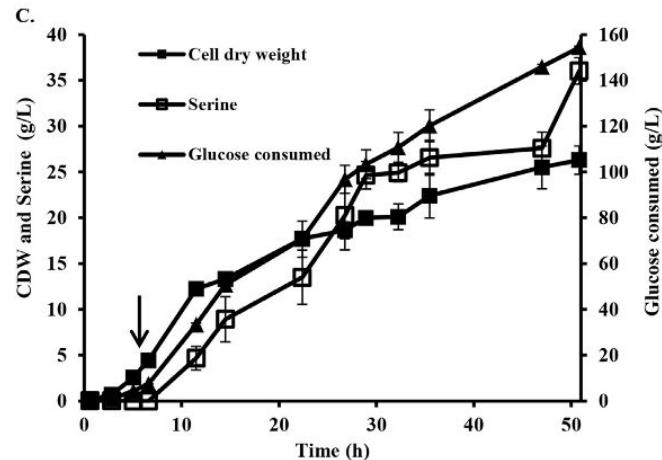
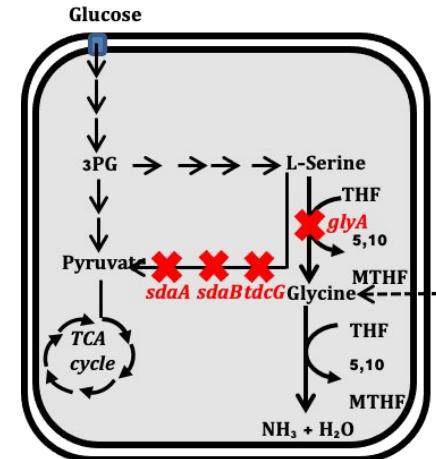


Kim et al. 2022, <https://doi.org/10.1016/j.csbj.2022.12.042>

Case Study: L-Serine Tolerance (TALE)

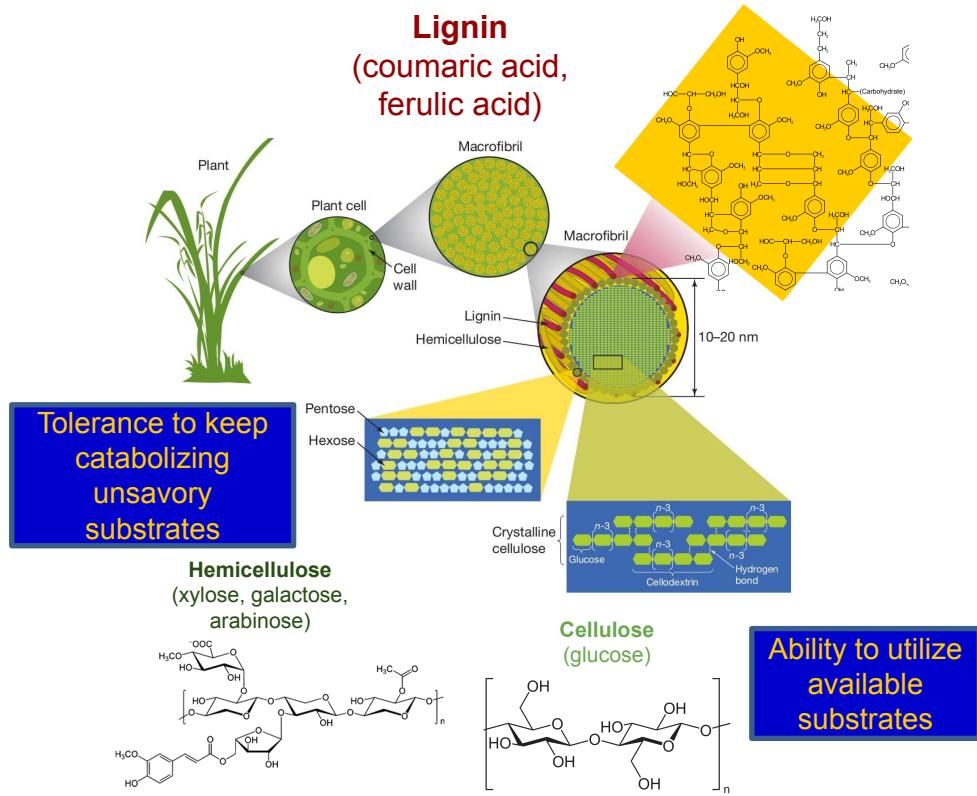
- Product Toxicity:** L-serine is a critical chemical but is highly toxic to *E. coli*, causing severe growth inhibition.
- Tolerance ALE:** Using TALE (Tolerance ALE), the host strain was evolved under increasing serine levels to "fix" growth defects - 3 different mutation targets.
- Record Titer:** The final evolved strain achieved a significant production titer of 37 g/L of L-serine.

Strain Design -
4 Knock Out

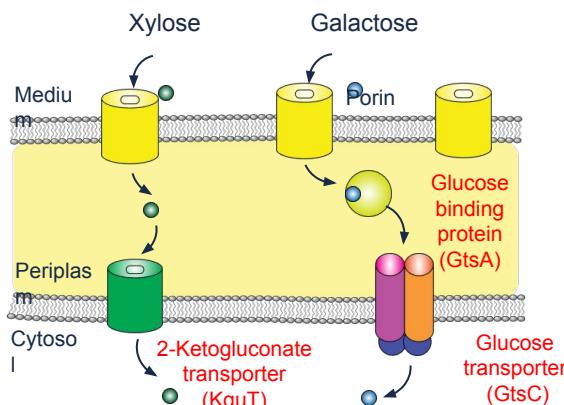
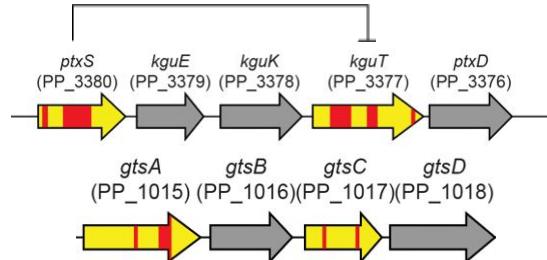
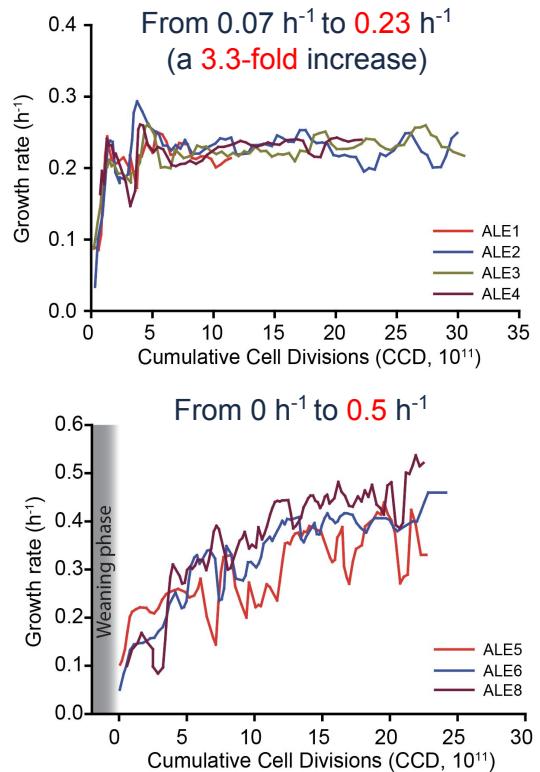
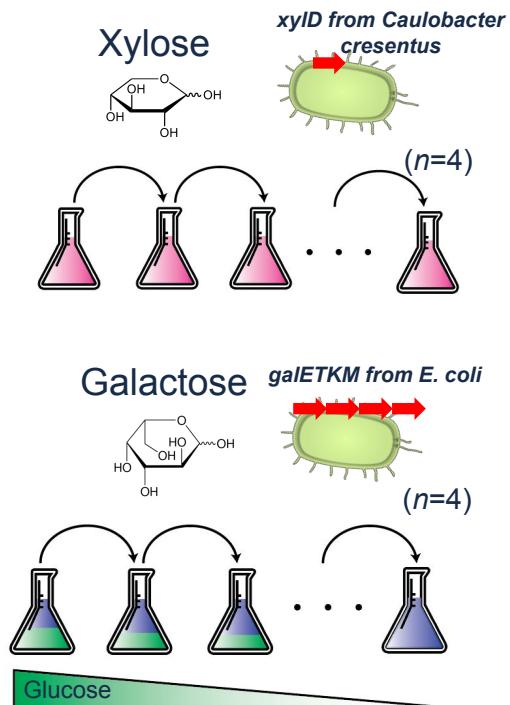


Case Study: Non-Native Sugars (ALE)

- Feedstock Barriers:** *P. putida* cannot naturally grow on xylose or galactose, which are major components of sustainable biomass.
- Pathway Activation:** Introduction of foreign genes was followed by ALE to activate latent transporters and optimize metabolic flux.
- Flux Efficiency:** The evolved strains showed a 3.3-fold growth rate increase, proving that ALE can "tune" newly introduced pathways.

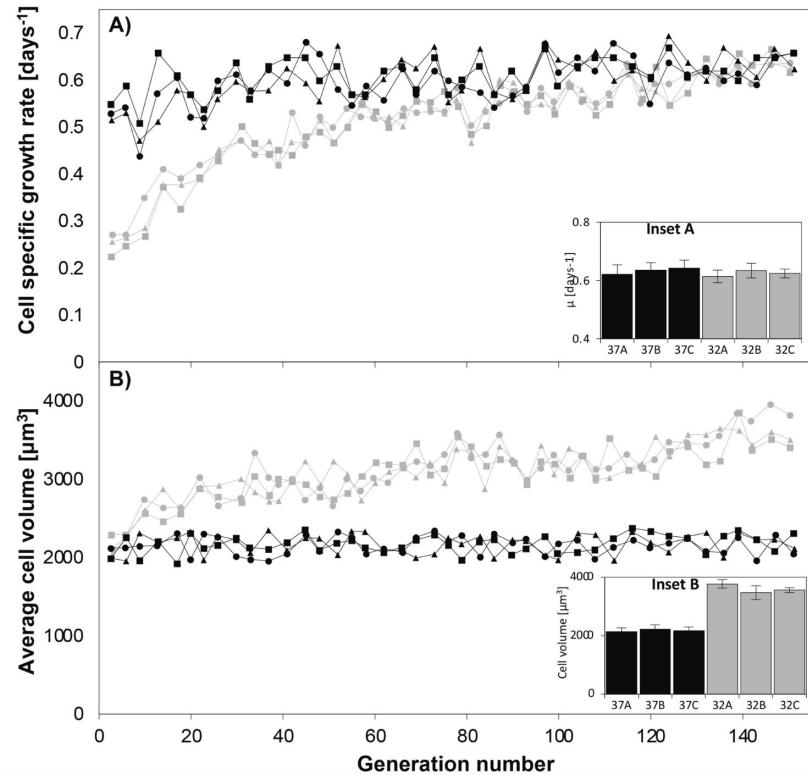


Case Study: Non-Native Sugars (ALE)



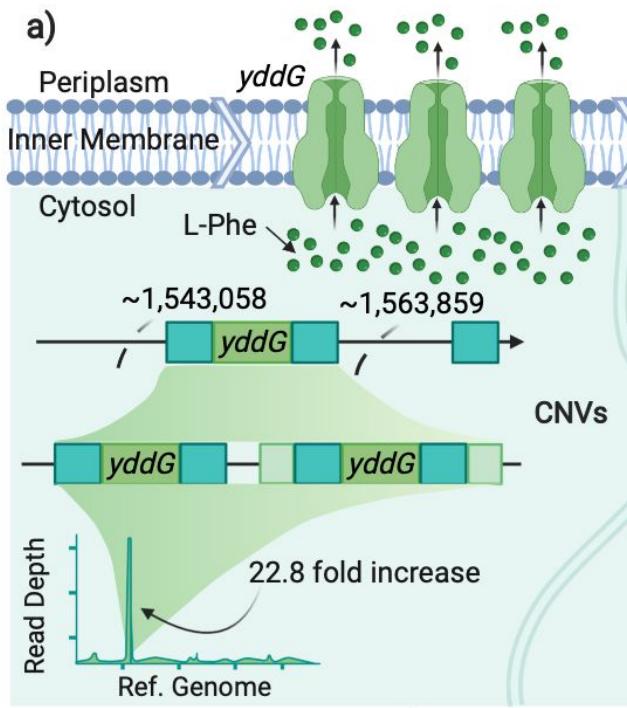
Case Study: CHO Cell Biomanufacturing

- **Mammalian Stalling:** CHO cells produce high-quality antibodies but grow slowly at the lower temperatures required for some folding.
- **Long-Term Propagation:** ALE was used for 150+ generations at 32°C to recover growth rates and enhance biomanufacturing capacity.
- **Titer Improvements:** The resulting cells were larger and produced significantly higher titers of therapeutic proteins like IgG4 and Fc-fusions.



Case Study: Transporter Modulation (TALE)

- **Hidden Exporters:** Transporters make up 10% of the genome, but 31% remain uncharacterized, hiding potential targets for product export.
- **Discovery Mode:** TALE was used to identify amino acid export pumps by selecting for mutations that increased cellular tolerance.
- **Enhanced Export:** Identifying mutations in the *yddG* support the export activity and the physiological impact of its overexpression.



**TALE
Approach**



Potential of Evolutionary Engineering

- **Uncultured Potential:** Since <1% of microbes can be cultured, ALE is used to "domesticate" new species for industrial use.
- **Domestication Strategy:** Evolutionary engineering allows us to utilize the "Uncultured World" discovered in Lecture 3 for green biotechnology.
- **Metabolic Shifting:** Case studies have used ALE to convert heterotrophic *E. coli* into fully autotrophic strains that consume CO₂
- **Industrial Robustness:** Evolved strains are stable for large-scale bioprocessing, contain patentable genetic intellectual property, and represent the next generation of biomanufacturing.

Summary: Protein Engineering Approaches

- Comparing Evolutionary Engineering Approaches with Rational Design

Table 8.4 A comparison of directed evolution, rational design and semi-rational design

	Directed evolution	Rational design	Semi-rational design
Parental gene	A single gene or a group of homologous sequences	A single gene	A single gene
<i>A priori</i> knowledge requirement	Not required	Required	Required
Genetic diversity creation	Random mutagenesis or DNA recombination	Focused mutagenesis	Focused mutagenesis
Library size	Large	Small	Small to medium
Screening	High to ultra high throughput	Low to high throughput	Low to high throughput
Advantages	<ul style="list-style-type: none">No prior knowledge of the enzyme structure and mechanism is required.Mutate the entire enzyme, and as such, it is possible to identify mutations distant to the active site that affect the enzymatic activity via allosteric interaction.	<ul style="list-style-type: none">Small library size.Less time and effort on screening.Particularly advantageous when there is no high-throughput screening system available.	<ul style="list-style-type: none">Library size is significantly reduced compared to directed evolution.A larger portion of the protein sequence space is explored compared to rational design.
Disadvantages	<ul style="list-style-type: none">Large library size.Impossible to explore the full protein sequence space, even with the most powerful selection or screening method.Time consuming to develop an assay and to screen large library.Resource intensive.	<ul style="list-style-type: none"><i>A priori</i> knowledge is required.Mutations are mainly targeted at the active site.	<ul style="list-style-type: none"><i>A priori</i> knowledge is required.Mutations are mainly targeted at the active site.

Video - Concepts in the module or a demonstration

- The Long Term Evolution Experiment
- Robotic system can fast-forward evolution to create useful microbes

The End